Environmental Inequality in Germany

Vom Fachbereich Sozialwissenschaften
der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Philosophie (Dr. phil.)
genehmigte

Dissertation

vorgelegt von
Tobias Rüttenauer

Tag der Disputation:	Kaiserslautern, 28. September 2018
Dekanin:	Prof. Dr. Shanley E. M. Allen
Vorsitzende/r:	Prof. Dr. Thomas Lachmann
Gutachter/in:	1. Prof. Dr. Henning Best
	2. Prof. Dr. Peter Preisendörfer

D 386

Kaiserslautern, Dezember 2018

Acknowledgements

Writing a dissertation is a quite interesting adventure, but can also be hard work from time to time. Most importantly, it is a long process and it would not be possible without the assistance of colleagues and friends. Thus, I am very thankful to all the people who supported me during the process of writing this book.

First and foremost, I would like to thank my advisor Henning Best. You did not only teach me how to organise every day life at the university, but also contributed a lot to my dissertation. I could always discuss important matters with you when I was stuck, and I received a lot of helpful feedback and new insights from you. Though it sometimes took me a while to follow your advise, I always appreciated your help. Thank you very much for the ongoing support.

Furthermore, I am very grateful to my colleagues and friends for not getting tired discussing various aspects of this work with me. A special thanks goes to Volker Ludwig and Scott Cook for helping me to understand several methodological issues occurring throughout this project, and to Peter Preisendörfer for the effort of reading and judging this dissertation. I would also like to thank Julia Schulte-Cloos for helpful feedback on several chapters of this book, and for reminding me of the question 'so what?' (besides other important comments). I also gained a lot from my next-door colleague Phil Kolbe, who listened to my monologues even when occupied with more important tasks. Furthermore, this project profited a lot from ongoing discussions with Henrik Andersen, Tanja Dannwolf, Christoph Giehl, Uta Landrock, Jochen Mayerl, Thomas Morris, and Ingmar Rapp. Many thanks also to Charlotte Haußmann, Leonore Röseler, and Elo Schneider for the assistance at several stages of my work.

Moreover, I would like to thank several anonymous reviewers for providing helpful feedback on single chapters of this book. Many thanks also to Joanne Hall, Michèle Parent, and members of the Social Sciences peer critique group for helping to improve the writing and the language of this book.

Last but not least, I am very grateful to my parents Ulrike and Clemens. Thank you a lot for making all this possible, for always having an open ear, and for supporting me all the long way in academia.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Aims and scope

Only recently, *The Lancet* published a report identifying environmental pollution as the major cause of diseases and premature deaths worldwide (Landrigan et al., 2018). The study attributes approximately 9 million premature deaths in 2015 (16% worldwide) to diseases caused by pollution, thereby exceeding the death toll of all wars or other violent acts within the same period. Besides the severe impacts of environmental pollution on human health, environmental pollution also affects other dimensions of social live. For instance, studies have shown significant effects of environmental pollution on psychological health (e.g. Gascon et al., 2018; Sass et al., 2017; Zhang, Zhang & Chen, 2017), physical activity (e.g. An, Zhang, Ji & Guan, 2018; Bresnahan, Dickie & Gerking, 1997; Roberts, Voss & Knight, 2014), and happiness (e.g. Ambrey, Fleming & Chan, 2014; Levinson, 2012; Schmitt, 2013). Yet, environmental pollution is not equally distributed across the society and mostly affects the poor and minorities. Because of its severe consequences for the population at risk, this unequal distribution of pollution constitutes a serious dimension of social inequality, which is even linked to mortality.

As one of the first research studies on the topic of environmental inequality, the report of the United Church of Christ - Commission for Racial Justice (1987) showed that minorities and the socio-economically disadvantaged live disproportionately close to hazardous industrial facilities. Ever since, environmental inequality has been on the agenda of social science research for several decades in the United States (e.g. Anderton, Anderson, Oakes & Fraser, 1994; Been, 1994; Bryant & Mohai, 1992; Bullard, 1990). Those and other studies (for an overview e.g. Mohai & Saha, 2015a; Ringquist, 2005) have documented a persistently high disadvantage of minority groups regarding the environmental quality of their living environment. As Mohai, Pellow and Roberts (2009, p. 406) put it: ‘Today, hundreds of studies conclude that, in general, ethnic minorities, indigenous persons, people of color, and low-income communities confront a higher burden of environmental exposure from air, water, and soil pollution from industrialization’.

Though environmental inequality research has gained extensive interest in the United States, it has received far less attention in Europe and Germany. Only recently, a few empirical studies (Diekmann & Meyer, 2010; Flacke, Schüle, Köckler & Bolte, 2016; Kabisch & Haase, 2014; Kohlhuber, Mielck, Weiland & Bolte, 2006; Raddatz & Mennis, 2013) have started to provide first evidence for a disproportionately high burden of minorities and (somewhat less consistent) of the socio-economically disadvantaged in the German speaking area. For instance, these studies demonstrate that foreigners in Germany generally perceive a higher impairment through air pollution than native German citizens (Kohlhuber et al., 2006), experience a lower provision of public green-space around their place of residence (Kabisch & Haase, 2014), or live closer to industrial facilities (Raddatz & Mennis, 2013). Similarly, high-income households report a lower impairment through air pollution (Kohlhuber et al., 2006), and are less exposed to chemical compounds and particulate matter (Flacke et al., 2016). Though these studies provide some evidence for the existence of environmental inequality in the German speaking area, research in Europe and especially in Germany lags far behind the state of research in the United States. Studies in Germany have analysed either single regions (Flacke et al., 2016; Kabisch & Haase, 2014; Raddatz & Mennis, 2013) or used subjective measures of air pollution (Kohlhuber et al., 2006). Especially in a cross-sectional design, it is thus difficult to evaluate the state-wide extent of environmental inequality and to draw conclusions about the causal mechanisms at work. Consequently, Germany lacks a nationwide and systematic analysis of the extent and the causal mechanisms of environmental inequality.

The main objective of this book is to extend the research on environmental inequality in Germany. This book aims to shed more light on the question of whether minorities in Germany are affected by a disproportionately high burden of environmental pollution, and to increase the general knowledge about the causal mechanisms, which contribute to the unequal distribution of environmental hazards across the population. In general, this book deals with two main research questions:

1. Are minorities in Germany disproportionately exposed to environmental pollution?
2. What causes this disproportionate exposure of minorities to environmental pollution?

To improve our knowledge about environmental inequality in Germany, this book extends previous research in several ways. First, to evaluate the extent of environmental inequality, this book relies on two different data sources. On the on hand, it uses household-level survey data and self-reports about the impairment through air pollution. On the other hand, it combines aggregated census data and objective register-based measures of industrial air pollution by using geographic information systems (GIS). *Consequently, this book offers the first analysis of environmental inequality on the national level that uses objective measures of air pollution in Germany.* Second, to evaluate the causes of environmental inequality, this book applies a panel data analysis on the household level, *thereby offering the first longitudinal analysis of selective migration processes outside the United States.* Third, it compares the level of environmental inequality between German metropolitan areas and evaluates to which extent the theoretical arguments of environmental inequality can explain differing levels of environmental inequality across the country. By doing so, this book not only investigates the impact of indicators derived by the standard strand of theoretical reasoning but also includes structural characteristics of the urban space.

Though Chapter 2 also investigates the socio-economic dimensions of selective migration processes and resulting disadvantages in environmental pollution, the main focus of this book lies on the disadvantage of minorities for two reasons. First, previous research in the United States has identified the minority-majority devide as the more consistent and robust finding (Ringquist, 2005). Second, data limitations hinder the use of socio-economic variables on a fine-grained spatial level in Germany. Still, the following sections will also discuss the theoretical arguments and previous findings regarding socio-economic indicators, as socio-economic resources constitute one major explanation of the minority-majority divide. Most prominently, Wilson (1978, 1987) has argued that disadvantages of racial minorities in the United States mostly stem from socio-economic disadvantages rather than from race or ethnicity itself.

Note that many scholars use the terms of environmental inequality and environmental justice interchangeably. Yet, this book is primarily concerned with environmental inequality per se, thereby concentrating on the unequal distribution of environmental pollution across the society. Consequently, this book does not deal with normative questions of justice or injustice. Whether a given distribution of pollution can be evaluated as just or fair depends on the underlying concept of justice. For example, it can be argued that justice does not only depend on the distribution of environmental burdens but also on the distribution of profits generated by emitting industries or the usage of polluting technologies (for a discussion see e.g. Preisendörfer, 2014). However, before discussing the dimensions of justice, it is important to offer an empirical foundation investigating the extent of environmental inequality and its mechanisms. Hence, I will deal with the analyses of inequality rather than justice in this book.

1.2 Theoretical mechanisms

When analysing the causal mechanisms of environmental inequality, most research in the United States points to two different explanation: selective siting and selective migration (e.g. Been & Gupta, 1997; Crowder & Downey, 2010; Hamilton, 1995; Sieg, Smith, Banzhaf & Walsh, 2004). The argument of selective siting assumes that industrial facilities are sited disproportionately close to areas with a high proportion of minorities and poor households. The argument of selective migration, in contrast, assumes that minorities and poor households selectively move into areas with an already high level of environmental pollution, while majority or affluent households selectively leave those areas. Accordingly, one of the main concerns in environmental inequality research is the question of 'which came first?' (Pastor, Sadd & Hipp, 2001). The following sections will outline both theoretical arguments in more detail.

1.2.1 Selective siting

The selective siting argument claims that hazardous facilities are disproportionately sited in neighbourhoods that already face a high proportion of minorities or low-income households (Been & Gupta, 1997; Mohai & Saha, 2015a; Pastor et al., 2001; Saha & Mohai, 2005; Wolverton, 2009). The reason for this selective siting behaviour can be threefold.

First, selective siting might be the results of taste-based discrimination. If decision makers belong predominantly to the majority group and are aware of the danger or the burden of industrial facilities, they might want to externalise unwanted facilities onto minority groups, while protecting members of their own group (Hamilton, 1995).

Second, the market explanation assumes that companies seek to minimise their land and housing costs when looking for the location of new facilities. Because of lower land prices and housing costs, poor regions are an attractive siting location for new facilities (Downey, 2005; Farber, 1998; Saha & Mohai, 2005; Wolverton, 2009, 2012). Furthermore, compensation costs for environmental pollution are lower in poor neighbourhoods, as low-income groups allegedly value environmental quality less than high-income groups do (Mohai & Saha, 2015a; Saha & Mohai, 2005; Wolverton, 2009, 2012). Therefore, it is a rational strategy for a profit-maximising company to place facilities in low-income areas. If minorities are overrepresented in low-income areas, they are disproportionately exposed to industrial facilities.

Third, the social and political capital explanation assumes that minorities and low-income households hold a lower social and political capital. Hence, inhabitants of regions with a high minority share are less likely to organise collective protests against hazardous facilities (Hamilton, 1995; Mohai & Saha, 2015a; Pastor et al., 2001). They are less likely to influence political decisions, to engage in collective action (to reach e.g. a ban of hazardous facilities), or to take legal actions (Wolverton, 2009). High-income neighbourhoods with a low minority share, in contrast, are much more likely to influence political actors not only due to their social ties, but also due to their political engagement or civic activism. Furthermore, affluent households are able to afford expensive legal proceedings. Because civil protest, political or legal actions against the siting of facilities may cause financial harms or bad publicity, the respective executive decision makers try to avoid those situations: they choose the 'path of least political resistance' (Saha & Mohai, 2005) and — in anticipation of potential protest in affluent neighbourhoods — selectively settle in socio-economically disadvantaged regions.

In sum, those fine-grained mechanisms predict that minorities face a higher burden of environmental pollution because facilities are disproportionately sited in areas with a high proportion of minorities. Temporally, this means that the minority population already resided in a specific area before this area received newly sited industrial facilities, and in consequence experienced a high amount of environmental pollution. Interestingly, only the first sub-mechanisms assumes that the siting decisions is a direct consequence of race or ethnicity itself. The second and third mechanism, in contrast, assume that minorities lack either economic or socio-political resources, leading to the disproportionate siting of facilities in areas with a high minority share.

1.2.2 Selective migration

The second major explanation of environmental inequality proposes a competing mechanism. In general, this second approach follows a more general strand of research on differential migration and neighbourhood-attainment patterns, showing that minorities face difficulties to access high-quality neighbourhoods and to escape low-quality neighbourhoods (e.g. Alba, Logan, Stults, Marzan & Zhang, 1999; Crowder & South, 2005; Crowder, South & Chavez, 2006; Logan & Alba, 1993; South & Crowder, 1997; South, Huang, Spring & Crowder, 2016). Based on these more general patterns, the selective migration argument assumes that the unequal exposure of disadvantaged citizens stems from selective migration processes into and out of polluted areas (Banzhaf & McCormick, 2012; Banzhaf & Walsh, 2008; Crowder & Downey, 2010; Mohai & Saha, 2015a; Pais, Crowder & Downey, 2014; Sieg et al., 2004). This means that socio-demographic changes in polluted areas sequentially follow the siting process. Again, two different sub-mechanisms support this theory.[1]

First, the 'racial residential discrimination thesis' assumes that minorities are steered into polluted areas because of discriminatory barriers on the housing market. On the one hand, housing agents or property owners may fear declining attractiveness or declining prices due to minority inhabitants. For instance, research has shown that the minority share has a significant impact on the perception of neighbourhood crime rates, indepedent of the actual crime rate (Massey & Denton, 1993; Semyonov, Grodzeisky & Glikman, 2012). Hence, housing agents or landlords may prefer renters belonging to the majority group and discriminate against minority applicants (Turner & Ross, 2005; Yinger, 1986). On the other hand, housing agents may also spuriously anticipate that minorities have lower preferences for high environmental quality and thus restrict the housing opportunities to a subset with lower quality (Ondrich, Ross & Yinger, 2003; Turner & Ross, 2005). Importantly, this mechanism assumes that the selective migration patterns of minority households are a direct consequence of the minority status and are independent of socio-economic resources.

[1] Note that similar sub-mechanisms in the more general literature on neighbourhood-attainment processes are called 'place stratification' and 'spatial assimilation' models (e.g. Crowder & South, 2005; Crowder et al., 2006; South et al., 2016).

Second, the 'racial income-inequality thesis' assumes that minorities are mainly forced to move into polluted areas as a side product of their lower economic resources. In general, this argument follows Tiebout's (1956) model of the 'consumer-voter'. Hereafter, households are assumed to have specific preferences for the provision of public goods and aim to satisfy these preferences. As local communities have relatively stable revenues and expenditure plans, households can adjust the level of public goods provision to their preferences mainly by moving between the communities. Households are 'voting with their feet'. Because households prefer a higher environmental quality over a lower one (Bayer, Keohane & Timmins, 2009), there is a higher demand for clean neighbourhoods. Given the finite amount of high environmental quality neighbourhoods, a higher demand for housing in clean areas also increases the housing and land prices in these areas (Banzhaf & McCormick, 2012; Bayer et al., 2009; Farber, 1998). Furthermore, previous research has shown that households are willing to pay more for a clean environment as their income rises (Banzhaf, Sidon & Walsh, 2012; Liebe, Preisendörfer & Meyerhoff, 2010; Meyer & Liebe, 2010). High-income households are therefore more likely to move out of low-quality neighbourhoods (selective out-migration) because they can afford to do so and are more willing to pay higher housing prices in clean neighbourhoods. Simultaneously, low-income households are more likely to move into low-quality neighbourhoods (selective in-migration), as these neighbourhoods are more affordable and, thus, more attractive to low-income households. In sum, minorities sort into low-quality neighbourhoods just because they cannot afford high-quality areas with high-priced housing opportunities. In contrast to the first explanation, the 'racial income-inequality thesis' posits that selective migration trajectories are not a function of minority status itself, but rather a consequence of differently distributed socio-economic resources.

Similar to the predictions of the selective siting argument, both sub-mechanisms of selective migration predict that minorities bear a disproportionate burden of environmental pollution. However, the argument of selective migration presumes a different temporal order than the argument of selective siting. According to the theory of selective migration, areas firstly hold a high amount of environmental pollution, and subsequently experience the in-flow of minority households. Consequently, both theories – selective siting and selective migration – predict that minorities end up in closer proximity to industrial pollution, though both theories posit a distinct temporal order. Interestingly and similar to the theory of selective siting, only one fine-grained mechanism of selective migration assumes that selective migration patterns are the consequence of race or ethnicity itself. The second explanation, again, builds on the idea that minorities are steered into low quality neighbourhoods because they posses lower economic resources.

1.3 Previous findings and gaps

As already sated above, previous research (mostly from the United States) has unambiguously shown that minorities bear a disproportionate exposure to environmental risks in various dimensions. For instance, a vast body of research shows that minorities are exposed to higher amounts of hazardous industrial air pollution (e.g. Ard, 2015; Ash & Fetter, 2004; Downey, 2005; Morello-Frosch, Pastor & Sadd, 2001; Pastor, Morello-Frosch & Sadd, 2005; Pastor, Sadd & Morello-Frosch, 2002). Other studies find that minorities live closer to industrial facilities (e.g. Downey, 2003; Mohai, Lantz, Morenoff, House & Mero, 2009; Raddatz & Mennis, 2013) or toxic waste sites (e.g. Hipp & Lakon, 2010; Mohai & Saha, 2007). Still others demonstrate that minorities are also disproportionately affected by pollution stemming from mobile sources like traffic (e.g. L. P. Clark, Millet & Marshall, 2014; Diekmann & Meyer, 2010; Moreno-Jiménez, Cañada-Torrecilla, Vidal-Domínguez, Palacios-García & Martínez-Suárez, 2016) or are provided with lower amounts of green-space around their residences (e.g. Casey, James, Cushing, Jesdale & Morello-Frosch, 2017; Kabisch & Haase, 2014). Though some studies provide none or only weak support for ethnic disparities (e.g. Anderton et al., 1994; Bowen, Salling, Haynes & Cyran, 1995), the majority of studies confirms the disadvantage of minorities. In a meta-analysis of 49 environmental inequality studies, Ringquist (2005) finds a consistent disadvantage of racial minorities throughout all studies, though the strength of this effect may vary with applied methods and study area.

Regarding the disproportionate exposure of economically disadvantaged groups, results seem to be less consistent. Though higher environmental risks, on average, are concentrated in socio-economically weak communities, this finding highly depends on the type of measurement and the applied method (Ringquist, 2005). For instance, studies show that aggregated income is negatively associated with the level of hazardous industrial pollution (e.g. Ash & Fetter, 2004; Downey & Hawkins, 2008) or the presence of industrial facilities (e.g. Mohai & Saha, 2007; Wolverton, 2009). This is in line with results on the individual level, reporting a lower burden with increasing household income (Crowder & Downey, 2010; Mohai, Lantz et al., 2009). However, there is also a quite large body of research that finds no or only weak evidence for the association between socio-economic indicators and environmental pollution (e.g. Arora & Cason, 1999; Baden & Coursey, 2002; Davidson & Anderton, 2000; Diekmann & Meyer, 2010; Havard, Deguen, Zmirou-Navier, Schillinger & Bard, 2009). Baden and Coursey (2002), for example, find that income correlates negatively with the presence of hazardous waste sites in earlier decades, but not with newly sited industrial facilities since 1990. In a later study, Baden, Noonan and Turaga (2007) demonstrate that conclusions crucially depend on the geographic scale, as income correlates with the presence of superfund sites only on some spatial scales. Besides studies from the United States, studies in Europe found rather weak or inconclusive results regarding the correlation between socio-economic status and environmental pollution. For instance, Diekmann and Meyer (2010) report only negligible effects of income on the level of pollution, and Havard et al. (2009) find a curvilinear relationship between the deprivation level of an area and pollution, with midlevel areas exhibiting the highest level of pollution. Similarly, Padilla et al. (2014) obtain opposing results in different French cities: While deprived areas in Lille and Marseille experience higher levels of nitrogen dioxide (NO_2), Paris exhibits a negative correlation between deprivation and pollution, and Lyon a curvilinear relationship. The fact that many studies report inconclusive results regarding the connection between socio-economic resources and environmental pollution seems somewhat surprising given that many explanations for the minority-majority divide rest upon the assumption that socio-economic resources are decisive.

In accordance with those mixed findings on the socio-economic dimension of environmental inequality, scholars offer a rather inconclusive picture of the causal mechanisms producing the disproportionate exposure of minorities. When looking at cross-sectional studies, a high discontent appears on whether the disadvantage of minorities is a function of the minority status itself or rather the result of economic resources. Some studies, for example, report a declining or even vanishing correlation between the minority share and pollution when controlling for the aggregated income (Ash & Fetter, 2004; Baden & Coursey, 2002; Bowen et al., 1995; Downey, 2006a). These findings support the argument that the disproportionate exposure of minorities is mainly a by-product of their lower socio-economic resources. Still, other studies do not find any reduction of the minority-majority divide due to accounting for income (Daniels & Friedman, 1999; Morello-Frosch et al., 2001), or show that conclusions depend on the level of aggregation (Downey, 1998). Similarly, income accounts only for a minor part of lower moving returns experienced by minority households (Crowder & Downey, 2010; Pais et al., 2014, see also below). These findings, in contrast, contradict the assumption that the disproportionate burden of minorities stems from the uneven allocation of socio-economic resources.

Also when turning to the question of 'which came first' – or more precisely, whether selective siting or selective migration is responsible for environmental inequality – results seem rather mixed. While some longitudinal studies investigating the temporal order of facility siting and socio-demographic changes find evidence for selective siting processes (Funderburg & Laurian, 2015; Mohai & Saha, 2015b; Pastor et al., 2001), others find none or only weak support (Been & Gupta, 1997; Downey, 2005; Oakes, Anderton & Anderson, 1996; Shaikh & Loomis, 1999). For instance, the results of Pastor et al. (2001) exhibit a significantly higher probability of receiving a hazardous waste facility within a quarter mile distance for areas with a higher share of minorities. In contrast, Been and Gupta (1997) find a correlation between minority share and facility presence prior to 1970, but no correlation for newly sited facilities, and Downey (2005) concludes that the change in the number of facilities is neither correlated with the total percentage of Black inhabitants nor with the change in the percentage of Black inhabitants. Furthermore, other studies show that historical patterns like agglomeration economies (Elliott & Frickel, 2013), waterways within the city (Baden & Coursey, 2002), or infrastructural opportunities (Wolverton, 2012) are more decisive for the appearance of new facilities than the socio-demographic composition of the nearby inhabitants. Thus, evidence for selective siting processes is rather mixed.

Similarly, longitudinal studies investigating selective migration patterns provide rather inconclusive results. Most studies using aggregated data rather contradict the selective migration argument (Downey, 2005; Funderburg & Laurian, 2015; Hunter, White, Little & Sutton, 2003; Mohai & Saha, 2015b; Oakes et al., 1996; Pastor et al., 2005, 2001). For instance, Mohai and Saha (2015b) show that the share of minorities in census tracts receiving a new facility increased temporally prior to the siting of facilities and not due to subsequent migration processes. Similarly, Pastor et al. (2001) find rather declining minority shares in areas that received a new facility, and Hunter et al. (2003) do not find selective out-flows from polluted areas, thereby contradicting the theoretical prediction of selective migration. However, those longitudinal studies relying on individual or household level data provide empirical evidence for selective migration patterns (Crowder & Downey, 2010; Pais et al., 2014). Crowder and Downey (2010), for example, reveal that Black households experience significantly higher environmental pollution in the area of destination when moving to a new place of residence. Pais et al. (2014) corroborate this finding by showing that Black households face a significantly higher probability of following a permanently high-pollution moving-trajectory. Interestingly, both studies on the individual level find only a marginal reduction of the minority effect when controlling for income. This finding is also in line with more general findings of migration research, demonstrating that differences in neighbourhood attainment cannot be explained by socio-economic differences (South et al., 2016) and that minorities experience lower payoffs from socio-economic resources (Logan & Alba, 1993). Thus, previous research indicates that lower socio-economic resources can only marginally explain differing moving trajectories.

Note that most of the studies investigating the causal mechanisms of environmental inequality were conducted in the United States. Thus, it is not clear to which extent findings regarding the causal mechanisms are transferable to the German context. First, environmental inequality in the United States is mostly concerned with ethnic minorities like Black Americans, Hispanics, or Asians. In contrast, minorities in Germany are mostly immigrant-minorities (Kalter & Granato, 2007) and stem from relatively recent immigration from other European countries or Turkey. Second, the United States exhibits much higher levels of residential segregation (Musterd, 2005), which might be a quite important factor in the occurrence and development of environmental inequality. Third, the urban structures differ between the United States and European countries. For instance, European cities are more densely populated and more centrally organised (Huang, Lu & Sellers, 2007; Schwarz, 2010), and consequently exhibit different patterns of urban transportation (Buehler, 2011). Fourth, economic inequality in the United States is higher than in Germany (Piketty & Saez, 2014), which might also point towards lower levels of inequality in other dimensions. All those differences could on the one hand affect the extent of environmental inequality, but on the other hand also play an important role regarding the causes of environmental inequality.

In Germany, we face two main gaps in the empirical literature on environmental inequality. First, we lack a national-level evaluation of the presence of environmental inequality that uses objective pollution data. Though a tremendous amount of empirical research in the United States supports the finding that minorities are disproportionately affected by environmental pollution on various scales, it is not clear whether this finding is easily transferable to Germany. As outlined above, previous studies in Germany either use subjective measures of air pollution or restrict the analysis to single regions, thereby not allowing for inference on the country level. Second, we lack a congruent picture on the causes of environmental inequality. This is not a particular problem of German research in this field, but a shortcoming of environmental inequality research in general. Though several studies in the United States use longitudinal data to explicitly investigate the causal mechanisms of environmental inequality, results offer a rather inconsistent picture. Some studies find support for selective siting, other do not. Similarly, some studies find support for selective migration patterns, others do not; and still others rather emphasise the importance of infrastructural characteristics. It is thus the aim of this book to contribute to these gaps in the literature and to advance the research on environmental inequality in Germany, but also to contribute to the more general discussion on the driving forces behind environmental inequality.

1.4 Levels of aggregation

From a methodological point of view, it is important to note that previous research on environmental inequality employed two distinct strategies to evaluate the distributions of environmental pollution across the society. While some studies rely on individual data, others use spatially aggregated data to investigate environmental inequality.

The individual approach follows the ‘standard procedure’ in social sciences and analyses individual or household-level survey data (e.g. Crowder & Downey, 2010; Diekmann & Meyer, 2010; Mohai, Lantz et al., 2009; Pais et al., 2014). This is especially important when analysing the process of selective migration, as the actual moving behaviour of households can be observed. However, data protection issues hinder the use of actual household locations, making it difficult to connect household level data to fine-grained neighbourhood characteristics and objective indicators of environmental quality like air pollution or the distance to industrial facilities. Furthermore, the pure description of the extent of environmental inequality might crucially depend on sampling procedures, as many studies use regionally stratified sampling techniques. The aggregated approach in contrast uses spatial units like neighbourhoods or census tracks as the level of observation (e.g. Banzhaf & Walsh, 2008; Been, 1994; Mohai & Saha, 2015b). This has the advantage that the analyses are based on large samples and provide information on the aggregated dynamics. Especially when analysing selective siting, the average socio-economic status or the aggregated minority share of an area are the theoretically relevant factors for choosing the location of industrial facilities. Still, relying on aggregated data makes it difficult to draw conclusions on individual-level behaviour. For instance, Banzhaf and Walsh (2013); Depro, Timmins and O'Neil (2015) show that the impacts of selective migration or sorting processes on aggregated income dynamics are theoretically hard to predict and empirically difficult to identify when relying on aggregated data. To overcome these problems, the current book employs both strategies. It uses household level survey data to investigate selective migration patterns, but aggregated data to investigate the level of environmental inequality in Germany.

However, using spatially aggregated data bears some potential problems that need to be accounted for in the analyses. One issue is the modifiable areal unit problem (MAUP; see e.g. Wong, 2009), actually consisting of two distinct problems. First, conclusions may depend on the level of aggregation (e.g. Baden et al., 2007; Downey, 1998). For example, we can analyse environmental inequality on the city level and conclude that minorities live in more polluted cities. However, this does not necessarily mean that minorities also live in more polluted districts within the cities. Thus, conclusions could vary when conducting similar analyses on the city and the city-district level. In general, a higher resolution increases the variance between units and decrease the (unobserved) variance within units, thereby increasing the accuracy of predictions (assuming that there is no theoretical reason to expect differing results for different levels). Second, the conclusions may also depend on the borders of the spatial units. Accordingly, we need to ask whether conclusions would change if we shift given borders by an arbitrary distance to the left or the right. In environmental inequality research, it seems plausible to assume that hazardous facilities are systematically sited at the borders of administrative spatial units (Mohai & Saha, 2006, 2007). If we allocate the pollution to the facility-hosting spatial unit only, selection of geographic borders may crucially influence the results. Furthermore, facilities located at the borders of a spatial unit can actually affect neighbouring units more than the hosting unit. To overcome this problem, Mohai and Saha (2006, 2007) propose to use a so called 'areal apportionment method', which weights the characteristics of nearby units by the spatial intersection between census tracts and a circle around the facility location (for a similar method see also Banzhaf & Walsh, 2008). The present books deals with these issues by 1) using very fine-grained spatial units to minimise the unobserved variance within units, and by 2) applying a matching algorithm between facilities and census units that follows Banzhaf and Walsh (2008), merging the emissions to census units weighted by the intersection between the census units and a circular buffer around the facility location.

A second issue of spatially aggregated data relates to the use of appropriate statistical methods. As has been formulated by Tobler (1970, p. 236): 'everything is related to everything else, but near things are more related than distant things'. It seems quite intuitive that geographically close units (whether aggregated or individual) are more similar to each other, more likely to influence each other, and also more likely to be affected by common shocks. Though this does not seem surprising, it has important implications for the statistical methods applied to analyse spatial data. As the units of observation cannot be handled as if they were independent of each other, a core assumption of 'standard' linear regression models ($E(\varepsilon_i \varepsilon_j) = E(\varepsilon_i)E(\varepsilon_j) = 0$, where ε is the regression disturbance for all observations i and j, $i \neq j$) is violated. As has been shown by previous research (e.g. Pace & LeSage, 2010), this can lead to erroneous inferential statistics, but also to biased point estimates when using non-spatial regression methods. Thus, appropriate spatial methods need to be applied when analysing spatial data (e.g. Anselin, 1988; Cliff & Ord, 1981; Elhorst, 2014; LeSage & Pace, 2009). Surprisingly, this issue has been widely ignored in environmental inequality research (exceptions are Elliott & Frickel, 2013, 2015; Havard et al., 2009; Pastor et al., 2005; Raddatz & Mennis, 2013), and studies using spatial regression techniques predominantly apply the most commonly used spatial autoregressive model (e.g. Elliott & Frickel, 2015; Havard et al., 2009; Raddatz & Mennis, 2013) to control for spatial dependence. However, a variety of spatial model specifications exist, which not only control for spatial dependence but also explicitly model theoretically relevant spatial patterns. Hence, the current book 1) summarises the most popular spatial regression techniques, and 2) performs several Monte Carlo experiments to compare the performance of spatial model specifications throughout different situations. This provides a theoretical and empirical foundation for the subsequent use of spatial regression models in the analysis of environmental inequality patterns in Germany.

1.5 Overview

This book is divided into four empirical studies and a final conclusion. Chapters 2, 4, and 5 deal with the research questions formulated in this chapter. Chapter 2 employs the 'standard' methodological approach and uses household-level data, while Chapters 4 and 5 use spatially aggregated data. Chapter 3, in contrast, deals with methodological issues when using spatially aggregated data. Though the four research studies can be seen as cumulative research and build on each other, all chapters provide the necessary theoretical background and can be read independently. This section provides a short overview of each study.

Study 1 (Chapter 2) scrutinises the main research questions of this book by analysing household level survey data of the German Socio-economic Panel (SOEP). In the beginning, it replicates previous findings (Kohlhuber et al., 2006) by showing that, in a cross-sectional design, minorities and socio-economically disadvantaged households report a higher impairment through air pollution. However, the main focus of this study lies on the causal mechanisms, especially on the process of selective migration and its influence on the perceived exposure to air pollution in Germany. As has been argued earlier, the theory of selective migration assumes that minorities selectively move into more polluted areas.

To test this argument, the study applies fixed-effects panel estimators and investigates the moving returns of households between 1986 and 2014. Specifically, we test two main questions. First, whether increases in income lead to higher moving returns, i.e. a stronger reduction in exposure to environmental pollution, and second, whether minority households experience different moving returns than native German households. In addition, the study analyses the more fine-grained mechanisms by testing if income explains the lower moving returns of minorities and by differentiating the moving returns of first- and second-generation immigrants. This provides the first empirical test of selective migration mechanisms outside of the United States that uses longitudinal household data.

Study 2 (Chapter 3) is more concerned with a methodological topic than with environmental inequality per se. In a first step, the study summarises the most popular spatial model specifications and shows theoretically under which conditions non-spatial regression methods produce biased estimates. This gives a strong motivation for the use of spatial regression models when analysing spatially aggregated data, a finding which becomes important in the following chapters. However, a main difficulty in applied spatial econometrics is the selection of the correct model specification, as empirical specification tests exhibit severe drawbacks.

In a second step, the study thus compares the performance of the most common spatial regression models in different scenarios of spatial dependence by using Monte Carlo experiments. The study extends previous simulation results by evaluating the bias of the impacts rather than the regression coefficients, as direct and indirect impacts are the measures of interest in applied spatial econometrics. Furthermore, this study provides results for situations in which the regressions suffer from an omitted variable bias. By doing so, Chapter 3 constitutes a comprehensive introduction into spatial regression methods used in the following chapter, but also contributes to the discussion of model selection in spatial econometrics. The results clearly indicate that some specifications are much more robust against misspecification than others.

Study 3 (Chapter 4) again deals with the topic of environmental inequality. In contrast to study 1, study 3 uses spatially aggregated data and objective measures of air pollution. Therefore, I combine spatial data of the 2011 German census and pollution measures of the European Pollutant Release and Transfer Register (E-PRTR). By using geo-locations of industrial facilities and geographic boundaries of the census units, I calculate the proportionate toxicity-weighted amount of air pollution for each census cell. In the end, the study relies on more than 90,000 census cells over Germany.

Based on the results of study 2 (Chapter 3), this study uses SLX and community-fixed SLX models to incorporate spatial spillover-effects into the analysis and to account for the spatial distribution of socio-demographic characteristics. Specifically, the study is concerned with three main questions. First, the presence and extent of environmental inequality in Germany when using objective measures of industrial air pollution. Second, the spatial patterns of environmental inequality. Theoretically, I argue that both causal mechanisms of environmental inequality predict a clustering effect of minorities around hazardous facilities, a hypothesis that can be tested by applying spatial regression models. Third, the paper investigates whether the correlation between the minority share and toxic air pollution differs significantly between urban and rural areas. Thus, study 3 gives a first nation-wide assessment of environmental inequality and its spatial pattern with objective indicators of environmental pollution in Germany.

Study 4 (Chapter 5) constitutes a follow-up study of study 3. While study 3 shows that minorities, on average, bear a disproportionate exposure to environmental pollution, study 4 takes a more detailed look at the variations of environmental inequality between German cities and analyses which structural conditions foster environmental inequality. This study uses similar data as in study 3, but enriches the data by additional city-specific characteristics from the 'Indicators, Maps and Graphics on Spatial and Urban Monitoring' (INKAR) database as well as infrastructural characteristics derived from OpenStreetMap.

The standard strand of theoretical reasoning in environmental inequality research as well as the findings of the previous study in this book suggest that residential segregation constitutes a main driver of environmental inequality. However, other studies also point to the importance of the geographic centrality of pollution in shaping high levels of environmental inequality. To test the influence of macro-structural characteristics, the study applies city-fixed effects multilevel models. Thereby, I identify a structural characteristic of the urban form, which has been mostly overlooked by previous research. Furthermore, the findings of this last study provide a different interpretation for the results of the preceding studies, and emphasise the importance of taking the urban context into account when analysing environmental inequality.

The book ends with a *final discussion and conclusion* (Chapter 6). The last chapter summarises the main findings of the four empirical studies, and develops an overarching conclusion. Thereby, it highlights how the findings of the empirical studies help to understand the main research questions of whether environmental inequality exists in Germany, and which causal mechanisms are responsible for the disproportionate burden of minorities. Furthermore, it discusses the limitations of this book and develops some suggestions for further research, which can help to bolster our understanding of the processes of environmental inequality.

Chapter 2

How Selective Migration Shapes Environmental Inequality in Germany: Evidence from Micro-Level Panel Data

Abstract

Socio-economically disadvantaged and ethnic minorities are affected by a disproportionately high exposure to environmental pollution. Yet, it is unclear if selective migration causes this disproportionate exposure experienced by low-income and minority households. The study uses longitudinal data from the German Socio-Economic Panel to investigate the process of selective migration and its connection to the perceived exposure to air pollution in Germany. Consistent with the selective migration argument, movers experience a decrease in exposure according to their income, while stationary households do not experience a reductive effect due to income. Furthermore, the moving returns differ by minority status. While native German households experience less exposure to pollution when moving to a new place of residence, minority households do not. Additional analyses show that this minority effect cannot be explained by socio-economic differences, but completely vanishes in the second immigrant generation.

This chapter has been published as:

Best, H., & Rüttenauer, T. (2018). How selective migration shapes environmental inequality in Germany: Evidence from micro-level panel data. *European Sociological Review, 34*(1), 52-63. DOI: 10.1093/esr/jcx082

2.1 Introduction

Environmental inequality, connecting the distribution of environmental hazards to socio-economic and ethnic characteristics, has received growing attention in the United States (for an overview: Mohai & Saha, 2015a; Pellow & Nyseth Brehm, 2013; Ringquist, 2005), as well as in continental Europe (Diekmann & Meyer, 2010; Funderburg & Laurian, 2015; Havard et al., 2009; Kohlhuber et al., 2006; Laurian & Funderburg, 2014; Padilla et al., 2014). Most of the empirical research focuses on the question whether there is an unequal distribution of environmental pollution; yet only a few studies explore the causal mechanisms. To our knowledge, only one study investigates these causal mechanisms of environmental inequality in continental Europe (Funderburg & Laurian, 2015), exclusively using aggregated data. This is a major shortcoming, as the United States is a rather special case with a high level of economic inequality (Piketty & Saez, 2014) and residential segregation (Musterd, 2005). Therefore, it is far from clear whether we can observe similar drivers of environmental inequality in Europe.

Previous studies, especially from the United States, have identified two causal mechanisms of environmental inequality: selective siting and selective migration (Mohai & Saha, 2015a). The first mechanism states that the increase of pollution follows already existing differences in the socio-demographic composition of neighbourhoods. Hazardous facilities are disproportionately sited in areas with low socio-economic resources and high minority shares (or disproportionately cleaned up in areas with high socio-economic resources and low minority shares). In contrast, the second mechanism assumes that differences in the socio-demographic neighbourhood composition emerge after pre-existing differences in pollution. This means that minority households and households with low socio-economic resources selectively move into polluted areas, while socio-economically advantaged households move out. Previous studies have investigated these two causal explanations on aggregated levels, such as neighbourhoods, zip code areas, or census tracts (Been & Gupta, 1997; Downey, 2005; Funderburg & Laurian, 2015; Mohai & Saha, 2015b; Oakes et al., 1996; Pastor et al., 2001; Richardson, Shorty & Mitchell, 2010; Shaikh & Loomis, 1999). Those macro-level studies offer mixed results regarding selective siting and provide only weak evidence regarding the selective migration mechanisms. However, the only two longitudinal studies on the household level find evidence for selective migration patterns (Crowder & Downey, 2010; Pais et al., 2014). Thus, it remains unclear whether selective migration causes the disproportionate exposure to environmental pollution experienced by low-income and minority households.

The present study adds to the discussion of environmental inequality in two ways. First, the present study is the first panel study on environmental inequality in Germany. Hence, we add new evidence to the question, whether selective migration causes environmental inequality in continental Europe. We argue that the exposure to environmental pollution depends on income and ethnic origin via individual moving decisions (selective migration). Secondly, we compare different immigrant groups (first- and second-generation immigrants and by country of origin) and analyse whether the disadvantage of ethnic minorities can be explained by their lower socio-economic status.

This study relies on household-level data of self-reported impairment through air pollution of the German Socio-Economic Panel (GSOEP) between 1986 and 2014. By using panel data and fixed-effects estimators, we ensure that our results are not affected by differing perception of pollution between the households. Though the micro-level data used in this study offer clear advantages when analysing selective migration patterns, they do not allow to investigate selective siting, as the latter depends on aggregated neighbourhood characteristics rather than individual household characteristics.

2.2 Theory and previous results

Numerous studies in the United States have shown that income and race are related to the amount of environmental pollution (for an overview: Mohai & Saha, 2015a; Pellow & Nyseth Brehm, 2013; Ringquist, 2005). In the German-speaking area, previous studies conclude that low-income households as well as ethnic minorities experience a higher exposure to environmental pollution (Bolte & Mielck, 2004; Diekmann & Meyer, 2010; Kabisch & Haase, 2014; Kohlhuber et al., 2006; Raddatz & Mennis, 2013).[1] Regardless, all these studies did not aim to analyse the causal mechanisms of environmental inequality. Therefore, the following outline of the causal mechanisms mostly relies on literature from the United States.

The selective migration argument assumes that the patterns of environmental inequality result from specific decisions on the individual or household level (Massey, 1990; Schelling, 1978). On the one hand, these residential choices are driven by individual preferences, e.g. the preference to live in a clean and unpolluted environment. On the other hand, individuals have to deal with the structural constraints of their actions. If we assume a similar preference for clean environment throughout the society — an assumption that our analytical strategy allows to relax — market or other allocation mechanisms regulate the access to scarce resources. When selecting a place of residence, individuals try to satisfy their preferences regarding the good ‘clean environment’, given their economic and structural constraints (Tiebout, 1956). Thus, we have to explain how these constraints differ by income and ethnicity to understand

[1] Supplementary cross-sectional analyses of our data confirm these results (see supplementary Table 2.3).

the causes of environmental inequality. First, we will outline how income affects the migration process and afterwards turn to the ethnic differences in migration patterns.

2.2.1 Income as the key to clean neighbourhoods

The 'market explanation' of environmental inequality considers a clean environment as an economic luxury good, which is available on the market for an additional price. Environmental quality influences the rents and housing prices: While housing opportunities in low-quality areas are relatively cheap, they are relatively costly in high-quality areas. As a result, tenants and homebuyers must pay for high environmental quality when making a migration decision (Banzhaf & McCormick, 2012; Been & Gupta, 1997; Diekmann & Meyer, 2010; Hanna, 2007; Hunter et al., 2003). Given the preference for a clean environment, households experiencing an increase in income are able to pay more for environmental quality when migrating and consequently will end up in neighbourhoods with lower pollution. Even if housing prices are not exogenously higher in clean neighbourhoods, prices should rise due to higher demand for housing in these areas, despite relatively constant housing opportunities (Banzhaf & McCormick, 2012; Kim, Campbell & Eckerd, 2014). On the other hand, higher environmental pollution (combined with an out-migration of the wealthier inhabitants and decreasing demand for housing) leads to decreasing housing prices in polluted areas and consequently attracting lower-income households who cannot afford the luxury good 'clean environment'. As a result, lower-income households will choose neighbourhoods with higher pollution when moving.

Following this market explanation, high pollution does not need to be the reason for out-migration. Even if households relocated for other reasons, high-income households sort into clean neighbourhoods, while low-income households sort into polluted neighbourhoods (Banzhaf & McCormick, 2012). Though pollution may trigger out-migration, all migrating households — independent of the reasons for relocation — need to choose a neighbourhood destination, which is assumed to happen selectively. Therefore, we postulate the following Hypothesis 1:

H1: An increase in income will lead to decreasing exposure to environmental pollution when moving to a new neighbourhood.

This is a causal formulation of the between hypothesis, that is high-income households will on average realise a stronger reduction in exposure. In line with this hypothesis, the majority of previous research has found a negative correlation between income and environmental pollution within the same spatial area (Ash & Fetter, 2004; Downey, 2006a; Downey & Hawkins, 2008; Mohai & Saha, 2007; Pastor et al., 2002), which supports the 'market explanation'. Despite this, other studies suggest that there is no relationship between income and environmental pollution (Been & Gupta, 1997; Morello-Frosch & Jesdale, 2005) or at least no linear relationship (Ash, Boyce, Chang & Scharber, 2013; Havard et al., 2009). Aside from these macro-level studies, only two studies have presented results from longitudinal analyses on the micro-level. Crowder and Downey (2010) used household-level data from the Panel of Income Dynamics and pollution data of the Toxics Release Inventory (TRI). They state that although pollution does not increase the probability of out-migration when controlling for other individual characteristics, income is associated with lower pollution in the neighbourhood of destination for movers. Thus, they find a conditional effect of income on pollution. If households move, higher-income households sort into neighbourhoods with a lower amount of pollution. Using similar data, Pais et al. (2014) compare different migration trajectories separated by the exposure to pollution. They find that the probability of being in a constantly high pollution trajectory compared to being in a constantly low pollution trajectory decreases with income. This indicates that households with higher income are more likely to continuously live in low pollution neighbourhoods.

2.2.2 Minority status as a barrier to clean neighbourhoods

Turning to the disadvantages of ethnic minorities, two explanations have been widely discussed by previous scholars: the 'racial income-inequality thesis' and the 'racial residential discrimination thesis'.[2]

The 'racial income-inequality thesis' relates to the 'market explanation' in explaining the effect of ethnicity on the exposure to environmental pollution (Been & Gupta, 1997; Campbell, Peck & Tschudi, 2010; Crowder & Downey, 2010; Oakes et al., 1996; Pais et al., 2014). Hereafter, the high exposure of ethnic minority groups is not a result of ethnicity itself, but rather, a result of the differences in the socio-economic resources of different ethnic groups. The hypothesis assumes that ethnic minority groups hold a relatively low income compared to the ethnic majority. Following the 'market explanation', minority groups are limited by their economic resources and cannot afford the housing prices in high-quality neighbourhoods. Thus, they are pushed to more affordable but more polluted areas. The ethnic majority, in contrast, holds a relatively high income and faces lower economic constrains when choosing a neighbourhood of destination. It follows that majority households sort into high-quality neighbourhoods, while minority households sort into low-quality neighbourhoods. If

[2] The terms 'race' and 'ethnicity' have the same meaning throughout this article. US scholars typically study inequalities based on race, while European researchers typically focus on inequalities along ethnic lines. For our purposes, both concepts refer to a fixed actor attribute that can be the basis for different housing market opportunities.

this is true, a higher exposure of minority households should dissipate after controlling for socio-economic resources.

In contrast, the second mechanism assumes a persisting effect of ethnicity on the exposure to environmental hazards, independent of socio-economic characteristics. The 'racial residential discrimination thesis' explains the unequal distribution of environmental pollution by discriminating actions of real estate agents or property owners (Crowder & Downey, 2010; Crowder, Pais & South, 2012; Pais et al., 2014; for discrimination on the housing market in general: Choi, Ondrich & Yinger, 2005; Ondrich et al., 2003; Pager & Shepherd, 2008; Turner & Ross, 2005). The reasons for housing discrimination could be twofold: first, native inhabitants could perceive minority groups as a threat due to prejudices about their criminal behaviour (Massey & Denton, 1993; Semyonov et al., 2012). Housing agents and property owners, in turn, fear declining desirability of neighbourhoods and, as a result, declining profits due to minority in-migration (Turner & Ross, 2005; Yinger, 1986). Thus, they prefer majority households as new inhabitants in high-quality neighbourhoods. Secondly, housing agents could spuriously anticipate the housing preferences of minority groups (Ondrich et al., 2003; Turner & Ross, 2005). If housing agents supposed minority groups have lower preferences for clean environments or neighbourhood quality in general, they could pre-select housing offers based on their prejudiced viewpoint. In both cases, discriminating behaviour sorts minority households into polluted neighbourhoods.

The theoretical explanations as well as the empirical results regarding the disproportionate exposure of minority groups stem mainly from the United States and refer to ethnic minorities like Asian, Mexican, or African-American groups. It is important to note that minorities in Germany are the result of relatively recent immigration, mainly from Turkey, Southern Europe, and later Ex-Yugoslavia and Eastern Europe (see Kalter & Granato, 2007, and the more detailed discussion of our data in the following section). Additionally, housing segregation by ethnic group is lower in Germany than in the United States (Musterd, 2005). Nonetheless, we assume that the mechanisms discussed above are transferable to the German context, as previous research has found general disadvantages of immigrants on the German housing market (Auspurg, Hinz & Schmid, 2017; Drever & Clark, 2002), plus a higher exposure of immigrant minorities to pollution (Diekmann & Meyer, 2010 for Switzerland; Kohlhuber et al., 2006; Raddatz & Mennis, 2013 for Germany). In sum, this leads to Hypothesis 2:

H2: Native Germans will realise a stronger reduction in the exposure to environmental pollution than immigrant minority households when moving to a new neighbourhood.

Based on the discussion of the fine-grained mechanisms of selective migration, we will extend our analysis of selective migration in two ways. First, we will test whether the differences between ethnic groups can be explained by socio-economic differences (H2a: 'racial income-inequality hypothesis'). If this explanation is true, the disadvantage of minorities should disappear or at least diminish when controlling for income. Secondly, we will separate the minority group into first- and second-generation immigrants. If discrimination causes the disproportionate burden of minority households (H2b: 'racial residential discrimination thesis'), we would assume second-generation immigrants to experience similar disadvantages as first-generation immigrants.

The disproportionate exposure of ethnic minorities to environmental pollution is well documented in previous research. Even after controlling for income, many studies document a significant correlation between ethnic minority share and environmental pollution within the same neighbourhood (Ash & Fetter, 2004; Downey, 2006a; Downey & Hawkins, 2008; Pastor et al., 2005). In addition, housing audit studies highlight the persisting discrimination against minorities on the housing market (Choi et al., 2005; Ondrich et al., 2003; Turner & Ross, 2005). This income-independent effect of ethnicity is in line with the 'racial residential discrimination thesis'. Yet, as most of the environmental inequality studies use cross-sectional data, they cannot — and do not aim to — identify the causal mechanisms leading to the disproportionate exposure of minority groups to environmental pollution. Nonetheless, some other studies (Been & Gupta, 1997; Downey, 2005; Mohai & Saha, 2015b; Oakes et al., 1996; Pastor et al., 2001; Richardson et al., 2010; Shaikh & Loomis, 1999) focus on the causal mechanisms by analysing longitudinal data on the aggregate level. Of these studies, only Richardson et al. (2010) find empirical evidence for the selective migration into polluted areas. All other macro studies rather contradict the selective migration argument.

In contrast, the two longitudinal studies on the household level find evidence for selective migration of ethnic groups. According to Crowder and Downey (2010), ethnic minorities are more likely to move into polluted areas when comparing mobile households. Even when controlling for socio-economic characteristics, the effect of ethnicity persists. Additionally, the analysis shows that income is a greater determining factor for Black than for White homeseekers in dictating the pollution at the place of destination. This finding is in line with the explanation of discriminating housing markets, as mobile minority households require a higher income to realise the same level of pollution than their majority counterparts do. The second household-level study by Pais et al. (2014) comes to similar conclusions. Although the race effect is lowered in magnitude when controlling for socio-economic factors, the odds of following a high-to-high pollution trajectory are still substantially higher for Black than for White households. These results support both the 'racial income-inequality thesis' and the 'racial residential discrimination thesis'. Furthermore, the number of inter-neighbourhood moves reduces the probability of being in a constantly high-pollution trajectory for White households, while having a contradicting effect for Black households.

2.3 Data, operationalisation, and method

We use household-level data from the GSOEP, a repeated and representative panel study in Germany (Wagner, Frick & Schupp, 2007). Our final sample comprises information on 12,037 households, participating between 1986 and 2014. This sample includes all households living in private households (households living in hostels or retirement homes were excluded) participating at least in two of the six relevant waves (unbalanced panel). All models were estimated without using sampling weights.[3]

As dependent variable we use the subjectively perceived impairment through air pollution. In 1986, 1994, 1999, 2004, 2009, and 2014, the GSOEP household questionnaire included the question 'How strongly do you feel you are affected by the followed environmental influences on your residential area: Through air pollution'. Respondents had to evaluate their answer on a five-point scale ranging from 'Not at all' to 'Very strongly' (coded 1–5).

The main explanatory variables are income and minority status. As we argue that the latter variables are crucial for the extent of environmental exposure only when people move, we include an interaction-term of these main explanatory variables with a dummy indicating a relocation of the household between two points of observation. This ensures that the report of perceived pollution temporally follows the relocation, and we do not run into trouble concerning the causality. As a measure of income, we calculated the equivalised household income by dividing the monthly net income by the square root of the household size (Burniaux et al., 1998).

In the German context it makes sense to specify minority as immigrant minority: the vast share of minorities — especially of minorities visible by their looks or their language — are immigrants and their descendants (Kalter & Granato, 2007). Most of the first-generation immigrants in our sample originate from countries covered by Germany's active labour immigration policy of the 1960's like Turkey (n = 348), Yugoslavia (n = 199), Italy (n = 171), Greece (n = 110), or Spain (n = 66). The second important group of first-generation immigrants originates from Eastern Europe (n = 566). In total, the sample contains 1,490 first- and 699 second-generation immigrants.[4] Third-generation immigrants are categorised as native Germans in the original data set. Due to the low number of cases, we need to collapse some of the origins and differentiate between Turkey, former Yugoslavia, and the following regions as classified by the United Nations Statistics Division: Southern Europe, Eastern Europe, rich Western Countries

[3] As panel attrition may be an issue for the present research question, we estimated additional models using inverse staying probability weights (not shown). These models yield very similar results (see Solon, Haider & Wooldridge, 2015, for a discussion on weighting).

[4] To separate the effect due to relocation from the effect of other events, we had to exclude 198 households (approximately 1.6% of the total sample) that experienced a change in minority status due to a change of the household head (32 additional households had to be excluded when differentiating between first and second generation of immigration).

(Northern and Western Europe, the United States, Canada, Australia) and a residual category for other countries (see Gresch & Kristen, 2011, for a discussion of different operationalisations).[5]

Finally, the relocation of the household is measured by three period dummies indicating the first, second, or third move of a household. Hence, the coefficient of the second move-dummy indicates the additional effect of a second move over the first, etc. For construction, all GSOEP waves between 1986 and 2014 were taken into account to capture relocations between the waves used in the final analyses.

Additionally, we include several variables controlling for confounders. Most importantly, we exclude the possibility of a perception bias and, consequently, a spurious decrease in the perception of environmental pollution after a relocation (e.g. due to cognitive-dissonance reduction). Hence, we include a control variable for households that have been living in their new home for less than 6 months. We also include a dummy that captures the change of the household head as perception of pollution might differ between the former and the new household head. In addition, several control variables may influence the pollution level as well as income and thus confound the effect of income on pollution: we include age squared, since income and living situation might notably improve in early adulthood (the linear age term is omitted because it cannot be separated from the year trend in two-way fixed-effects models) but stagnate or decrease in old age. In the same line, we control for children living in the household, considering children influence the household income (by definition) and might influence the residential choice when households move.[6] Since we are interested in the total effect, we generally do not control for mediators. An important exception is controlling for income when testing the 'racial income-inequality hypothesis'.

To identify the causal mechanisms of environmental inequality, we use fixed-effects panel estimators. These estimators use the variance within the household over time while excluding the variance between the households (Allison, 2009; Brüderl & Ludwig, 2015). Time-constant household-specific characteristics are excluded and no longer affect the estimation results. Furthermore, the inclusion of year-dummies (two-way fixed-effects) ensures that people without relevant within-variance — households without any relocation — serve as a control group. This controls for a change in pollution level over time independent of our explanatory variables. For ease of interpretation, we use linear fixed-effects estimators, though the dependent variable is measured on an ordinal scale. To ensure that this model selection does not influence our results, we conducted

[5] Our operationalisation of minority status follows the status of the GSOEP household head. This decision ignores the individual migration histories of all other household members. However, our approach avoids questionable exclusions (of mixed native and immigrant households) and leads to more conservative estimates, i.e. a bias towards a non-finding.

[6] Note that we do not include control variables that might cause moving decisions. This would be important if we investigated the question whether pollution induces relocations. However, for our dependent variable the reasons for the moving decision do not confound the results.

Table 2.1: Summary statistics

Variable	Native German household Mean	Native German household SD (within)	Minority household Mean	Minority household SD (within)
% moved at least once	49.04		63.76	
Perceived impairment through air pollution	1.825	0.587	1.923	0.639
% experienced an improvement	48.62		49.62	
Monthly equivalence income (in 1,000 Euros)	1.614	0.527	1.367	0.426
Age of household head	52.683	5.928	48.963	6.088
Children living in household	0.263	0.273	0.379	0.296
Number of households	9816		2221	
N	31267		7118	

additional sensitivity checks using an additive index of air and noise pollution[7] as well as fixed-effects ordered-logit models (see supplementary Tables 2.4 and 2.5). Both models support the results of the linear fixed-effects estimators of air pollution.

2.4 Results

Table 2.1 presents descriptive statistics by the household's minority status. First, the data show that most households report a low perceived impairment through air pollution (mean of 1.8). Secondly, the data indicate a higher mobility level for minority households. Approximately 50% of the households with a native German household head moved at least once within our observation period compared to 64% of the minority households. Furthermore, minority households perceive a higher average level of impairment through air pollution. However, in both groups nearly 50% of households experienced at least one improvement in air quality over time. In addition, minority households report on average a 250 EUR lower household income, which indicates that the higher pollution experienced by minority households might be a function of socio-economic status ('racial income-inequality hypothesis').

To draw conclusions about the mechanisms leading to environmental inequality, Table 2.2 presents the results of the panel regressions. All models are two-way fixed-effects models with cluster robust standard errors. We begin with a discussion of income inequality and then turn to the minority effect.

Model FE 1 includes household income, moving behaviour, and their interaction as explanatory variables. The important construct for testing the effects of selective migration is the interaction of moving behaviour and income which separates the effect of movers and non-movers. In line with our H1, FE 1 shows a significant reduction of the perceived pollution due to income for movers only. This confirms the 'market explanation'. When moving, households experience a higher reduction of perceived

[7] The subjective impairment by ambient noise was measured using a question very similar to air pollution (see above).

Table 2.2: Fixed-effects estimation of perceived impairment through air pollution

	FE 1	FE 2	FE 3	FE 4
Monthly equivalence income (in 1,000 Euros)	−0.008		−0.009	−0.009
	(0.009)		(0.009)	(0.009)
1. Move	−0.203***	−0.282***	−0.247***	−0.249***
	(0.030)	(0.022)	(0.031)	(0.032)
2. Move	−0.034	−0.080**	−0.039	−0.045
	(0.039)	(0.025)	(0.042)	(0.042)
3. Move	−0.014	−0.015	−0.023	−0.034
	(0.086)	(0.041)	(0.087)	(0.088)
1. Move × income	−0.027*		−0.022	−0.021
	(0.013)		(0.013)	(0.013)
2. Move × income	−0.019		−0.020	−0.018
	(0.020)		(0.020)	(0.020)
3. Move × income	0.009		0.009	0.014
	(0.044)		(0.044)	(0.045)
1. Move × minority		0.165***	0.159***	
		(0.044)	(0.044)	
2. Move × minority		0.023	0.015	
		(0.049)	(0.049)	
3. Move × minority		0.047	0.048	
		(0.097)	(0.096)	
1. Move × minority (1st-generation immigrant)				0.226***
				(0.051)
2. Move × minority (1st-generation immigrant)				0.083
				(0.060)
3. Move × minority (1st-generation immigrant)				0.141
				(0.135)
1. Move × minority (2nd-generation immigrant)				−0.005
				(0.078)
2. Move × minority (2nd-generation immigrant)				−0.047
				(0.075)
3. Move × minority (2nd-generation immigrant)				−0.026
				(0.129)
R^2	0.056	0.057	0.057	0.058
Adj. R^2	0.056	0.056	0.057	0.057
AIC	67185	67166	67153	66880
Number of households	12037	12037	12037	12005
N	38385	38385	38385	38268

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Cluster robust standard errors in parentheses. Additional controls: Age^2, children living in household, change of household head, duration of residence, year.

air pollution as their income rises. Figure 2.1 shows how the reduction of pollution due to moving increases with income. In contrast, the effect of income for stationary households is close to zero and not significant: only when moving, households can use their income to reduce their exposure to pollution. Note, however, that the effect is statistically significant but low in magnitude. An increase in income by 1,000 Euro increases the effect of mobility on impairment through air pollution by only 0.03 points (which equals 0.05 standard deviations). This income interaction effect equals three times the presented effect if we interact moving behaviour with the household's average income over time, which would compare the moving returns between 'rich' and 'poor'

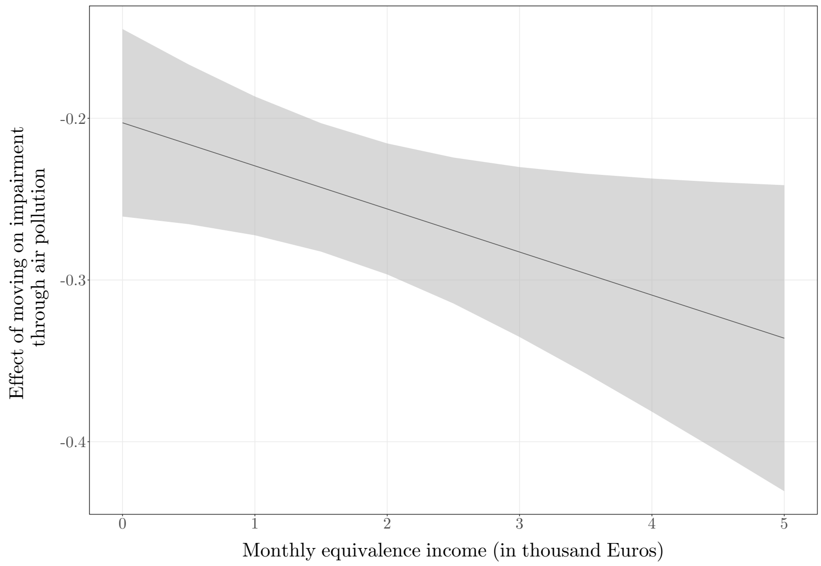

Figure 2.1: Conditional effect of the first move on the impairment through pollution with 95% confidence interval (FE 2)

households (see supplementary Table 2.6).[8] Independent of model choice, our results confirm selective migration as a causal mechanism of environmental inequality.

Turning to the minority effect, model FE 2 includes the main moving effects and an interaction between moving behaviour and minority status. While the interaction terms in combination with the main effects represent the effect of migration for minority households, the main moving effects represent migration returns for native German households. In consonance with H2, we find a significant and negative effect of migration for native German households: the first observed move reduces the perceived impairment through air pollution by nearly 0.28 points on a scale from 1 to 5 (which equals approximately 0.47 within standard deviations). In contrast, minority members experience a much lower reduction in pollution when moving (reduction by approximately 0.12 points or 0.20 within standard deviations). This is a substantial difference and confirms our H2: the improvement due to mobility is more than twice as strong for native German households than for minority households. Figure 2.2 depicts the results graphically. Even the effect of the second move exceeds the 1% significance level for native German households and points to an average additional improvement due to a second move, while it is non-significant for minority households. These results confirm selective migration as a cause of the disproportionate exposure to air pollution of minority households (H2).

Model FE 3 tests the 'racial income-inequality hypothesis' and includes income as well as minority status with its moving interactions. If the effect of the household's

[8] We argue that comparing different income levels within the same household over time is the adequate strategy to model the effect, as it uses within-variance only. The strategy may produce conservative point estimates but is less prone to biases induced by unobservable confounders.

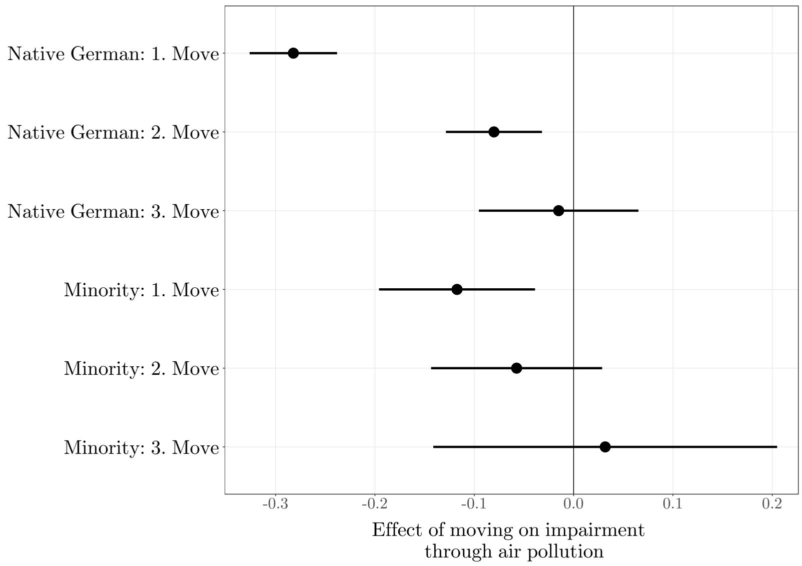

Figure 2.2: The effect of moving for native German and immigrant minority households on the impairment through pollution with a 95% confidence interval (FE 2)

minority status was a result of lower socio-economic status, the minority effect should disappear or diminish under control of income. Still, the difference in reduction of air pollution between native German and minority movers remains nearly unaffected when controlling for income (comparing FE 2 and FE 3). Independent of income, native German households can experience a much higher reduction in air pollution than their minority counterparts do. In contrast to the 'racial income-inequality hypothesis', we do not find any significant effect of household income and only a slight decrease of the minority effect in model FE 3. As in previous research, we observe a persistently lower effect of mobility on the exposure to pollution for minority households, even when controlling for income. Interestingly, the interaction between moving and income fails to reach the 5% significance level in model FE 3, which is in line with the results of Ringquist (2005) who concludes that the minority effect is much more robust than the income effect.

The last model (FE 4) separates the effect of minority status for first- and second-generation immigrants. The results show that the disadvantage of minority households identified in Models FE 2 and FE 3 completely stems from the disadvantage of first-generation immigrants. Figure 2.3 compares the effect of the first two moves for native Germans, separating first- and second-generation immigrants (under control of income). It turns out that the improvement of the pollution level from moving to a new place of residence experienced by second-generation immigrants is comparable to the improvements experienced by native Germans. First-generation immigrants, in contrast, do not experience any improvement in exposure to pollution when moving. Even under control for income, native Germans, as well as second-generation immigrants, sort into neighbourhoods with lower environmental pollution when moving, while first-

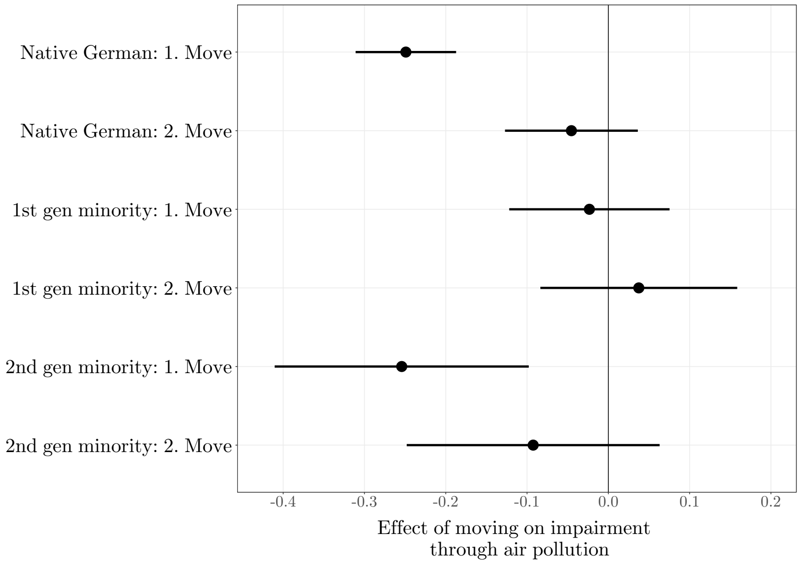

Figure 2.3: The effect of moving for native German, first-, and second-generation immigrant households on the impairment through pollution with a 95% confidence interval (FE 4, controlling for income)

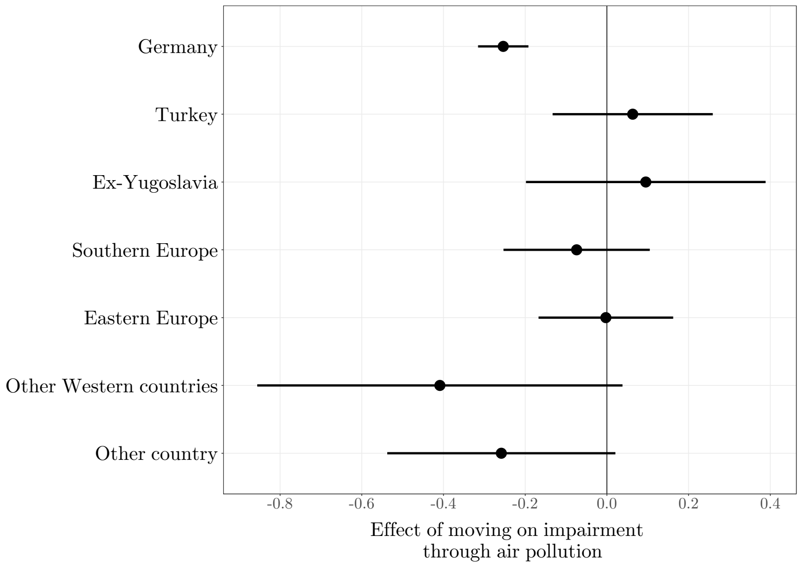

Figure 2.4: The effect of the first move for native German and first-generation immigrant households separated by their country of origin on the impairment through pollution with a 95% confidence interval

generation immigrants face the same exposure to pollution after moving to a new place of residence. Additional analyses of first-generation minority households (Figure 2.4, see supplementary Table 2.7) reveal that especially households originating from Turkey, Ex-Yugoslavia, and Eastern Europe are disadvantaged compared to native German households ($p \leq 0.05$). Households from those countries are not able to improve environmental quality due to mobility. Immigrants from wealthy Western countries, such as France, the United Kingdom, or the United States, in contrast, seem to experience the similar average improvement in environmental conditions due to moving as native Germans do, but the estimation of this effect is based on few cases only and statistically not significant.

2.5 Summary and conclusion

Environmental inequality has become an important topic of sociological research in the United States. This research has led to several important findings on the differential exposure of minorities to pollutants. Despite this, previous studies have led to inconsistent results regarding the causes of environmental inequality, and empirical studies in Europe are rare. In this article we present the first causal-analytic panel study of environmental inequality in Germany.

To investigate causal mechanisms, this article uses household-level panel data and fixed-effects estimators. We find that income has a significant impact on the level of perceived air pollution for movers, while it has no effect for stationary households: An increase in income leads to a significant reduction of air pollution when the household moves to a new place of residence. This confirms the existence of selective moving processes that shape environmental inequality. However, the effect of income is relatively low in magnitude and sensitive to model specification. When simultaneously including income and nationality, the effect of income loses significance. Regarding minority status, we find a significant improvement in air quality due to migration for native German households. Minority households, in contrast, only experience weak improvements. The disadvantages are especially strong for first-generation immigrants, who do not improve their situation at all when moving. This minority-difference in mobility patterns is robust against model specifications. Thus, selective migration behaviour operates as a causal mechanism, shaping the difference in pollution regarding minority status and, to a lower extent, income. These findings are consistent with previous results on the individual level from the United States (Crowder & Downey, 2010; Pais et al., 2014).

The fact that we find a moderate income-effect but a strong difference between native German and minority households is especially interesting in the German context. As German minorities are far less segregated than minority groups in the United States, we would have expected to observe lower minority differences. However, our data show that first-generation immigrant minorities in Germany are confronted with a high disadvantage when relocating and, thus, experience a higher exposure to pollution. This is true especially for first-generation immigrants from Turkey, former Yugoslavia, and Eastern Europe.

In summary, our analyses confirm selective migration as a causal mechanism of environmental inequality. Nevertheless, this study can only be a first step towards understanding the fine-grained mechanisms triggering selective migration. Our data allow to rule out some mechanisms prominently discussed in the literature: first, we do not find a noteworthy reduction of the minority effect when controlling for income, which indicates that minority disadvantages cannot be explained solely by their relatively low socio-economic status. This contradicts the ‘racial income-inequality hypothesis’. Secondly, the disadvantage of immigrants completely vanishes in the second generation, indicating that the disadvantage does not stem from discriminative behaviour triggered by simple ethnic markers like the look or the name of a person. This contradicts the ‘racial residential discrimination thesis’, but results are not fully conclusive, as discrimination could still occur based on other characteristics like language skills or citizenship. These characteristics are predominantly visible in the first immigrant generation and, thus, would lead to disadvantages in the first but not in the second generation.

At the same time, we cannot rule out that mechanisms other than income or discrimination play a role: first, successful integration of second-generation immigrants could also lead to informal network structures that support the search for high-quality housing. Following this argument, the disadvantage of first-generation immigrants would stem from their ethnic networks and their lack of sufficient ties to the mainstream. Secondly, homogeneity preferences might play an important role in residential choice (Kim et al., 2014), affecting predominantly first-generation immigrants. Thirdly, preferences for environmental quality might differ between Germans/second-generation immigrants on one side and first-generation immigrants on the other. Addressing these potential explanations in further research could produce important insights of the fine-grained mechanism of selective migration, hence bolstering our understanding of the causal mechanisms of environmental inequality.

A shortcoming of the present study is the subjectivity of the pollution measure. Though a strong correlation with noise pollution and proximity to the city centre indicates a connection to traffic-related air pollution, it remains a subjective measure. Thus, its use may be critical for cross-sectional analyses because of differing perceptions of pollution (as criticized by Diekmann & Meyer, 2010). Fixed-effects panel estimators, in contrast, do not rely on the consistency of perception between respondents, which strongly reduces the concerns brought forward against the use of perceived impairment. Nevertheless, further research should validate the results by combining objective and subjective measures of air pollution in a panel study (note that the cross-sectional study by Diekmann and Meyer (2010) found a correlation between objective and subjective pollution data).

Finally, it is important to note that we did not address the question ‘which came first?’ (Pastor et al., 2001): we did not study facility siting or clean-up behaviour, which was identified as another cause of environmental inequality by previous research, stating that facilities are sited disproportionately close to minorities. Our analysis offers strong evidence for selective migration as a causal mechanism but cannot claim selective migration as the exclusive or the most important mechanism. To complete the picture of causality producing environmental inequality, further research needs to combine micro panel data and longitudinal macro data to analyse selective migration as well as selective siting in one single study. This could also help to explain the fact that macro-level studies were not able to find evidence for selective migration processes, while micro-level studies do.

2.A Appendix Chapter 2

Table 2.3: Pooled OLS estimation of perceived impairment through air pollution

	POLS 1	POLS 2	POLS 3	POLS 4
Monthly equivalence income (in 1,000 Euros)	-0.065^{***}		-0.064^{***}	-0.064^{***}
	(0.008)		(0.008)	(0.008)
Minority		0.073^{***}	0.068^{***}	
		(0.017)	(0.017)	
Minority (1st-generation)				0.075^{***}
				(0.021)
Minority (2nd-generation)				0.057^{*}
				(0.027)
Female household	0.062^{***}	0.080^{***}	0.065^{***}	0.066^{***}
	(0.014)	(0.014)	(0.014)	(0.014)
Education^{a}				
(Ref: high)				
Other	0.708	0.829	0.722	0.719
	(0.444)	(0.445)	(0.443)	(0.444)
Drop out	0.002	0.026	-0.043	-0.048
	(0.048)	(0.049)	(0.050)	(0.050)
Low	-0.028	0.029	-0.032	-0.033
	(0.019)	(0.017)	(0.019)	(0.019)
Medium	-0.029	0.015	-0.026	-0.026
	(0.019)	(0.018)	(0.019)	(0.019)
Constant	1.719^{***}	1.601^{***}	1.690^{***}	1.695^{***}
	(0.099)	(0.099)	(0.099)	(0.099)
R^2	0.067	0.063	0.069	0.069
Adjusted R^2	0.066	0.062	0.068	0.068
AIC	24843	24898	24829	24826
N	37886	37886	37886	37881

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors in parentheses. Additional controls: Age, age^2, children living in household, year.

^{a} Categorised Casmin classification: Other (unspecified); Drop out (1a); Low (1b, 1c); Medium (2a, 2b, 2c_gen, 2c_voc), High (3a, 3b).

Table 2.4: Fixed-effects estimation of additive pollution index (perceived impairment through air pollution and perceived impairment through noise)

	FE 1	FE 2	FE 3	FE 4
Monthly equivalence income (in 1,000 Euros)	-0.001		-0.003	-0.003
	(0.015)		(0.015)	(0.015)
1. Move	-0.421^{***}	-0.573^{***}	-0.521^{***}	-0.525^{***}
	(0.056)	(0.043)	(0.060)	(0.060)
2. Move	-0.090	-0.118^{*}	-0.088	-0.099
	(0.074)	(0.047)	(0.078)	(0.079)
3. Move	0.017	-0.061	-0.022	-0.054
	(0.175)	(0.082)	(0.178)	(0.181)
1. Move × income	-0.044		-0.033	-0.031
	(0.025)		(0.025)	(0.025)
2. Move × income	-0.008		-0.012	-0.009
	(0.036)		(0.036)	(0.037)
3. Move × income	-0.019		-0.017	-0.002
	(0.089)		(0.089)	(0.090)
1. Move × minority		0.372^{***}	0.365^{***}	
		(0.085)	(0.085)	
2. Move × minority		-0.001	-0.007	
		(0.093)	(0.094)	
3. Move × minority		0.170	0.167	
		(0.189)	(0.189)	
1. Move × minority (1st-generation)				0.482^{***}
				(0.098)
2. Move × minority (1st-generation)				0.128
				(0.116)
3. Move × minority (1st-generation)				0.417
				(0.262)
1. Move × minority (2nd-generation)				0.067
				(0.145)
2. Move × minority (2nd-generation)				-0.142
				(0.139)
3. Move × minority (2nd-generation)				-0.039
				(0.255)
R^2	0.052	0.054	0.054	0.055
Adjusted R^2	0.052	0.053	0.053	0.054
AIC	113589	113536	113537	113091
Number of households	12037	12037	12037	12005
N	38371	38371	38371	38254

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Cluster robust standard errors in parentheses. Additional controls: Age^2, children living in household, change of household head, duration of residence, year.

Table 2.5: Fixed-effects ordered-logit (BUC) estimation of perceived impairment through air pollution

	FE 1	FE 2	FE 3	FE 4
Monthly equivalence income (in 1,000 Euros)	-0.015		-0.020	-0.018
	(0.038)		(0.038)	(0.038)
1. Move	-0.391^{***}	-0.729^{***}	-0.524^{***}	-0.524^{***}
	(0.093)	(0.063)	(0.098)	(0.099)
2. Move	-0.018	-0.171^{*}	-0.046	-0.074
	(0.126)	(0.073)	(0.134)	(0.134)
3. Move	-0.049	0.037	-0.055	-0.064
	(0.219)	(0.118)	(0.224)	(0.226)
1. Move × income	-0.154^{**}		-0.133^{**}	-0.134^{**}
	(0.050)		(0.050)	(0.050)
2. Move × income	-0.052		-0.054	-0.042
	(0.070)		(0.071)	(0.070)
3. Move × income	0.081		0.075	0.076
	(0.115)		(0.115)	(0.115)
1. Move × minority		0.465^{***}	0.434^{***}	
		(0.116)	(0.116)	
2. Move × minority		0.108	0.087	
		(0.134)	(0.134)	
3. Move × minority		0.074	0.075	
		(0.247)	(0.246)	
1. Move × minority (1st-generation)				0.577^{***}
				(0.135)
2. Move × minority (1st-generation)				0.280
				(0.158)
3. Move × minority (1st-generation)				0.282
				(0.337)
1. Move × minority (2nd-generation)				0.062
				(0.200)
2. Move × minority (2nd-generation)				-0.121
				(0.210)
3. Move × minority (2nd-generation)				-0.096
				(0.329)
Pseudo R^2	0.068	0.068	0.069	0.070
AIC	28934	28932	28905	28764
Number of households	8053	8053	8053	8026
N	27814	27814	27814	27711

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Cluster robust standard errors in parentheses. Additional controls: Age^2, children living in household, change of household head, duration of residence, year.

Table 2.6: Fixed-effects estimation of perceived impairment through air pollution with time-constant average income over time per household

	FE 1	FE 2	FE 3	FE 4
Monthly equivalence income (in 1,000 Euros)	-0.017^{*}		-0.017^{*}	-0.016^{*}
	(0.008)		(0.008)	(0.008)
1. Move	-0.136^{***}	-0.282^{***}	-0.186^{***}	-0.194^{***}
	(0.040)	(0.022)	(0.041)	(0.042)
2. Move	-0.001	-0.080^{**}	-0.009	-0.018
	(0.049)	(0.025)	(0.051)	(0.051)
3. Move	-0.070	-0.015	-0.081	-0.106
	(0.104)	(0.041)	(0.105)	(0.107)
1. Move $\times$ $\overline{\text{income}}$	-0.072^{**}		-0.061^{**}	-0.057^{**}
	(0.022)		(0.022)	(0.022)
2. Move $\times$ $\overline{\text{income}}$	-0.046		-0.044	-0.040
	(0.028)		(0.028)	(0.028)
3. Move $\times$ $\overline{\text{income}}$	0.043		0.043	0.059
	(0.065)		(0.065)	(0.066)
1. Move $\times$ minority		0.165^{***}	0.151^{***}	
		(0.044)	(0.044)	
2. Move $\times$ minority		0.023	0.010	
		(0.049)	(0.049)	
3. Move $\times$ minority		0.047	0.055	
		(0.097)	(0.097)	
1. Move $\times$ minority (1st-generation)				0.213^{***}
				(0.051)
2. Move $\times$ minority (1st-generation)				0.076
				(0.060)
3. Move $\times$ minority (1st-generation)				0.159
				(0.136)
1. Move $\times$ minority (2nd-generation)				-0.001
				(0.078)
2. Move $\times$ minority (2nd-generation)				-0.050
				(0.075)
3. Move $\times$ minority (2nd-generation)				-0.029
				(0.128)
R^2	0.057	0.057	0.057	0.058
Adjusted R^2	0.056	0.056	0.057	0.058
AIC	67168	67166	67140	66869
Number of households	12037	12037	12037	12005
N	38385	38385	38385	38268

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Cluster robust standard errors in parentheses. Additional controls: Age^2, children living in household, change of household head, duration of residence, year.

Table 2.7: Fixed-effects estimation of perceived impairment through air pollution (full models)

	FE 1	FE 2	FE 3	FE 4	FE 5
Monthly equivalence income	−0.008		−0.009	−0.009	−0.010
	(0.009)		(0.009)	(0.009)	(0.009)
1. Move	−0.203***	−0.282***	−0.247***	−0.249***	−0.254***
	(0.030)	(0.022)	(0.031)	(0.032)	(0.031)
2. Move	−0.034	−0.080**	−0.039	−0.045	−0.050
	(0.039)	(0.025)	(0.042)	(0.042)	(0.041)
3. Move	−0.014	−0.015	−0.023	−0.034	−0.041
	(0.086)	(0.041)	(0.087)	(0.088)	(0.089)
1. Move × income	−0.027*		−0.022	−0.021	−0.017
	(0.013)		(0.013)	(0.013)	(0.013)
2. Move × income	−0.019		−0.020	−0.018	−0.018
	(0.020)		(0.020)	(0.020)	(0.020)
3. Move × income	0.009		0.009	0.014	0.015
	(0.044)		(0.044)	(0.045)	(0.045)
1. Move × minority		0.165***	0.159***		
		(0.044)	(0.044)		
2. Move × minority		0.023	0.015		
		(0.049)	(0.049)		
3. Move × minority		0.047	0.048		
		(0.097)	(0.096)		
1. Move × minority (1st-gen.)				0.226***	
				(0.051)	
2. Move × minority (1st-gen.)				0.083	
				(0.060)	
3. Move × minority (1st-gen.)				0.141	
				(0.135)	
1. Move × minority (2nd-gen.)				−0.005	
				(0.078)	
2. Move × minority (2nd-gen.)				−0.047	
				(0.075)	
3. Move × minority (2nd-gen.)				−0.026	
				(0.129)	
1. Move × Turkey					0.317**
					(0.101)
2. Move × Turkey					0.096
					(0.112)
3. Move × Turkey					0.328
					(0.225)
1. Move × Ex-Yugoslavia					0.349*
					(0.150)
2. Move × Ex-Yugoslavia					−0.163
					(0.154)
3. Move × Ex-Yugoslavia					0.287
					(0.377)

To be continued.

Table 2.7 continued

	FE 1	FE 2	FE 3	FE 4	FE 5
1. Move × Southern Europe					0.180
					(0.092)
2. Move × Southern Europe					0.309^{**}
					(0.113)
3. Move × Southern Europe					−0.005
					(0.232)
1. Move × Eastern Europe					0.251^{**}
					(0.084)
2. Move × Eastern Europe					−0.038
					(0.103)
3. Move × Eastern Europe					0.177
					(0.218)
1. Move × Other Western					−0.155
					(0.225)
2. Move × Other Western					0.116
					(0.282)
3. Move × Other Western					−0.780
					(0.481)
1. Move × Other country					−0.005
					(0.142)
2. Move × Other country					0.238
					(0.191)
3. Move × Other country					0.038
					(0.409)
Residence < 6 month	-0.138^{***}	-0.136^{***}	-0.138^{***}	-0.140^{***}	-0.141^{***}
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Change of household head	-0.088^{***}	-0.083^{***}	-0.087^{***}	-0.085^{***}	-0.084^{***}
	(0.024)	(0.024)	(0.024)	(0.025)	(0.025)
Age^2	$-7.8e^{-05***}$	$-7.3e^{-05***}$	$-7.7e^{-05***}$	$-8.0e^{-05***}$	$-7.9e^{-05***}$
	$(1.6e^{-05})$	$(1.6e^{-05})$	$(1.6e^{-05})$	$(1.6e^{-05})$	$(1.6e^{-05})$
Children in household	0.053^{***}	0.061^{***}	0.056^{***}	0.061^{***}	0.062^{***}
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
R^2	0.056	0.057	0.057	0.058	0.059
Adjusted R^2	0.056	0.056	0.057	0.057	0.058
AIC	67185	67166	67153	66880	66696
Number of households	12037	12037	12037	12005	11977
N	38385	38385	38385	38268	38158

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Cluster robust standard errors in parentheses.

Chapter 3

A Systematic Comparison of Spatial Regression Models Using Monte Carlo Experiments

Abstract

This study summarises the most popular spatial model specifications and offers a comparison of their performance in different scenarios by using Monte Carlo experiments. In contrast to previous simulations, this study evaluates the bias of the impacts rather than the regression coefficients and also provides results for situations with a non-spatial omitted variable bias. Results reveal that the most commonly used SAR and SEM specifications yield severe drawbacks. In contrast, spatial Durbin specifications (SDM and SDEM) as well as the simple SLX provide accurate estimates of direct impacts even in the case of misspecification. However, regarding the indirect spillover effects, several situations exist in which the SLX outperforms the more complex SDM and SDEM specifications.

3.1 Introduction

The increasing availability of spatially aggregated and georeferenced data have led to raising interest of social scientist in spatial analyses (Logan, 2012). For instance, social scientist investigate the influence of contextual conditions (e.g. Crowder, Hall & Tolnay, 2011; Friedrichs, Galster & Musterd, 2003; Kling, Liebman & Katz, 2007; Sampson, Morenoff & Earls, 1999; Sampson, Morenoff & Gannon-Rowley, 2002), or deal with explicitly spatial research questions like segregation, neighbourhood boundaries, or the exposure to environmental conditions (e.g. Dokshin, 2016; Downey, 2006b; Legewie & Schaeffer, 2016; Lichter, Parisi & Taquino, 2015; Reardon et al., 2008).

Still, researchers need to be aware of the fact that analysing spatial data also bears new challenges for the applied methods. In many cases, the spatial dynamics or relationships are of specific interest, and thus require the use of spatial methods. However, even when spatial relationships are not of explicit interest, spatial units are often not independent. As Tobler's first law of geography puts it: 'everything is related to everything else, but near things are more related than distant things' (Tobler, 1970, p. 236). For instance, it seems plausible to assume that the unemployment rate in one region is correlated to or even influenced by the unemployment rate of the neighbouring regions. In consequence, the observations cannot be considered independent and identically distributed (i.i.d.), thus violating one of the assumptions for standard linear regression models: $\mathrm{E}(\varepsilon_i\varepsilon_j) = \mathrm{E}(\varepsilon_i)\mathrm{E}(\varepsilon_j) = 0$, where ε is the random error term for each observation i and j $(i \neq j)$. An intuitive implication of this violation is that the standard OLS estimator of the equation $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ produces erroneous inferential statistics. However, spatial autocorrelation can also lead to biased point estimates, depending on the spatial process underlying the spatial autocorrelation (e.g. Pace & LeSage, 2010).

Several spatial model specifications exist to deal with this issue by explicitly modelling the spatial dependence of the data. On the one hand, those models provide a way to obtain unbiased point estimates of the true parameters. On the other hand, spatial models offer the opportunity to estimate spatial spillover coefficients, thereby informing the researcher about processes of spatial correlation or influence. Still, researchers have to deal with several decisions regarding the spatial model specifications, all using different ways to model the spatial dependence. Unfortunately, empirical specification tests of spatial models yield severe drawbacks, thus impeding a general rule of selecting the correct model specification in practical research. Therefore, it is of substantial interest to evaluate the performance of spatial model specifications under different scenarios of misspecification.

This study conducts a systematic comparison of different spatial model specifications under different scenarios of spatial dependence by using Monte Carlo experiments. This demonstrates under which conditions standard linear models yield biased estimates and how spatial model specifications perform throughout different scenarios of spatial dependence. The study extends previous simulations in several ways. First, the analysis evaluates the bias of the model specifications by relying on the impact estimates rather than the point estimates, as the model impacts are the measures of interest in applied research (LeSage & Pace, 2017). Second, it systematically evaluates the performance of the spatial model specifications under absence and presence of a non-spatial omitted variable bias. Third, it incorporates multiple explanatory variables with distinct spatial effects, as this resembles the case in applied research.

The results of the Monte Carlo experiments reveal that the most commonly used spatial models – spatial autoregressive model (SAR) and spatial error model (SEM) – have severe drawbacks in applied research. In line with previous findings, those models are outperformed by more flexible Durbin specifications. However, the results also reveal that, under quite realistic scenarios, SLX offers a better performance than the Durbin specifications, especially regarding the indirect spillover effects.

3.2 Theoretical background

As a first step in spatial econometrics, the researcher has to specify the spatial relationship between the units of observation, or more precisely, to define which units j are neighbours of unit i for all units $i = \{1, 2, ..., N\}$. This is done by setting up an $N \times N$ dimensional neighbours weights matrix $\boldsymbol{W}$, where all elements $w_{ij} > 0$ for all neighbouring units i and j $(i \neq j)$, and 0 otherwise. This study relies on a row-standardised contiguity weights matrix, defining all units as neighbours that share at least one common border. Several specifications for $\boldsymbol{W}$ exist, as for example k nearest neighbour or distance based approaches (see e.g. Dubin, 2009), and choosing the correct or incorrect specification may be vital for the results. However, these aspects have been discussed elsewhere (Corrado & Fingleton, 2012; LeSage & Pace, 2014; Neumayer & Plümper, 2016), and the focus of this study remains at the model specifications, thereby assuming a correctly specified $\boldsymbol{W}$.

3.2.1 Model specifications

As mentioned above, spatial dependence can be modelled in various ways (for a comprehensive introduction see e.g. Elhorst, 2014; LeSage & Pace, 2009). The most popular spatial model specification is the spatial autoregressive model (SAR), which incorporates the spatially weighted dependent variable $\boldsymbol{y}$ as an endogenous regressor at the right-hand side of the equation. The SAR model is defined as:

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{3.1}$$

where $\boldsymbol{y}$ is an $N \times 1$ vector of the dependent variable, $\boldsymbol{W}$ as defined above, $\boldsymbol{X}$ an $N \times K$ matrix of $k = \{1, 2, ...K\}$ covariates, and $\boldsymbol{\varepsilon}$ an $N \times 1$ vector of normally distributed disturbances. $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameter estimates and ρ represents the autoregressive scalar parameter.[1] This SAR specification assumes that the dependent variable of unit i is directly influenced by the spatially weighted dependent variable of neighbouring units j.

Another specification of spatial models is the spatial error model (SEM). In contrast to the SAR specification, the SEM explicitly models spatial autocorrelation between the disturbances $\boldsymbol{u}$, represented by the scalar parameter λ. The SEM is defined as:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \\ \boldsymbol{u} &= \lambda \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}. \end{aligned} \tag{3.2}$$

A third approach does not incorporate the spatial dependence as an autoregressive term of the dependent variable or the error term, but directly models so called spatial 'spillover' effects by including the spatially lagged covariates into the equation. This spatial lag of X (SLX) specification is defined as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{3.3}$$

where $\boldsymbol{\theta}$ is an $K \times 1$ vector of spatial spillover parameters. This model incorporates the direct effect of the covariates $\boldsymbol{\beta}$ as well as the indirect spillover effects $\boldsymbol{\theta}$ from the covariates of neighbouring units. An important property of the SLX model is that $\boldsymbol{\theta}$ constitutes a $K \times 1$ vector, thus including a distinct spatial effect for each covariate.

The three specifications shown above represent the most basic specifications of spatial models. Yet, there are further specifications which combine the models mentioned above. The spatial autoregressive combined model (SAC) comprises an autocorrelated dependent variable and an autocorrelated disturbance, resulting in:

$$\begin{aligned} \boldsymbol{y} &= \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \\ \boldsymbol{u} &= \lambda \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}. \end{aligned} \tag{3.4}$$

[1] The model intercept is omitted in all models for simplicity.

The spatial Durbin model (SDM), in contrast, combines the spatial spillover specification of the covariates (SLX) with the spatial autoregressive term of the dependent variable, resulting in:

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{W} \boldsymbol{X} \boldsymbol{\theta} + \boldsymbol{\varepsilon}. \tag{3.5}$$

A third combined model is the spatial Durbin error model (SDEM), combining the specifications of SEM and SLX:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{W} \boldsymbol{X} \boldsymbol{\theta} + \boldsymbol{u}, \\ \boldsymbol{u} &= \lambda \boldsymbol{W} \boldsymbol{u} + \boldsymbol{\varepsilon}, \end{aligned} \tag{3.6}$$

thereby comprising the spatial spillover effects of the covariates as well as an autocorrelated disturbance term.

Combining all three basic model specifications mentioned above leads to the general nesting spatial model (GNS):

$$\begin{aligned} \boldsymbol{y} &= \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{W} \boldsymbol{X} \boldsymbol{\theta} + \boldsymbol{u}, \\ \boldsymbol{u} &= \lambda \boldsymbol{W} \boldsymbol{u} + \boldsymbol{\varepsilon}. \end{aligned} \tag{3.7}$$

Though the GNS specification combines all the spatial processes of the previous specifications, this model plays only a minor role in applied research, as this specification – analogous to Manski's neighbourhood effects model (Manski, 1993) – is not or only weakly identifiable (Cook, Hays & Franzese, 2015; Gibbons & Overman, 2012).[2]

Note that most of the spatial model specifications cannot be estimated by Least Squares (LS), as using (constrained) LS estimators for models containing a spatially lagged dependent variable or disturbance leads to inconsistent results (e.g. Anselin & Bera, 1998; Franzese & Hays, 2007). However, an extensive amount of econometric literature discusses different estimation methods based on (quasi-) maximum likelihood (e.g. Anselin, 1988; Lee, 2004; Ord, 1975) or instrumental variable approaches using generalized methods of moments (e.g. Drukker, Egger & Prucha, 2013; Kelejian & Prucha, 1998, 2010), in which the endogenous lagged variables can be instrumented by q higher order lags of the exogenous regressors $(\boldsymbol{X}, \boldsymbol{W}\boldsymbol{X}, \boldsymbol{W}^2\boldsymbol{X}, ..., \boldsymbol{W}^q\boldsymbol{X})$ (Kelejian & Prucha, 1998).

3.2.2 Local and global spatial impacts

At first glance, the specifications presented above seem relatively similar in the way of modelling spatial effects. Yet, they differ in very important aspects. First, models with an endogenous spatial term (SAR, SAC, and SDM) assume a very different spatial

[2] Gibbons and Overman (2012) even argue that the SDM is only weakly identified in practice.

dependence structure than models with only exogenous spatial terms as SLX and SDEM specifications. While the first three assume global spatial dependence, the second two assume local spatial dependence (Anselin, 2003; Halleck Vega & Elhorst, 2015; LeSage & Pace, 2009). Second, the interpretation of the coefficients differs greatly between models with and without endogenous effects. This becomes apparent when considering the reduced form of the equations above. Exemplary using the SAR model of (3.1), the reduced form is given by:

$$\begin{aligned} \boldsymbol{y} - \rho \boldsymbol{W}\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \\ (\boldsymbol{I}_N - \rho \boldsymbol{W})\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \\ \boldsymbol{y} &= (\boldsymbol{I}_N - \rho \boldsymbol{W})^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}). \end{aligned} \tag{3.8}$$

When subsequently taking the first derivative of the explanatory variable $\boldsymbol{x}_k$ of the reduced form in (3.8) to interpret the partial effect of a unit change in variable $\boldsymbol{x}_k$ on $\boldsymbol{y}$, we receive

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}_k} = (\boldsymbol{I}_N - \rho \boldsymbol{W})^{-1}\beta_k, \tag{3.9}$$

for each covariate $k = \{1, 2, ..., K\}$. As can be seen from (3.9), the partial derivative with respect to $\boldsymbol{x}_k$ produces an $N \times N$ matrix, thereby representing the partial effect of each unit i onto the focal unit i itself and all other units $j = \{1, 2, ...N-1\}$. The diagonal elements of (3.9) indicate how each unit i influences itself (change of x_i on change of y_i), and each off-diagonal elements in column i represents the effect of i on each other unit j (change of x_i on change of y_j). Since reporting the individual partial effects is usually not of interest, LeSage and Pace (2009) proposed to average over these effect matrices. While the average diagonal elements of the effects matrix resulting from (3.9) represent the direct impacts of variable $\boldsymbol{x}_k$, the average column-sums of the off-diagonal elements represent the indirect impacts (or spatial spillover effects). Hence, the direct impacts can be interpreted as the average effect of a change in x_i on y_i, while the indirect impacts exhibit how a change in x_i, on average, influences all neighbouring units y_j.

Table 3.1 summarises the direct and indirect impacts for all model specifications outlined above (adopted from Halleck Vega & Elhorst, 2015, p. 345). For OLS, SEM, SLX, and SDEM, the point estimates obtained in the regression models can be interpreted as partial (direct and indirect) impacts. However, in case of SAR, SAC, and SDM, point estimates differ from the partial derivatives (or impacts). Two important consequences follow from the impact estimates presented in Table 3.1.

First, for SAR, SAC, and SDM even the direct impacts differ from the point estimates. This results from the fact that an endogenous term of the dependent variable $\boldsymbol{W}\boldsymbol{y}$ contains feedback loops through the system of neighbours (Franzese &

Table 3.1: Direct and indirect impacts, adopted from Halleck Vega and Elhorst (2015)

	Direct Impacts	Spatial Spillovers	Type
OLS/SEM	β_k	0	none
SAR/SAC	Diagonal elements of $(\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\beta_k$	Off-diagonal elements of $(\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\beta_k$	global
SLX/SDEM	β_k	$\boldsymbol{W}\theta_k$	local
SDM/GNS	Diagonal elements of $(\boldsymbol{I} - \rho \boldsymbol{W})^{-1}[\beta_k + \boldsymbol{W}\theta_k]$	Off-diagonal elements of $(\boldsymbol{I} - \rho \boldsymbol{W})^{-1}[\beta_k + \boldsymbol{W}\theta_k]$	global

Hays, 2007; Halleck Vega & Elhorst, 2015). A change of x_{ik} in the focal unit i influences the focal unit i itself, but also the neighbouring unit j, which in turn influences the focal unit i in a feedback loop. This feedback loop is part of the direct impact.

Second, the kind of indirect spillover effects in SAR, SAC, and SDM models differs from the kind of indirect spillover effects in SLX and SDEM models: while the first three specifications represent global spillover effects, the latter three represent local spillover effects (Anselin, 2003; LeSage, 2014; LeSage & Pace, 2009). In case of SLX and SDEM the spatial spillover effects can be interpreted as the effect of a one unit change of $\boldsymbol{x}_k$ in the spatially weighted neighbouring observations on the dependent variable of the focal unit; when using a row-standardised contiguity weights matrix, $\boldsymbol{W}\boldsymbol{x}_k$ is the average value of $\boldsymbol{x}_k$ in the neighbouring units. Thus, only direct neighbours – as defined in $\boldsymbol{W}$ – contribute to those local spillover effects. In contrast, spillover effects in SAR, SAC, and SDM models do not only include direct neighbours but also neighbours of neighbours (second order neighbours) and further higher-order neighbours. This can be seen by rewriting the inverse in (3.9) as power series:[3]

$$(\boldsymbol{I}_N - \rho \boldsymbol{W})^{-1}\beta_k = (\boldsymbol{I}_N + \rho \boldsymbol{W} + \rho^2 \boldsymbol{W}^2 + \rho^3 \boldsymbol{W}^3 + ...)\beta_k = (\boldsymbol{I}_N + \sum_{h=1}^{\infty} \rho^h \boldsymbol{W}^h)\beta_k, \quad (3.10)$$

where the identity matrix represents the direct effects and the sum represents the first and higher order indirect effects (including the above mentioned feedback loops). This implies that a change in one unit i does not only affect the direct neighbours but passes through the whole system towards higher-order neighbours, where the impact declines with distance within the neighbouring system. Note furthermore, that all diagonal elements $\text{diag}(\boldsymbol{W}) = w_{ii} = 0$, whereas the diagonal elements $\text{diag}(\boldsymbol{W}^2) = \text{diag}(\boldsymbol{W}\boldsymbol{W}) \neq 0$. Intuitively, $\rho \boldsymbol{W}$ only represents the effects between direct neighbours (and the focal unit is no neighbour of the focal unit itself), whereas $\rho^2 \boldsymbol{W}^2$ contains the effects of second order neighbours, where the focal unit is a second order neighbour of the focal unit itself. Thus, equation (3.10) includes feedback effects from $\rho^2 \boldsymbol{W}^2$ on.

[3] A power series of $\sum_{k=0}^{\infty} \boldsymbol{W}^k$ converges to $(\boldsymbol{I} - \boldsymbol{W})^{-1}$ if the maximum absolute eigenvalue of $\boldsymbol{W} < 1$, which is ensured by standardising $\boldsymbol{W}$.

In consequence, local and global spillover effects represent two distinct kinds of spatial spillover effects (LeSage, 2014). The interpretation of local spillover effects is straightforward: it represents the effect of all neighbours as defined by $\boldsymbol{W}$ (the average over all neighbours in case of a row-standardised weights matrix). For example, the environmental quality in the focal unit itself but also in neighbouring units could influence the attractiveness of a district and thus lead to in- or out-migration flows. In contrast, interpreting global spillover effects can be a bit more difficult. Intuitively, the global spillover effects can be seen as a kind of diffusion process. For example, an exogenous event might increase the housing prices in one district of a city, thus leading to an adaptation of housing prices in neighbouring districts, which than leads to further adaptations in other units and also to a feedback effect and further increases in the focal unit. Yet, those processes happen in time. In a cross-sectional framework, the global spillover effects are hard to interpret. Anselin (2003) proposes an interpretation as an equilibrium outcome, where the partial impact represents an estimate of how this long-run equilibrium would change due to a change in $\boldsymbol{x}_k$ (see also LeSage, 2014).

3.2.3 Bias in non-spatial OLS

So far, this study has summarised different spatial model specifications and discussed the types of spatial effects defined by those specifications. However, even if spatial effects are not of specific interest, spatial dependence can influence the estimation results. Non-spatial OLS models may not only exhibit erroneous inference but also biased estimates in some cases of spatial correlation. Still, it is important to distinguish between two kinds of biases, resulting from the discussion of direct and indirect impacts above. First, one could say that the unbiased estimate is the non-spatial parameter β_k. Second, one could also say that the unbiased estimate is the direct impact of $\boldsymbol{x}_k$, which does not only include the non-spatial effect but also the feedback loops. As discussed elsewhere (Gibbons & Overman, 2012; Gibbons, Overman & Patacchini, 2015), if a researcher is interested in the treatment effect (e.g. of a political intervention), the total direct impact including feedback loops might be of more interest than the non-spatial effect. The non-spatial parameter β_k does not include feedback effects, and thus actually underestimates the impact of a chance in x_{ik} on y_i. Note that the non-spatial parameter reflects the first derivative of the non-reduced spatial regression formula as in (3.1), while the direct impacts (with feedback effects) are given by the diagonal elements of the first derivative of the reduced form as in (3.8). Thus, the choice of which effect is the correct 'treatment' effect does also affect the discussion under which conditions a non-spatial OLS model produces biased estimates of a spatial data generating process (DGP).

Suppose the DGP follows a GNS as defined in (3.7), but we erroneously assume the DGP was $\boldsymbol{y} = \boldsymbol{x}\beta + \boldsymbol{\varepsilon}$, and use the OLS estimator $\hat{\beta} = (\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal\boldsymbol{y}$ for estimation of the parameter β. As shown by Franzese and Hays (2007), using the non-reduced form of (3.7) as DGP leads to the following estimate for β:

$$
\begin{aligned}
\hat{\beta} =&(\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal(\rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{x}\beta + \boldsymbol{W}\boldsymbol{x}\theta + \lambda\boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon})\\
=&(\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal\boldsymbol{x}\beta + (\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal\rho\boldsymbol{W}\boldsymbol{y} + (\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal\boldsymbol{W}\boldsymbol{x}\theta\\
&+ (\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal\lambda\boldsymbol{W}\boldsymbol{u} + (\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal\boldsymbol{\varepsilon}\\
\operatorname{plim}\hat{\beta} =&\beta + \rho\frac{\operatorname{Cov}(\boldsymbol{x}, \boldsymbol{W}\boldsymbol{y})}{\operatorname{Var}(\boldsymbol{x})} + \theta\frac{\operatorname{Cov}(\boldsymbol{x}, \boldsymbol{W}\boldsymbol{x})}{\operatorname{Var}(\boldsymbol{x})} + \lambda\frac{\operatorname{Cov}(\boldsymbol{x}, \boldsymbol{W}\boldsymbol{u})}{\operatorname{Var}(\boldsymbol{x})} + \frac{\operatorname{Cov}(\boldsymbol{x}, \boldsymbol{\varepsilon})}{\operatorname{Var}(\boldsymbol{x})},
\end{aligned}
\tag{3.11}
$$

where $\boldsymbol{W}\boldsymbol{y}$, $\boldsymbol{W}\boldsymbol{x}$ and $\boldsymbol{W}\boldsymbol{u}$ are omitted variables, producing a bias similar to the standard omitted variable bias resulting from $\operatorname{Cov}(\boldsymbol{x}, \boldsymbol{\varepsilon}) \neq 0$. Still, if we assume that $\boldsymbol{\varepsilon}$ is independent and randomly distributed, $\operatorname{E}(\boldsymbol{\varepsilon}|\boldsymbol{x}) = 0$, therefore also $\operatorname{E}(\boldsymbol{W}\boldsymbol{u}|\boldsymbol{x}) = 0$, which leads to an expectation of 0 for the last two terms in (3.11), as long we do not suffer from a standard (non-spatial) omitted variable bias ($\operatorname{Cov}(\boldsymbol{x}, \boldsymbol{\varepsilon}) = 0$). As in the standard case of an omitted variable, the bias resulting from estimating a non-spatial OLS if spatial dependence is present depends on $\operatorname{Cov}(\boldsymbol{x}, \boldsymbol{z})$ and $\operatorname{Cov}(\boldsymbol{y}, \boldsymbol{z})$, where $\boldsymbol{z}$ is the omitted variable $\boldsymbol{W}\boldsymbol{y}$ or $\boldsymbol{W}\boldsymbol{x}$.

Still, if one argues that the unbiased causal effect of a change in $\boldsymbol{x}$ is given by the direct impacts as described in Table 3.1, we need to rewrite the DGP in reduced form:

$$
\boldsymbol{y} =(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}(\boldsymbol{x}\beta + \boldsymbol{W}\boldsymbol{x}\theta + \boldsymbol{L}(\lambda)\boldsymbol{\varepsilon}), \tag{3.12}
$$

where $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$ are independent and randomly distributed $\mathcal{N}(0, \sigma_x^2)$ and $\mathcal{N}(0, \sigma_\varepsilon^2)$ with a mean of zero, and for simplicity we define $\boldsymbol{L}(\lambda) = (\boldsymbol{I}_N - \lambda\boldsymbol{W})^{-1}$. Consequently, we derive at a different expectation for the estimate in non-spatial OLS models:

$$
\begin{aligned}
\hat{\beta} =&(\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}(\boldsymbol{x}\beta + \boldsymbol{W}\boldsymbol{x}\theta + \boldsymbol{L}(\lambda)\boldsymbol{\varepsilon})\\
=&(\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\boldsymbol{x}\beta\\
&+ (\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\boldsymbol{W}\boldsymbol{x}\theta\\
&+ (\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\boldsymbol{L}(\lambda)\boldsymbol{\varepsilon}.
\end{aligned}
\tag{3.13}
$$

For independent random vectors $\boldsymbol{x}$ and symmetric real matrices $\boldsymbol{A}$, Girard (1989); Pace and LeSage (2010) show that $\operatorname{E}((\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal\boldsymbol{A}\boldsymbol{x}) = N^{-1}\operatorname{tr}(\boldsymbol{A})$. Furthermore, we know that $\operatorname{E}((\boldsymbol{x}^\intercal\boldsymbol{x})^{-1}\boldsymbol{x}^\intercal\boldsymbol{L}(\lambda)\boldsymbol{\varepsilon}) = 0$, as $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$ are independent, and therefore $\operatorname{E}(\boldsymbol{\varepsilon}|\boldsymbol{x}) = 0$. Thus, (3.13) can be simplified to

$$
\begin{aligned}
\operatorname{plim}\hat{\beta} =& \frac{1}{N}\mathrm{tr}((\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\beta) \\
&+ \frac{1}{N}\mathrm{tr}((\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\boldsymbol{W}\theta) \\
&= \frac{1}{N}\mathrm{tr}((\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}(\beta + \boldsymbol{W}\theta)),
\end{aligned} \tag{3.14}
$$

which equals the average diagonal elements of $(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}(\beta + \boldsymbol{W}\theta)$. Note that this equals exactly the summary measure for the direct impacts in the GNS model as defined by LeSage and Pace (2009) and described in Table 3.1. Thus, the non-spatial OLS provides an unbiased estimate of the direct impacts, though it does not provide an unbiased estimate of the parameter β.[4]

However, bias arises when not only the dependent but also the covariate $\boldsymbol{x}$ is spatially autocorrelated. To see this, we define $\boldsymbol{x} = \delta\boldsymbol{W}\boldsymbol{x} + \boldsymbol{\upsilon}$, or in reduced form $\boldsymbol{x} = (\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}\boldsymbol{\upsilon}$, where $\boldsymbol{\upsilon} \sim \mathcal{N}(0, \sigma_\upsilon^2)$, and δ denotes the autocorrelation in $\boldsymbol{x}$. Consequently, we can rewrite (3.13) as:

$$
\begin{aligned}
\hat{\beta} =& \frac{\boldsymbol{\upsilon}^\intercal((\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1})^\intercal(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}\boldsymbol{\upsilon}}{\boldsymbol{\upsilon}^\intercal((\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1})^\intercal(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}\boldsymbol{\upsilon}}\beta \\
&+ \frac{\boldsymbol{\upsilon}^\intercal((\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1})^\intercal(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}\boldsymbol{W}\boldsymbol{\upsilon}}{\boldsymbol{\upsilon}^\intercal((\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1})^\intercal(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}\boldsymbol{\upsilon}}\theta \\
&+ \frac{\boldsymbol{\upsilon}^\intercal((\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1})^\intercal(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}\boldsymbol{L}(\lambda)\boldsymbol{\varepsilon}}{\boldsymbol{\upsilon}^\intercal((\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1})^\intercal(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}\boldsymbol{\upsilon}}.
\end{aligned} \tag{3.15}
$$

Multiplying both numerator and denominator by $(\boldsymbol{\upsilon}^\intercal\boldsymbol{\upsilon})^{-1}$ (Pace & LeSage, 2010) and using $\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{x}) = 0$ leads to

$$
\begin{aligned}
\operatorname{plim}\hat{\beta} =& \frac{\mathrm{tr}\Big[(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}((\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1})^\intercal(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\Big]}{\mathrm{tr}\Big[((\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1})^\intercal(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}\Big]}\beta \\
&+ \frac{\mathrm{tr}\Big[(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}((\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1})^\intercal(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\boldsymbol{W}\Big]}{\mathrm{tr}\Big[((\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1})^\intercal(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}\Big]}\theta.
\end{aligned} \tag{3.16}
$$

To see that (3.16) $> N^{-1}\mathrm{tr}((\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}(\beta + \boldsymbol{W}\theta))$ for positive values of ρ and δ, we can rewrite the traces in (3.16) as the sum of the elements of the Hadamard product

$$
\begin{aligned}
\operatorname{plim}\hat{\beta} =& \frac{\sum_{ij}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal \circ \boldsymbol{M}(\rho))_{ij}}{\mathrm{tr}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal)}\beta \\
&+ \frac{\sum_{ij}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal \circ \boldsymbol{M}(\rho)\boldsymbol{W})_{ij}}{\mathrm{tr}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal)}\theta,
\end{aligned} \tag{3.17}
$$

[4] Betz, Cook and Hollenbach (2018) show why the parameter estimates of a non-spatial model are even biased if the covariates are randomly distributed.

where $\circ$ denotes the Hadamard product, $\boldsymbol{M}(\delta) = (\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}$, and $\boldsymbol{M}(\rho) = (\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}$. Now recall that $\boldsymbol{M}(\delta) = \boldsymbol{I}_N + \delta\boldsymbol{W} + \delta^2\boldsymbol{W}^2 + ...$, thus all diagonal elements of $\boldsymbol{M}(\delta) > 1$ and all off-diagonal elements of $\boldsymbol{M}(\delta) \geq 0$ for $\delta > 0$, therefore $\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal$ is a non-negative matrix with all diagonal elements > 1. Similarly, $\boldsymbol{M}(\rho)$ is non-negative with diagonal elements > 1. It follows that

$$\sum_{ij}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal \circ \boldsymbol{M}(\rho))_{ij} > \mathrm{tr}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal \circ \boldsymbol{M}(\rho)), \tag{3.18}$$

and using $\mathrm{E}(\boldsymbol{a} \circ \boldsymbol{b}) = \mathrm{E}(\boldsymbol{a})\mathrm{E}(\boldsymbol{b}) + \mathrm{Cov}(\boldsymbol{a}, \boldsymbol{b})$ shows that when taking only the traces of (3.17) instead of the total sum (and leaving aside the positive off-diagonal elements):

$$\begin{aligned} \mathrm{E}\left(\frac{\mathrm{tr}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal \circ \boldsymbol{M}(\rho))}{\mathrm{tr}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal)}\right) =& \frac{1}{N}\mathrm{tr}(\boldsymbol{M}(\rho)) \\ &+ \frac{\mathrm{Cov}(\mathrm{diag}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal), \mathrm{diag}(\boldsymbol{M}(\rho))}{N^{-1}\mathrm{tr}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal)}. \end{aligned} \tag{3.19}$$

As both $\boldsymbol{M}(\delta)$ and $\boldsymbol{M}(\rho)$ are constructed from the same weights matrix $\boldsymbol{W}$, $\mathrm{Cov}(\mathrm{diag}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal), \mathrm{diag}(\boldsymbol{M}(\rho)) > 0$ for positive ρ, δ. Thus, it follows from (3.18) and (3.19) that $\mathrm{tr}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal)^{-1}\sum_{ij}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal \circ \boldsymbol{M}(\rho))_{ij}\beta > N^{-1}\mathrm{tr}(\boldsymbol{M}(\rho)\beta)$. Similarly, $\mathrm{tr}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal)^{-1}\sum_{ij}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal \circ \boldsymbol{M}(\rho)\boldsymbol{W})_{ij}\theta > N^{-1}\mathrm{tr}(\boldsymbol{M}(\rho)\boldsymbol{W}\theta)$ for positive values of δ and ρ, which adds an additional bias when $\theta > 0$. Thus, $\hat{\beta}_{OLS}$ exceeds the direct impacts of $N^{-1}\mathrm{tr}(\boldsymbol{M}(\rho)[\beta + \boldsymbol{W}\theta])$, and is upwardly biased for positive values of ρ and δ. Furthermore, this bias in the impacts is a non-linear function of the parameter estimates, thereby giving a strong motivation to compare the biases in impacts rather than coefficients.

Note that the first part of (3.17) goes to $N^{-1}\mathrm{tr}(\boldsymbol{M}(\rho)\beta)$ only if either $\rho = 0$ or $\delta = 0$ (leading to $\boldsymbol{M}(\rho) = \boldsymbol{I}_N$, or $\boldsymbol{M}(\delta) = \boldsymbol{I}_N$ respectively), and the second term goes to $N^{-1}\mathrm{tr}(\boldsymbol{M}(\rho)\boldsymbol{W}\theta)$ only if $\delta = 0$. Obviously, the latter part of the bias also disappears if $\theta = 0$. Furthermore, LeSage and Pace (2009); Pace and LeSage (2010) show that in the presence of an omitted variable and $\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{x}) \neq 0$, spatial correlation in the disturbances leads to an amplification of the non-spatial omitted variable bias. Replacing the random disturbance in (3.15) by $\boldsymbol{\varepsilon} = \gamma\boldsymbol{x} + \boldsymbol{\eta}$, where γ defines the covariance between the error term (or an omitted variable) and the covariate $\boldsymbol{x}$, adds an additional bias of the form

$$+\frac{\mathrm{tr}(\boldsymbol{M}(\delta)\boldsymbol{M}(\delta)^\intercal\boldsymbol{M}(\rho)\boldsymbol{L}(\lambda))}{\mathrm{tr}(\boldsymbol{M}(\delta)^\intercal\boldsymbol{M}(\delta))}\gamma \tag{3.20}$$

to equation (3.17). Following the same argument as above, (3.20) is positive and $> \gamma$ for positive parameters ρ, δ, $\lambda > 0$, but also in the case of ρ, $\delta = 0$ and $\lambda > 0$. The term (3.20) goes to γ only if both ρ, $\lambda = 0$. Thus, the non-spatial OLS estimator β_{OLS} is biased in the presence of either:

1. Spatial autocorrelation in the dependent variable ($\rho \neq 0$) and spatial autocorrelation in the covariate ($\delta \neq 0$). This bias increases with ρ, δ, and β.
2. Local spatial spillover effects ($\theta \neq 0$) and spatial autocorrelation in the covariate ($\delta \neq 0$). This is analogous to the omitted variable bias resulting from the omission of $\boldsymbol{Wx}$. It increases with θ and δ, but additionally with ρ if $\theta \neq 0$ and $\delta \neq 0$.
3. An omitted variable and $\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{x}) \neq 0$. This non-spatial omitted variable bias γ is amplified by spatial dependence in the disturbances (λ) and spatial autocorrelation in the dependent variable (ρ), but also increases with positive values of δ if either $\rho \neq 0$ or $\lambda \neq 0$. Obviously, it also increases with γ.

3.3 Model selection

By showing that the non-spatial OLS estimates are biased in some constellations of spatial dependence, the previous chapter gives a strong motivation for the use of spatial regression models. However, as described in Chapter 3.2.1, a variety of spatial model specifications exist that can be used to account for the spatial structure of the data. Thus, selecting the correct model specification remains a crucial task in applied research.

3.3.1 Specification tests

One way of selecting the model specification is the application of empirical specification tests. In general, there are two different strategies: a specific-to-general or a general-to-specific approach (Florax, Folmer & Rey, 2003; Mur & Angulo, 2009).

The specific-to-general approach is the more common way in spatial econometrics. This approach starts with the most basic non-spatial model and tests for possible misspecifications due to omitted autocorrelation in the error term or the dependent variable. Therefore, Anselin, Bera, Florax and Yoon (1996) proposed to use Lagrange multiplier (LM) tests for the hypotheses H_0: $\lambda = 0$ and H_0: $\rho = 0$, which are robust against the alternative source of spatial dependence. Though the specific-to-general approach based on the robust LM test offers a quite good performance in distinguishing between SAR, SEM, and non-spatial OLS (Florax et al., 2003), in their original paper, Anselin et al. (1996) already note the declining power of the robust LM_λ test for spatial error dependence with increasing autocorrelation in the dependent variable (indicating some uncertainty under a SAC-like DGP). Furthermore, Mur and Angulo (2009) demonstrate strong drawbacks of the specific-to-general approach under non-optimal conditions like heteroscedasticity or endogeneity. Another shortcoming of this approach is its disregard of spatial dependence from local spillover effects, as resulting from an SLX-like process. Cook et al. (2015), for instance, show theoretically that an

SLX-like dependence structure leads to the rejection of both hypotheses H_0: $\lambda = 0$ and H_0: $\rho = 0$, though no autocorrelation is present. This is also validated by simulation results: the LM test is not helpful in distinguishing between SDM- and SDEM-like processes, and produces heavily biased results in case of a GNS-like structure (see Table 3.12 of the Appendix).

The general-to-specific approach depicts the opposite way of specification search. This approach starts with the most general model and stepwise imposes restrictions on the parameters of this general model. In theory, one would start with a GNS specification and subsequently restrict the model to simplified specifications based on the significance of parameters in the GNS. The problem with this strategy is that the GNS is only weakly identified and, thus, is of little help in selecting the correct restrictions. The most intuitive alternative would be to start with one of the two-source models SDM, SDEM, or SAC. This, however, bears the risk of imposing the wrong restriction in the first place (Cook et al., 2015). Furthermore, Cook et al. (2015) show that more complicated restrictions are necessary to derive all single-source models from SDEM or SAC specifications.

Thus, both ways of specification testing suffer from different sources of uncertainty, thereby making it hard to develop a general way of empirically finding the correct model specification. Consequently, scholars have developed other strategies to select the spatial model specification.

3.3.2 Alternative strategies

Some argue that the best way of choosing the appropriate model specification is to exclude one or more sources of spatial dependence – autocorrelation in the dependent variable, autocorrelation in the disturbances, or spatial spillover effects of the covariates – by design (Gibbons & Overman, 2012; Gibbons et al., 2015). Natural experiments are probably the best way of making one or more sources of spatial dependence unlikely, thereby restricting the model alternatives to a subset of all available models. However, the opportunities to use natural experiments are restricted in social sciences, making it a favourable but often impractical way of model selection. Therefore, Cook et al. (2015) argue that theoretical considerations should guide the model selection. First, it might be possible to rule out some sources of spatial dependence by theory, and thus restricting the specification alternatives to a subset. Second, theoretical mechanisms might guide the choice of either global or local spillover effects.

Still, others (Elhorst, 2014; LeSage, 2014; LeSage & Pace, 2009) argue that there are strong analytical reasons to restrict the model specifications to a subset, as the SDM subsumes the SLX and SAR model, and the SDEM subsumes SLX and SEM. It is easily observed that SDM reduces to SLX if $\rho = 0$ and to SAR if $\boldsymbol{\theta} = 0$, while the SDEM reduces to SLX if $\lambda = 0$ and to SEM if $\boldsymbol{\theta} = 0$. Less intuitively, Anselin (1988) has also shown that the SDM subsumes the SEM. Therefore, we can express the reduced form of (3.2) and rearrange terms:

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X\beta} + (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}\boldsymbol{\varepsilon} \\
(\boldsymbol{I}_N - \lambda \boldsymbol{W})\boldsymbol{y} &= (\boldsymbol{I}_N - \lambda \boldsymbol{W})\boldsymbol{X\beta} + \boldsymbol{\varepsilon} \\
(\boldsymbol{I}_N - \lambda \boldsymbol{W})\boldsymbol{y} &= \boldsymbol{X\beta} - \lambda \boldsymbol{WX\beta} + \boldsymbol{\varepsilon} \\
\boldsymbol{y} &= (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}(\boldsymbol{X\beta} + \boldsymbol{WX\theta} + \boldsymbol{\varepsilon}).
\end{aligned} \tag{3.21}$$

Thus, the SEM constitutes a special case of an SDM with the relative simple restriction $\boldsymbol{\theta} = -\lambda\boldsymbol{\beta}$, meaning direct and indirect effects are constrained to a common factor (Anselin, 1988, 2003). The fact that SDM subsumes SAR, SLX, and SEM leads to the conclusion that applied research should only consider SDM and SDEM as model specifications (LeSage, 2014). Especially in the case of a likely omitted variable bias, LeSage and Pace (2009, p. 68) argue in favour of using the SDM.

Nonetheless, others propose to use the SLX specification as point of departure (Gibbons & Overman, 2012; Halleck Vega & Elhorst, 2015). First, scholars have argued that SAC and SDM models are only weakly identified in practice (Gibbons & Overman, 2012; Pinkse & Slade, 2010). Second, the global spillover specification in SAR, SAC, and SDM often seem to be theoretically implausible. Recall, for instance, the example of 'diffusing' housing prices, where the housing price in one district influences housing prices in neighbouring districts. Specifying a SAR-like process means that housing prices directly influence each other. Yet, it might be more plausible that housing prices are driven by the demand within the focal, but also by the demand of neighbouring units, which reflects an SLX-like structure. Third, the SLX is computationally efficient, as it can be estimated by using Least Squares. Fourth, it turns attention back to the question of whether $\boldsymbol{X}$ and $\boldsymbol{WX}$ are exogenous, which should be the main focus when investigating the dependence between $\boldsymbol{X}$ and $\boldsymbol{y}$. Furthermore, SLX, SDM, and SDEM share the advantage that all three models estimate flexible indirect spillover effects, which are not bound to a common ratio between direct and indirect effects for all covariates (as in SAR and SAC).

3.4 Monte Carlo experiment

As outlined in the previous section, quite diverging advises exist of which model to select in applied research. To further improve the discussion on the selection of spatial model specification, this study compares the performance of different spatial model specifications by using a Monte Carlo experiment. Following the discussion above, a Monte Carlo experiment should consider several aspects to provide implications for practical research. First, as proposed by LeSage and Pace (2017), it is important to evaluate the bias of the impacts rather than the point estimates. The impacts are the measures of interest in applied research and the bias in impacts follows a non-linear function of the bias in parameter estimates. Second, it is important to incorporate more than one covariate, as bias in one parameter β_k could be counterbalanced by a bias in ρ, θ_k or λ, thus producing unbiased impacts for covariate k. However, a counterbalancing bias in ρ or λ could, in turn, affect the impacts of other covariates, as the autoregressive parameters affect all covariates. Third, we should evaluate the performance in two different worlds, one without omitted variable (omv) bias and one with omv bias, as this is likely to occur in applied research.

The DGP of the Monte Carlo simulation follows a GNS and is shown in (3.22) to (3.24), where $\boldsymbol{\upsilon}_k$ and $\boldsymbol{\varepsilon}$ are independent and randomly distributed $\mathcal{N}(0, \sigma_\upsilon^2)$ and $\mathcal{N}(0, \sigma_\varepsilon^2)$ with a mean of zero, and $\boldsymbol{x}_k$ is the kth column-vector of $\boldsymbol{X}$ for $k = 1, ..., K$ covariates (K is fixed at 2 in the simulations). The parameter ρ represents the autocorrelation in the dependent variable, λ the autocorrelation in the disturbances, and δ_k the autocorrelation in covariate k.

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{u}, \tag{3.22}$$

$$\boldsymbol{u} = \lambda \boldsymbol{W} \boldsymbol{u} + \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{3.23}$$

$$\boldsymbol{x}_k = \delta_k \boldsymbol{W} \boldsymbol{x}_k + \boldsymbol{\upsilon}_k. \tag{3.24}$$

The parameter-vector $\boldsymbol{\gamma}$ in (3.23) specifies the correlation between $\boldsymbol{x}$ and the disturbance vector $\boldsymbol{u}$, thereby defining the strength of an omitted variable bias. In reduced form, this DGP can be written as

$$\begin{aligned} \boldsymbol{y} =& (\boldsymbol{I}_N - \rho \boldsymbol{W})^{-1} \Big[(\boldsymbol{I}_N - \delta_k \boldsymbol{W})^{-1} \boldsymbol{\upsilon}_k \beta_k \\ & + \boldsymbol{W} (\boldsymbol{I}_N - \delta_k \boldsymbol{W})^{-1} \boldsymbol{\upsilon}_k \theta_k \\ & + (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1} ((\boldsymbol{I}_N - \delta_k \boldsymbol{W})^{-1} \boldsymbol{\upsilon}_k \gamma_k + \boldsymbol{\varepsilon}) \Big]. \end{aligned} \tag{3.25}$$

The simulations build on a square grid of 900 observations, where all units sharing a common border are defined as neighbours. Thus, the DGP uses a row-standardised contiguity weights matrix $\boldsymbol{W}$. The parameter vector $\boldsymbol{\beta}$ was fixed at $\boldsymbol{\beta} = \begin{pmatrix} 0.2 & 0.5 \end{pmatrix}^\intercal$, and the noise parameters were fixed at σ_υ^2, $\sigma_\varepsilon^2 = 1$ for all trials. All other parameters vary between the following two options for each parameter (vector):

- $\rho \in \{0, 0.5\}$,
- $\lambda \in \{0, 0.5\}$,
- $\boldsymbol{\delta} \in \left\{ \begin{pmatrix} 0 & 0 \end{pmatrix}^\intercal, \begin{pmatrix} 0.4 & 0.7 \end{pmatrix}^\intercal \right\}$,
- $\boldsymbol{\theta} \in \left\{ \begin{pmatrix} 0 & 0 \end{pmatrix}^\intercal, \begin{pmatrix} 0.1 & 0.8 \end{pmatrix}^\intercal \right\}$,
- $\boldsymbol{\gamma} \in \left\{ \begin{pmatrix} 0 & 0 \end{pmatrix}^\intercal, \begin{pmatrix} 0.3 & 0 \end{pmatrix}^\intercal \right\}$,

leading to a total of 32 distinct combinations. Note that this selection of parameters intentionally violates the common ratio assumption between direct and indirect effects, as this should be a more common case in practical research. All combinations were simulated in 1000 trials, with the same starting seed for each combination. All spatial models were estimated using *R*'s package *spdep* (Bivand & Piras, 2015).

3.4.1 Results without omv

Figure 3.1 shows the bias of the direct and indirect impacts for the simulations without a non-spatial omitted variable bias. Respective numbers and the root mean squared error (RMSE) are shown in Appendix 3.A. Several findings can be observed in the plot.

First and less surprising, all models perform reasonably well when correctly specifying the DGP. For instance, when the DGP follows a SAR-like process (e.g. line 2 and 4), the SAR model yields very precise estimates of direct and indirect impact. Similarly, SDEM yields the lowest bias if the DGP contains positive error-correlation $\lambda > 0$ and positive local spillover effects $\boldsymbol{\theta} > 0$, but no correlation in the dependent variable $\rho = 0$ (see line 13 and 15). These findings hold over all model specifications (though SAC could be seen as an exception).

Second, OLS yields an unbiased estimate of the direct impacts in many situations. The results confirm the theoretical predictions of Chapter 3.2.3: OLS estimates of the direct impacts are only biased in case of either positive autocorrelation in the dependent variable $\rho > 0$ *and* autocorrelation in the covariate $\delta_k > 0$ or local spillover effects $\theta_k > 0$ *and* autocorrelation in the covariate $\delta_k > 0$. Furthermore, this bias is rather moderate for low values of δ, θ, and β as can been seen in the first column of Figure 3.1. Note, however, that the bias is a conservative estimate, because the simulations use a relatively symmetric neighbours weights matrix. For instance, when increasing the variance in the number of neighbours per unit (thereby increasing the covariance between the diagonals of the inverted matrices $\boldsymbol{M}(\delta)$ and $\boldsymbol{M}(\rho)$), the bias in non-spatial OLS becomes more severe (see Equation 3.19 for the theoretical explanation).

Third, SLX, SDM, and SDEM all provide quite accurate estimates of the direct impacts (most visible in column 2). SAR, SEM, and SAC, in contrast, yield some drawbacks: especially in the presence of local spillover effects, these three specifications are biased (see lower part of Figure 3.1). Furthermore, SAR and SEM suffer from bias if both autocorrelation in the disturbance as well as autocorrelation in the dependent variable are present (see line 6 and 8). Though SLX is downwardly biased in case of autocorrelation in the dependent variable and the covariates (e.g. line 12 and 16), and SDM as well as SDEM yield some bias in case of a GNS-like process (line 14 and 16), those biases are rather moderate. This indicates that SLX, SDM, and SDEM are most robust against misspecification regarding the direct impacts.

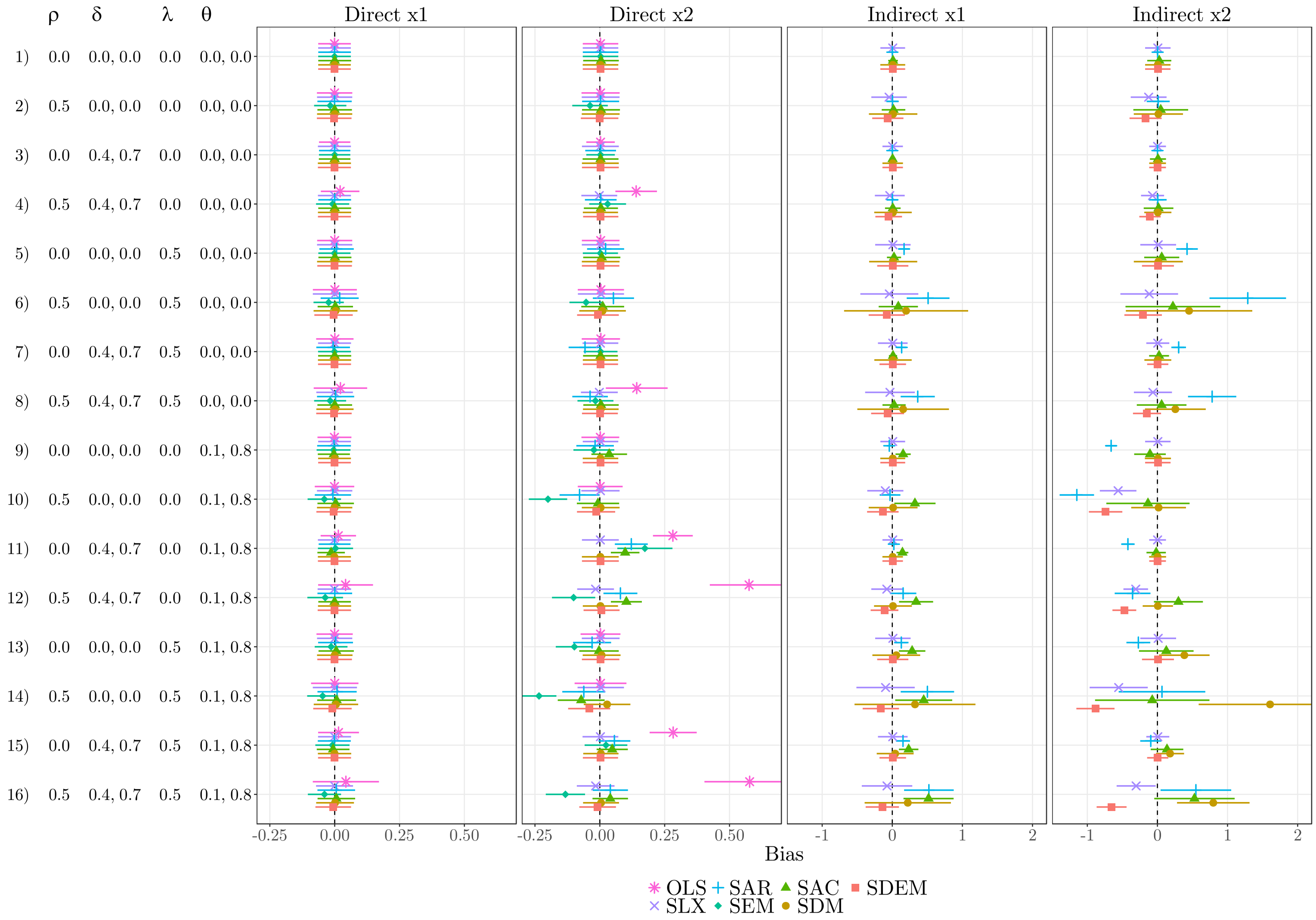

Figure 3.1: Bias of impacts and 95% confidence interval of empirical standard deviation without omv: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0, 0)^\intercal$

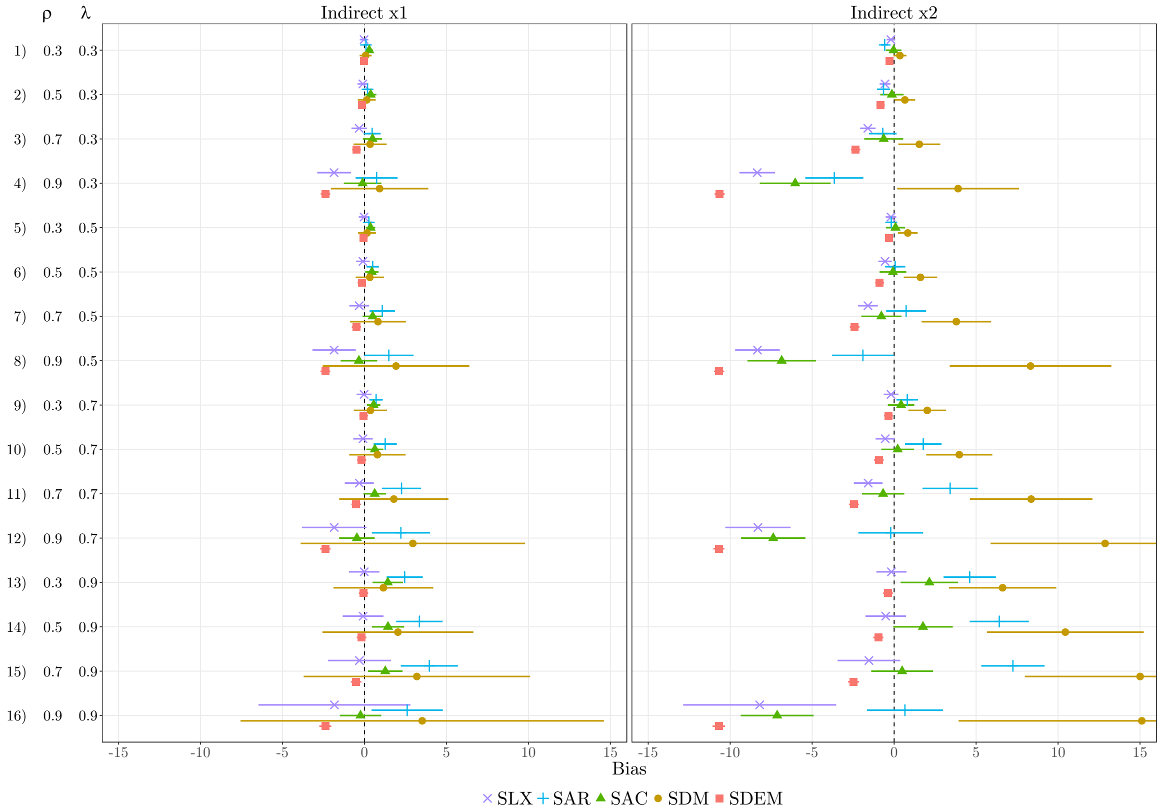

Figure 3.2: Bias of indirect impacts and 95% confidence interval of empirical standard deviation for different strengths of autocorrelation: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0, 0)^\intercal$, $\boldsymbol{\delta} = (0, 0)^\intercal$, $\boldsymbol{\theta} = (0.1, 0.8)^\intercal$

Fourth, several differences exist regarding the indirect impacts. Most obviously, the often used SAR specification suffers from considerable bias: it overestimates indirect impacts in case of autocorrelation in the disturbances, and offers biased estimates if local spillover effects exist (which are not restricted to a common ratio). The latter also applies to SAC: thought SAC offers relatively accurate estimates for $\boldsymbol{x}_2$, it overestimates indirect impacts for $\boldsymbol{x}_1$. Regarding the remaining three specifications – SLX, SDM, and SDEM – conclusions are less obvious. SDM and SDEM suffer from large bias for high values of $\boldsymbol{\theta}$ (see $\boldsymbol{x}_2$) if the DGP follows a GNS-like process (line 14 and 16): SDM overestimates the indirect impacts, while SDEM underestimates the indirect impacts. In addition, SDM performs badly if the true DGP is SDEM (line 13), and SDEM performs badly if the true DGP is SDM (line 10), whereat the bias increases with higher values of θ_k in both cases. Similar to SDEM, SLX underestimates the indirect impacts in presence of global spillovers / autocorrelation in the dependent variable.

For a better comparison, Figure 3.2 shows the bias of the indirect impacts in case of a GNS-like processes for different strengths of ρ and λ (for simplicity, we restrict the autocorrelation in covariates to zero and keep the local spillover effects fixed at $\boldsymbol{\theta} = (0.1, 0.8)^\intercal$). Respective numbers are shown in Appendix 3.B. Three findings are outstanding. First, in a GNS-like situation, the bias in SDM increases with increasing autocorrelation in $\boldsymbol{y}$ (ρ) and increasing autocorrelation in the disturbances (λ). Second, the bias in SLX and SDEM increases with higher values of ρ, but is unaffected from the strength of λ. Third, though SLX and SDEM suffer from the same problem, the bias from omitting global autocorrelation is less severe in SLX than in SDEM. Thus, the SLX outperforms SDEM. Furthermore, SLX outperforms SDM in most situations; only if the autocorrelation in the dependent variable is much stronger than the autocorrelation in the disturbances ($\rho = 0.9$, $\lambda = 0.3$), SDM yields lower bias than SLX.[5]

In sum, the results of the Monte Carlo experiments show that the three flexible model specifications of SLX, SDM, and SDEM offer an accurate estimate of the direct impacts, and all three specifications are relatively robust against misspecification regarding the direct impacts. However, results regarding the indirect impacts cast some doubt on the advice to consider only SDM or SDEM if no prior knowledge about the cause of spatial correlation is available. Especially in a 'mixed world' (where the true DGP is GNS), the results reveal that SLX offers a good alternative, which is more robust against misspecification in many situations.

3.4.2 Results with omv

So far, we have only considered the situation where $\boldsymbol{X}$ is perfectly exogenous in a non-spatial sense. However, in applied research one might often face situations in which the covariates might be correlated with the disturbances. Thus, simulations in Figure 3.3 replicate previous simulations with an omitted variable correlated with $\boldsymbol{x}_1$ ($\boldsymbol{\gamma} = \begin{pmatrix}0.3 & 0\end{pmatrix}^\intercal$) in a non-spatial way.

Comparing column 1 of Figures 3.1 and 3.3 reveals that OLS suffers from a larger bias due to spatial autocorrelation if $\boldsymbol{x}$ is correlated with the disturbance, conforming results by Pace and LeSage (2010). Regarding the direct impacts of variable $\boldsymbol{x}_1$, which is affected by the omv bias, SEM exhibits the best estimation results. Yet, the good performance of SEM regarding one variable ($\boldsymbol{x}_1$) is somewhat offset by a relatively large bias in the second variable ($\boldsymbol{x}_2$), as the SEM tends to underestimate the impact of the second variable not affected by the omitted variable in case of positive spatial autocorrelation in $\boldsymbol{y}$. As in previous results, the bias in SLX, SDM, and SDEM is comparable and lower than in the remaining specifications.

Turning to the indirect impacts, SDEM reduces the bias in the variable affected by the omv compared to other model specifications. Still, it suffers (as in the case without omv) from a bad performance in the second variable if the DGP contains an autoregressive parameter of the dependent variable ($\rho > 0$). Furthermore, it becomes apparent that the underestimation of the indirect impacts in case of an GNS-like

[5] Note that the SAC yields relatively low biases for the indirect impacts in GNS-like processes, but at the same time produces relative large biases in the direct impacts (see Table 3.10).

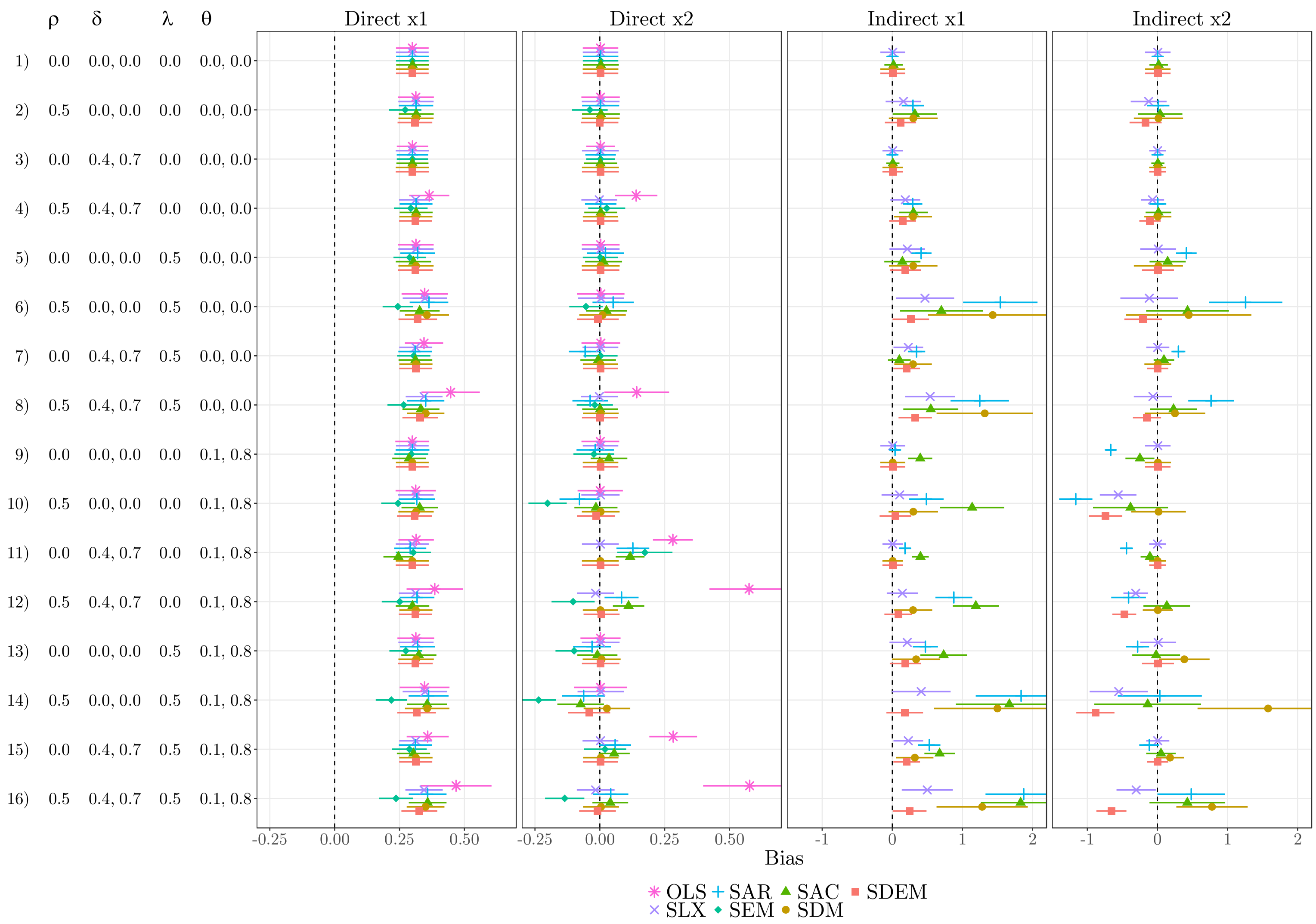

Figure 3.3: Bias of impacts and 95% confidence interval of empirical standard deviation with omv: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0.3, 0)^\intercal$

process without an omv bias (Figure 3.1) is somewhat counterbalanced by the positive omv bias leading to an overestimation of the impacts. For instance, in additional Monte Carlo experiments defining a downward bias ($\gamma_{x1} = -0.1$), SDEM amplifies the downward bias. Thus, it seems to depend on the constellation between the impacts and the omv bias whether SDEM reduces or increases the bias. However, a bias towards zero seems to be less severe than an upwardly biased parameter estimate. Another interesting finding is that the estimates of indirect impacts in SDM are most strongly affected by the omitted variable bias. Comparing column 3 in Figures 3.1 and 3.3 reveals that estimates in SDM show stronger changes than estimates of SLX and SDEM due to the non-spatial omitted variable (e.g. line 8 and 14). In sum, the second set of Monte Carlo simulations demonstrate that indirect impact estimates of SDEM and SLX are less affected by a non-spatial omv bias than SDM. Furthermore, in specific situations, SDEM may even help to reduce the non-spatial bias.

Additional simulations (available on request) reveal that the conclusions made above are robust to different variations of the parameters chosen in the DGP.[6] Obviously and as can be seen in Figure 3.2, the performance of SDM increases with increasing values of ρ relative to SDEM and SLX, and decreases with increasing values of λ (and vice versa). Still, SDM, SDEM, and SLX yield nearly equal biases regarding the direct impacts, while SLX and SDEM outperform the SDM in terms of the indirect impacts.

3.5 Conclusion

The increasing availability of spatial or georeferenced data provides the possibility to investigate spatial research questions. However, analysing spatial data also requires careful thoughts regarding the model specification. As has been shown theoretically and empirically, different constellations of the data exist that lead to biased estimates in non-spatial OLS models. Furthermore, non-spatial models disregard the spatial processes inherent in the data, thereby loosing interesting information. To overcome these problems, several spatial model specifications can be employed, which take the spatial dependence into account.

Nevertheless, the variety of specifications also bears the problem of selecting the correct specification in applied research, and specification tests are of little help in many situations. Therefore, this study employs a Monte Carlo experiment to systematically compare the bias of the most common spatial regression models in different situations of misspecification. In addition, this study extends previous simulations by relying on the impacts rather than the regression coefficients, as those impacts are the parameters of interest in applied research.

[6] Variations were specified as follows: $\rho \in \{0.3, 0.7\}$, $\lambda \in \{0.3, 0.7\}$, $\boldsymbol{\theta} = (0.1\ 0.4)^\intercal$, and $\boldsymbol{\gamma} \in \{(-0.1\ 0)^\intercal, (-0.3\ 0)^\intercal, (0.3\ 0.2)^\intercal\}$.

In line with previous studies (Elhorst, 2014; LeSage, 2014; LeSage & Pace, 2009), the Monte Carlo experiment reveals that the most commonly used SAR, SEM, and SAC specifications are outperformed by the more flexible specifications of SDM, SDEM, and SLX. Still, the results contradict the recommendation to consider only SDM and SDEM in applied research. While all three SDM, SDEM, and SLX show only marginal differences in the direct impacts, there are notable differences in the indirect spillover effects. Especially in a 'mixed world', in which the DGP follows a GNS-like structure, SLX produces less biased estimates of indirect impacts than SDM and SDEM. In the presence of non-spatial endogeneity, SDEM can help to reduce an upward bias due to omitted variables. Nonetheless, this finding needs further investigation, as conclusions seem to depend on the constellation of impacts and omv bias.

Given its good performance regarding direct as well as indirect impacts, the SLX seems to provide a worthwhile alternative to SDM and SDEM. Especially if it is not possible to eliminate one of the three sources of spatial dependence, SLX seems a good choice. Furthermore, the SLX is computational simple and intuitively interpretable. In contrast to global spillover effects in SDM, the local spillover effects can be interpreted as the effects of the average neighbours. Beyond that, the spatial spillover effects in SLX can also be 'globalised' to some extent by including separate terms for second or even higher order neighbours, which, in addition, would be a more flexible function than higher order terms of one autoregressive parameter ρ. This provides some motivation for practical research to rely on the simple SLX specification rather than the more complex SDM.

Still, it is important to keep in mind that spatial models only give parameter estimates for (conditional) correlations. Simply estimating a model from cross-sectional observational data hardly tells anything about the causal mechanisms underlying these correlations. The causal process underlying the spatial correlation can be the result of either 1) spatial interdependence in the depended variable, 2) spillover effects in the covariates, or 3) common unobserved shocks. To identify the (spatial) causal effects between two variables of interest, it is necessary to use designs or methods following the counterfactual approach like natural experiments (Angrist & Pischke, 2009, 2015; Morgan & Winship, 2015). However, this applies to all empirical research, also non-spatial observational studies. Thus, as is the case with all observational studies, spatial models can provide interesting insights into the correlational structure of the data, but can only be a first step in evaluating the causal mechanisms.

3.A Appendix Chapter 3 A: Results of Monte Carlo experiment

Table 3.2: RMSE of direct impacts without omv: $\boldsymbol{\beta} = (0.2, 0.5)^{\intercal}$, $\boldsymbol{\gamma} = (0, 0)^{\intercal}$

				OLS		SLX		SAR		SEM		SAC		SDM		SDEM	
ρ	$\boldsymbol{\delta}$	λ	$\boldsymbol{\theta}$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$
0.0	0, 0	0.0	0, 0	0.0321	0.0350*	0.0322	0.0350	0.0321	0.0351	0.0321	0.0350	0.0323	0.0353	0.0321*	0.0351	0.0322	0.0350
0.5	0, 0	0.0	0, 0	0.0348	0.0376	0.0346	0.0373	0.0337*	0.0365*	0.0361	0.0514	0.0339	0.0375	0.0346	0.0374	0.0341	0.0369
0.0	0.4, 0.7	0.0	0, 0	0.0308*	0.0281*	0.0325	0.0363	0.0308	0.0304	0.0308	0.0282	0.0310	0.0354	0.0326	0.0363	0.0325	0.0363
0.5	0.4, 0.7	0.0	0, 0	0.0433	0.1459	0.0325	0.0353	0.0322*	0.0313*	0.0335	0.0471	0.0323	0.0338	0.0326	0.0347	0.0325	0.0348
0.0	0, 0	0.5	0, 0	0.0345	0.0372	0.0346	0.0373	0.0346	0.0427	0.0317*	0.0346*	0.0325	0.0373	0.0346	0.0372	0.0341	0.0370
0.5	0, 0	0.5	0, 0	0.0432	0.0456	0.0433	0.0455	0.0421	0.0663	0.0377	0.0623	0.0348*	0.0436	0.0433	0.0470	0.0380	0.0416*
0.0	0.4, 0.7	0.5	0, 0	0.0368	0.0379	0.0324	0.0352	0.0336	0.0655	0.0323*	0.0334*	0.0325	0.0344	0.0326	0.0348	0.0325	0.0347
0.5	0.4, 0.7	0.5	0, 0	0.0570	0.1547	0.0360	0.0363	0.0367	0.0513	0.0360	0.0392	0.0341*	0.0351*	0.0366	0.0352	0.0350	0.0353
0.0	0, 0	0.0	0.1, 0.8	0.0334	0.0372	0.0322	0.0350*	0.0333	0.0411	0.0336	0.0463	0.0317*	0.0504	0.0323	0.0350	0.0322	0.0350
0.5	0, 0	0.0	0.1, 0.8	0.0388	0.0441	0.0348	0.0377	0.0363	0.0875	0.0517	0.2033	0.0361	0.0422	0.0347	0.0373*	0.0344*	0.0402
0.0	0.4, 0.7	0.0	0.1, 0.8	0.0375	0.2843	0.0325	0.0363	0.0313	0.1256	0.0347	0.1820	0.0303*	0.1016	0.0325	0.0363*	0.0325	0.0363
0.5	0.4, 0.7	0.0	0.1, 0.8	0.0684	0.5806	0.0329	0.0397	0.0341	0.0861	0.0507	0.1096	0.0320*	0.1068	0.0326	0.0347*	0.0328	0.0360
0.0	0, 0	0.5	0.1, 0.8	0.0357	0.0393	0.0346	0.0373	0.0345	0.0478	0.0351	0.1047	0.0352	0.0390	0.0349	0.0379	0.0341*	0.0370*
0.5	0, 0	0.5	0.1, 0.8	0.0466	0.0510	0.0434	0.0458*	0.0396	0.0749	0.0555	0.2369	0.0390	0.0853	0.0440	0.0537	0.0388*	0.0583
0.0	0.4, 0.7	0.5	0.1, 0.8	0.0428	0.2866	0.0324*	0.0352	0.0325	0.0644	0.0347	0.0482	0.0327	0.0568	0.0327	0.0349	0.0325	0.0347*
0.5	0.4, 0.7	0.5	0.1, 0.8	0.0781	0.5844	0.0364	0.0404	0.0375	0.0533	0.0512	0.1381	0.0375	0.0532	0.0372	0.0356*	0.0355*	0.0370
Average				0.0434	0.1518	0.0348	0.0379	0.0347	0.0587	0.0386	0.0875	0.0336*	0.0517	0.0350	0.0377*	0.0340	0.0379
				0.0976		0.0363		0.0467		0.0631		0.0427		0.0363		0.0359*	

* Lowest RMSE for $\boldsymbol{x}_k$ within the parameter combination. Number of observations=900, repetitions=1000.

Table 3.3: RMSE of indirect impacts without omv: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0, 0)^\intercal$

				SLX		SAR		SAC		SDM		SDEM	
ρ	$\boldsymbol{\delta}$	λ	$\boldsymbol{\theta}$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$
0.0	0, 0	0.0	0, 0	0.0903	0.0929	0.0121*	0.0302*	0.0366	0.0914	0.0909	0.0927	0.0907	0.0928
0.5	0, 0	0.0	0, 0	0.1371	0.1807	0.0449*	0.0843*	0.0876	0.2049	0.1766	0.1796	0.1302	0.2053
0.0	0.4, 0.7	0.0	0, 0	0.0745	0.0610	0.0113*	0.0280*	0.0240	0.0587	0.0748	0.0610	0.0748	0.0609
0.5	0.4, 0.7	0.0	0, 0	0.1143	0.1090	0.0406*	0.0644*	0.0572	0.1091	0.1368	0.0988	0.1098	0.1315
0.0	0, 0	0.5	0, 0	0.1292	0.1307	0.1728	0.4297	0.0570*	0.1415	0.1754	0.1789	0.1136	0.1165*
0.5	0, 0	0.5	0, 0	0.2150	0.2402*	0.5339	1.3185	0.1666	0.4097	0.4925	0.6443	0.1535*	0.2453
0.0	0.4, 0.7	0.5	0, 0	0.1082	0.0838	0.1377	0.3076	0.0317*	0.0760*	0.1363	0.0980	0.0968	0.0773
0.5	0.4, 0.7	0.5	0, 0	0.1837	0.1526*	0.3835	0.8000	0.0882*	0.1904	0.3674	0.3380	0.1374	0.1802
0.0	0, 0	0.0	0.1, 0.8	0.0903	0.0929	0.0467*	0.6617	0.1640	0.1566	0.0909	0.0953	0.0907	0.0928*
0.5	0, 0	0.0	0.1, 0.8	0.1648	0.5762	0.0820*	1.1564	0.3560	0.3315	0.1781	0.1999*	0.1771	0.7493
0.0	0.4, 0.7	0.0	0.1, 0.8	0.0745	0.0610	0.0341*	0.4238	0.1511	0.0726	0.0749	0.0618	0.0748	0.0609*
0.5	0.4, 0.7	0.0	0.1, 0.8	0.1373*	0.3220	0.1817	0.3776	0.3614	0.3485	0.1381	0.1104*	0.1500	0.4798
0.0	0, 0	0.5	0.1, 0.8	0.1292	0.1307	0.1388	0.2855	0.2987	0.2354	0.1831	0.4253	0.1136*	0.1165*
0.5	0, 0	0.5	0.1, 0.8	0.2328	0.5913	0.5372	0.3217*	0.4937	0.4227	0.5462	1.6884	0.2103*	0.8949
0.0	0.4, 0.7	0.5	0.1, 0.8	0.1082	0.0838	0.1610	0.1220	0.2438	0.1807	0.1404	0.2072	0.0968*	0.0773*
0.5	0.4, 0.7	0.5	0.1, 0.8	0.1983	0.3353*	0.5526	0.6058	0.5483	0.6037	0.3837	0.8393	0.1870*	0.6627
Average				0.1367	0.2028*	0.1919	0.4386	0.1979	0.2271	0.2116	0.3324	0.1254*	0.2652
				0.1697*		0.3153		0.2125		0.2720		0.1953	

* Lowest RMSE for $\boldsymbol{x}_k$ within the parameter combination. Number of observations=900, repetitions=1000.

Table 3.4: RMSE of direct impacts with omv: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0.3, 0)^\intercal$

				OLS		SLX		SAR		SEM		SAC		SDM		SDEM	
ρ	$\boldsymbol{\delta}$	λ	$\boldsymbol{\theta}$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$
0.0	0, 0	0.0	0, 0	0.3011	0.0350*	0.3010*	0.0350	0.3014	0.0350	0.3011	0.0350	0.3021	0.0354	0.3010	0.0352	0.3010	0.0350
0.5	0, 0	0.0	0, 0	0.3148	0.0379	0.3152	0.0373	0.3152	0.0366*	0.2736*	0.0520	0.3162	0.0374	0.3154	0.0374	0.3112	0.0369
0.0	0.4, 0.7	0.0	0, 0	0.3014	0.0281*	0.3007	0.0363	0.3015	0.0301	0.3014	0.0282	0.3014	0.0330	0.3007*	0.0364	0.3007	0.0363
0.5	0.4, 0.7	0.0	0, 0	0.3668	0.1462	0.3137	0.0353	0.3153	0.0311*	0.2951*	0.0451	0.3156	0.0326	0.3147	0.0348	0.3137	0.0348
0.0	0, 0	0.5	0, 0	0.3151	0.0373	0.3153	0.0373	0.3213	0.0424	0.2906*	0.0347*	0.3056	0.0391	0.3154	0.0374	0.3128	0.0370
0.5	0, 0	0.5	0, 0	0.3499	0.0466	0.3504	0.0456	0.3655	0.0655	0.2453*	0.0627	0.3301	0.0474	0.3583	0.0469	0.3226	0.0416*
0.0	0.4, 0.7	0.5	0, 0	0.3464	0.0384	0.3141	0.0352	0.3121	0.0651	0.3067*	0.0336*	0.3130	0.0355	0.3147	0.0347	0.3141	0.0347
0.5	0.4, 0.7	0.5	0, 0	0.4507	0.1559	0.3466	0.0363	0.3527	0.0511	0.2680*	0.0405	0.3345	0.0352*	0.3531	0.0353	0.3318	0.0353
0.0	0, 0	0.0	0.1, 0.8	0.3009	0.0372	0.3010	0.0350*	0.3018	0.0408	0.2975	0.0463	0.2884*	0.0505	0.3009	0.0351	0.3010	0.0350
0.5	0, 0	0.0	0.1, 0.8	0.3149	0.0445	0.3152	0.0377	0.3184	0.0876	0.2464*	0.2050	0.3297	0.0449	0.3155	0.0373*	0.3094	0.0403
0.0	0.4, 0.7	0.0	0.1, 0.8	0.3159	0.2843	0.3007	0.0363	0.2929	0.1314	0.3051	0.1820	0.2472*	0.1203	0.3007	0.0363*	0.3007	0.0363
0.5	0.4, 0.7	0.0	0.1, 0.8	0.3896	0.5807	0.3135	0.0397	0.3195	0.0902	0.2530*	0.1114	0.3018	0.1151	0.3147	0.0347*	0.3131	0.0360
0.0	0, 0	0.5	0.1, 0.8	0.3149	0.0395	0.3153	0.0373	0.3210	0.0478	0.2762*	0.1057	0.3262	0.0405	0.3158	0.0380	0.3128	0.0370*
0.5	0, 0	0.5	0.1, 0.8	0.3499	0.0522	0.3505	0.0459*	0.3641	0.0754	0.2205*	0.2382	0.3588	0.0870	0.3593	0.0534	0.3181	0.0584
0.0	0.4, 0.7	0.5	0.1, 0.8	0.3609	0.2867	0.3141	0.0352	0.3128	0.0669	0.2900*	0.0463	0.3055	0.0633	0.3147	0.0349	0.3141	0.0347*
0.5	0.4, 0.7	0.5	0.1, 0.8	0.4732	0.5848	0.3464	0.0404	0.3603	0.0545	0.2388*	0.1409	0.3603	0.0535	0.3523	0.0356*	0.3283	0.0372
Average				0.3479	0.1522	0.3196	0.0379	0.3235	0.0595	0.2756*	0.0880	0.3148	0.0544	0.3217	0.0377*	0.3128	0.0379
				0.2501		0.1787		0.1915		0.1818		0.1846		0.1797		0.1754*	

* Lowest RMSE for $\boldsymbol{x}_k$ within the parameter combination. Number of observations=900, repetitions=1000.

Table 3.5: RMSE of indirect impacts with omv: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0.3, 0)^\intercal$

				SLX		SAR		SAC		SDM		SDEM	
ρ	$\boldsymbol{\delta}$	λ	$\boldsymbol{\theta}$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$
0.0	0, 0	0.0	0, 0	0.0903	0.0929	0.0288*	0.0290*	0.0690	0.0693	0.0907	0.0927	0.0907	0.0928
0.5	0, 0	0.0	0, 0	0.2049	0.1807	0.3058	0.0818*	0.3601	0.1662	0.3477	0.1797	0.1629*	0.2053
0.0	0.4, 0.7	0.0	0, 0	0.0745	0.0610	0.0265*	0.0266*	0.0486	0.0482	0.0747	0.0611	0.0748	0.0609
0.5	0.4, 0.7	0.0	0, 0	0.2153	0.1091	0.2993	0.0615*	0.3197	0.0945	0.3266	0.0988	0.1764*	0.1315
0.0	0, 0	0.5	0, 0	0.2492	0.1307	0.4181	0.4194	0.1948*	0.1961	0.3467	0.1791	0.2189	0.1164*
0.5	0, 0	0.5	0, 0	0.5124	0.2405*	1.5642	1.2857	0.7623	0.5244	1.5075	0.6378	0.2954*	0.2454
0.0	0.4, 0.7	0.5	0, 0	0.2529	0.0839	0.3516	0.3032	0.1300*	0.1194	0.3261	0.0983	0.2270	0.0773*
0.5	0.4, 0.7	0.5	0, 0	0.5703	0.1533*	1.2656	0.7828	0.5845	0.2854	1.3652	0.3336	0.3490*	0.1807
0.0	0, 0	0.0	0.1, 0.8	0.0903	0.0929	0.0562*	0.6666	0.4080	0.2711	0.0910	0.0954	0.0907	0.0928*
0.5	0, 0	0.0	0.1, 0.8	0.1685	0.5761	0.5023	1.1710	1.1624	0.4710	0.3491	0.1994*	0.1234*	0.7496
0.0	0.4, 0.7	0.0	0.1, 0.8	0.0745*	0.0610	0.1867	0.4442	0.4065	0.1267	0.0750	0.0618	0.0748	0.0609*
0.5	0.4, 0.7	0.0	0.1, 0.8	0.1833	0.3220	0.8884	0.4307	1.2033	0.2167	0.3272	0.1098*	0.1325*	0.4804
0.0	0, 0	0.5	0.1, 0.8	0.2492	0.1307	0.4806	0.2946	0.7523	0.1748	0.3824	0.4252	0.2189*	0.1164*
0.5	0, 0	0.5	0.1, 0.8	0.4647	0.5912	1.8671	0.3076*	1.7147	0.4123	1.5691	1.6597	0.2219*	0.8959
0.0	0.4, 0.7	0.5	0.1, 0.8	0.2529	0.0839	0.5336	0.1372	0.6841	0.1186	0.3490	0.2067	0.2270*	0.0773*
0.5	0.4, 0.7	0.5	0.1, 0.8	0.5318	0.3355*	1.8952	0.5415	1.8538	0.5077	1.3240	0.8206	0.2762*	0.6655
Average				0.2616	0.2028*	0.6669	0.4365	0.6659	0.2376	0.5533	0.3287	0.1850*	0.2656
				0.2322		0.5517		0.4518		0.4410		0.2253*	

* Lowest RMSE for $\boldsymbol{x}_k$ within the parameter combination. Number of observations=900, repetitions=1000.

Table 3.6: Bias of direct impacts without omv: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0, 0)^\intercal$

				OLS		SLX		SAR		SEM		SAC		SDM		SDEM	
ρ	$\boldsymbol{\delta}$	λ	$\boldsymbol{\theta}$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$
0.0	0, 0	0.0	0, 0	−0.0006	0.0027	−0.0007	0.0027	−0.0004	0.0030	−0.0007	0.0028	−0.0000*	0.0045	−0.0007	0.0026*	−0.0007	0.0027
0.5	0, 0	0.0	0, 0	−0.0003	0.0029	−0.0002	0.0031	−0.0005	0.0031	−0.0172	−0.0378	−0.0001*	0.0045	−0.0001	0.0033	−0.0019	−0.0009*
0.0	0.4, 0.7	0.0	0, 0	−0.0002	0.0029	−0.0011	0.0021	−0.0001*	0.0031	−0.0002	0.0029	−0.0002	0.0028	−0.0011	0.0021*	−0.0011	0.0022
0.5	0.4, 0.7	0.0	0, 0	0.0209	0.1401	−0.0010	−0.0021*	0.0000*	0.0035	−0.0080	0.0301	0.0001	0.0042	−0.0006	0.0023	−0.0010	0.0035
0.0	0, 0	0.5	0, 0	0.0001*	0.0034	−0.0001	0.0034	0.0073	0.0223	−0.0011	0.0023*	0.0009	0.0075	−0.0002	0.0033	−0.0003	0.0031
0.5	0, 0	0.5	0, 0	0.0007	0.0041*	0.0007*	0.0044	0.0192	0.0524	−0.0232	−0.0530	0.0024	0.0112	0.0033	0.0106	−0.0044	−0.0072
0.0	0.4, 0.7	0.5	0, 0	0.0006	0.0040	−0.0006	0.0023*	−0.0060	−0.0570	−0.0008	0.0029	−0.0002*	0.0026	−0.0006	0.0025	−0.0006	0.0025
0.5	0.4, 0.7	0.5	0, 0	0.0222	0.1422	−0.0003*	−0.0018	0.0035	−0.0376	−0.0176	−0.0168	0.0003	0.0046	0.0011	0.0037	−0.0028	0.0004*
0.0	0, 0	0.0	0.1, 0.8	−0.0010	0.0020*	−0.0007	0.0027	−0.0025	−0.0182	−0.0044	−0.0229	−0.0032	0.0363	−0.0006*	0.0026	−0.0007	0.0027
0.5	0, 0	0.0	0.1, 0.8	−0.0007	0.0015*	−0.0002	0.0024	−0.0068	−0.0781	−0.0400	−0.1998	0.0039	−0.0071	−0.0001*	0.0034	−0.0036	−0.0144
0.0	0.4, 0.7	0.0	0.1, 0.8	0.0140	0.2816	−0.0011	0.0021	−0.0003*	0.1214	0.0032	0.1737	−0.0137	0.0976	−0.0011	0.0019*	−0.0011	0.0022
0.5	0.4, 0.7	0.0	0.1, 0.8	0.0420	0.5754	−0.0012	−0.0164	0.0004*	0.0794	−0.0366	−0.1011	−0.0005	0.1023	−0.0005	0.0023*	−0.0014	0.0067
0.0	0, 0	0.5	0.1, 0.8	−0.0003	0.0026*	−0.0001*	0.0034	0.0033	−0.0298	−0.0138	−0.0981	0.0058	−0.0032	0.0003	0.0070	−0.0003	0.0031
0.5	0, 0	0.5	0.1, 0.8	0.0003*	0.0027*	0.0007	0.0036	0.0092	−0.0617	−0.0464	−0.2344	0.0075	−0.0718	0.0049	0.0280	−0.0086	−0.0410
0.0	0.4, 0.7	0.5	0.1, 0.8	0.0148	0.2828	−0.0006	0.0023*	−0.0016	0.0562	−0.0087	0.0237	−0.0061	0.0478	−0.0006*	0.0041	−0.0006	0.0025
0.5	0.4, 0.7	0.5	0.1, 0.8	0.0432	0.5776	−0.0006*	−0.0160	0.0061	0.0405	−0.0394	−0.1326	0.0065	0.0402	0.0014	0.0047*	−0.0056	−0.0085
Average (absolute bias)				0.0101	0.1268	0.0006*	0.0044*	0.0042	0.0417	0.0163	0.0709	0.0032	0.0280	0.0011	0.0053	0.0022	0.0065
				0.0684		0.0025*		0.0230		0.0436		0.0156		0.0032		0.0043	

* Lowest bias for $\boldsymbol{x}_k$ within the parameter combination. Number of observations=900, repetitions=1000.

Table 3.7: Bias of indirect impacts without omv: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0, 0)^\intercal$

				SLX		SAR		SAC		SDM		SDEM	
ρ	$\boldsymbol{\delta}$	λ	$\boldsymbol{\theta}$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$
0.0	0, 0	0.0	0, 0	0.0064	0.0060	0.0013*	0.0030*	0.0105	0.0264	0.0067	0.0058	0.0065	0.0058
0.5	0, 0	0.0	0, 0	−0.0457	−0.1252	0.0030*	0.0110*	0.0175	0.0474	0.0120	0.0138	−0.0637	−0.1689
0.0	0.4, 0.7	0.0	0, 0	0.0055	0.0022	0.0006*	0.0008*	0.0043	0.0078	0.0057	0.0023	0.0055	0.0021
0.5	0.4, 0.7	0.0	0, 0	−0.0358	−0.0696	0.0016*	0.0051*	0.0073	0.0159	0.0098	0.0059	−0.0514	−0.1060
0.0	0, 0	0.5	0, 0	0.0086	0.0108	0.1679	0.4225	0.0243	0.0613	0.0127	0.0130	0.0068*	0.0076*
0.5	0, 0	0.5	0, 0	−0.0421*	−0.1169*	0.5107	1.2887	0.0857	0.2210	0.1971	0.4513	−0.0786	−0.2038
0.0	0.4, 0.7	0.5	0, 0	0.0070	0.0053	0.1335	0.3030	0.0104	0.0233	0.0102	0.0059	0.0067*	0.0044*
0.5	0.4, 0.7	0.5	0, 0	−0.0331	−0.0641	0.3629	0.7805	0.0256*	0.0600*	0.1546	0.2556	−0.0672	−0.1484
0.0	0, 0	0.0	0.1, 0.8	0.0064*	0.0060	−0.0427	−0.6605	0.1543	−0.1062	0.0065	0.0068	0.0065	0.0058*
0.5	0, 0	0.0	0.1, 0.8	−0.0995	−0.5606	−0.0322	−1.1496	0.3229	−0.1349	0.0119*	0.0165*	−0.1349	−0.7393
0.0	0.4, 0.7	0.0	0.1, 0.8	0.0055*	0.0022	0.0223	−0.4209	0.1448	−0.0178	0.0056	0.0025	0.0055	0.0021*
0.5	0.4, 0.7	0.0	0.1, 0.8	−0.0777	−0.3093	0.1544	−0.3544	0.3395	0.2994	0.0096*	0.0057*	−0.1110	−0.4720
0.0	0, 0	0.5	0.1, 0.8	0.0086	0.0108	0.1286	−0.2720	0.2828	0.1264	0.0570	0.3836	0.0068*	0.0076*
0.5	0, 0	0.5	0.1, 0.8	−0.0958*	−0.5523	0.5008	0.0649*	0.4480	−0.0739	0.3238	1.6066	−0.1637	−0.8841
0.0	0.4, 0.7	0.5	0.1, 0.8	0.0070	0.0053	0.1526	−0.0948	0.2333	0.1365	0.0413	0.1804	0.0067*	0.0044*
0.5	0.4, 0.7	0.5	0.1, 0.8	−0.0750*	−0.3038*	0.5222	0.5486	0.5172	0.5282	0.2208	0.7967	−0.1425	−0.6537
Average (absolute bias)				0.0350*	0.1344	0.1711	0.3988	0.1643	0.1179*	0.0678	0.2345	0.0540	0.2135
				0.0847*		0.2849		0.1411		0.1512		0.1338	

* Lowest bias for $\boldsymbol{x}_k$ within the parameter combination. Number of observations=900, repetitions=1000.

Table 3.8: Bias of direct impacts with omv: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0.3, 0)^\intercal$

				OLS		SLX		SAR		SEM		SAC		SDM		SDEM	
ρ	$\boldsymbol{\delta}$	λ	$\boldsymbol{\theta}$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$
0.0	0, 0	0.0	0, 0	0.2994	0.0027	0.2993*	0.0027	0.2997	0.0030	0.2993	0.0028	0.3004	0.0040	0.2993	0.0027*	0.2993	0.0027
0.5	0, 0	0.0	0, 0	0.3128	0.0029	0.3133	0.0032	0.3133	0.0032	0.2718*	−0.0384	0.3143	0.0043	0.3135	0.0034	0.3093	−0.0009*
0.0	0.4, 0.7	0.0	0, 0	0.2998	0.0029	0.2989	0.0021*	0.2999	0.0030	0.2998	0.0029	0.2998	0.0026	0.2989*	0.0022	0.2989	0.0022
0.5	0.4, 0.7	0.0	0, 0	0.3647	0.1401	0.3120	−0.0021*	0.3137	0.0034	0.2933*	0.0267	0.3139	0.0036	0.3130	0.0024	0.3120	0.0035
0.0	0, 0	0.5	0, 0	0.3132	0.0033	0.3134	0.0034	0.3195	0.0217	0.2889*	0.0022*	0.3036	0.0145	0.3135	0.0033	0.3109	0.0031
0.5	0, 0	0.5	0, 0	0.3469	0.0041*	0.3477	0.0045	0.3635	0.0514	0.2435*	−0.0534	0.3277	0.0255	0.3557	0.0105	0.3204	−0.0072
0.0	0.4, 0.7	0.5	0, 0	0.3443	0.0041	0.3124	0.0023*	0.3104	−0.0566	0.3050*	0.0028	0.3112	−0.0064	0.3130	0.0025	0.3124	0.0025
0.5	0.4, 0.7	0.5	0, 0	0.4470	0.1423	0.3447	−0.0017	0.3508	−0.0373	0.2661*	−0.0194	0.3325	0.0012	0.3512	0.0036	0.3299	0.0004*
0.0	0, 0	0.0	0.1, 0.8	0.2990	0.0020*	0.2993	0.0027	0.2999	−0.0177	0.2956	−0.0229	0.2865*	0.0354	0.2992	0.0027	0.2993	0.0027
0.5	0, 0	0.0	0.1, 0.8	0.3124	0.0015*	0.3133	0.0024	0.3164	−0.0782	0.2442*	−0.2015	0.3277	−0.0150	0.3135	0.0034	0.3075	−0.0145
0.0	0.4, 0.7	0.0	0.1, 0.8	0.3140	0.2816	0.2989	0.0021	0.2912	0.1274	0.3032	0.1737	0.2454*	0.1169	0.2990	0.0019*	0.2989	0.0022
0.5	0.4, 0.7	0.0	0.1, 0.8	0.3857	0.5754	0.3117	−0.0163	0.3177	0.0838	0.2505*	−0.1030	0.3000	0.1109	0.3131	0.0022*	0.3114	0.0067
0.0	0, 0	0.5	0.1, 0.8	0.3128	0.0026*	0.3134	0.0034	0.3191	−0.0299	0.2743*	−0.0991	0.3244	−0.0094	0.3139	0.0070	0.3109	0.0031
0.5	0, 0	0.5	0.1, 0.8	0.3465	0.0027*	0.3477	0.0037	0.3620	−0.0624	0.2184*	−0.2358	0.3566	−0.0739	0.3566	0.0274	0.3158	−0.0412
0.0	0.4, 0.7	0.5	0.1, 0.8	0.3585	0.2828	0.3124	0.0023*	0.3112	0.0591	0.2881*	0.0200	0.3038	0.0554	0.3130	0.0040	0.3124	0.0025
0.5	0.4, 0.7	0.5	0.1, 0.8	0.4680	0.5776	0.3445	−0.0159	0.3584	0.0421	0.2365*	−0.1355	0.3584	0.0404	0.3504	0.0047*	0.3264	−0.0089
Average (absolute bias)				0.3453	0.1268	0.3177	0.0044*	0.3217	0.0425	0.2736*	0.0713	0.3129	0.0325	0.3198	0.0052	0.3110	0.0065
				0.2361		0.1611		0.1821		0.1724		0.1727		0.1625		0.1588*	

* Lowest bias for $\boldsymbol{x}_k$ within the parameter combination. Number of observations=900, repetitions=1000.

Table 3.9: Bias of indirect impacts with omv: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0.3, 0)^\intercal$

				SLX		SAR		SAC		SDM		SDEM	
ρ	$\boldsymbol{\delta}$	λ	$\boldsymbol{\theta}$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$
0.0	0, 0	0.0	0, 0	0.0064	0.0060	0.0032*	0.0031*	0.0176	0.0177	0.0067	0.0059	0.0065	0.0058
0.5	0, 0	0.0	0, 0	0.1588	−0.1251	0.2947	0.0113*	0.3220	0.0388	0.2989	0.0143	0.1165*	−0.1689
0.0	0.4, 0.7	0.0	0, 0	0.0055	0.0022	0.0018*	0.0011*	0.0083	0.0064	0.0057	0.0021	0.0055	0.0021
0.5	0.4, 0.7	0.0	0, 0	0.1858	−0.0696	0.2910	0.0058	0.3019	0.0143	0.2960	0.0058*	0.1471*	−0.1060
0.0	0, 0	0.5	0, 0	0.2131	0.0109	0.4112	0.4127	0.1437*	0.1455	0.2985	0.0137	0.1871	0.0077*
0.5	0, 0	0.5	0, 0	0.4664	−0.1164*	1.5406	1.2576	0.6997	0.4293	1.4320	0.4455	0.2638*	−0.2040
0.0	0.4, 0.7	0.5	0, 0	0.2286	0.0053	0.3459	0.2990	0.1003*	0.0930	0.2961	0.0061	0.2054	0.0045*
0.5	0.4, 0.7	0.5	0, 0	0.5404	−0.0639*	1.2476	0.7650	0.5493	0.2298	1.3194	0.2504	0.3272*	−0.1488
0.0	0, 0	0.0	0.1, 0.8	0.0064*	0.0060	0.0396	−0.6655	0.3985	−0.2498	0.0066	0.0068	0.0065	0.0058*
0.5	0, 0	0.0	0.1, 0.8	0.1050	−0.5605	0.4865	−1.1647	1.1389	−0.3839	0.2992	0.0162*	0.0445*	−0.7396
0.0	0.4, 0.7	0.0	0.1, 0.8	0.0055*	0.0022	0.1817	−0.4417	0.4020	−0.1071	0.0056	0.0024	0.0055	0.0021*
0.5	0.4, 0.7	0.0	0.1, 0.8	0.1439	−0.3093	0.8781	−0.4118	1.1915	0.1341	0.2960	0.0057*	0.0855*	−0.4725
0.0	0, 0	0.5	0.1, 0.8	0.2131	0.0109	0.4718	−0.2825	0.7329	−0.0184	0.3400	0.3834	0.1871*	0.0077*
0.5	0, 0	0.5	0.1, 0.8	0.4127	−0.5518	1.8377	0.0338*	1.6697	−0.1388	1.4996	1.5784	0.1775*	−0.8851
0.0	0.4, 0.7	0.5	0.1, 0.8	0.2286	0.0053	0.5274	−0.1162	0.6751	0.0508	0.3217	0.1799	0.2054*	0.0045*
0.5	0.4, 0.7	0.5	0.1, 0.8	0.4985	−0.3036*	1.8747	0.4823	1.8309	0.4265	1.2818	0.7785	0.2474*	−0.6565
Average (absolute bias)				0.2137	0.1343*	0.6521	0.3971	0.6364	0.1553	0.5002	0.2309	0.1387*	0.2138
				0.1740*		0.5246		0.3958		0.3656		0.1763	

* Lowest bias for $\boldsymbol{x}_k$ within the parameter combination. Number of observations=900, repetitions=1000.

3.B Appendix Chapter 3 B: Different combinations of autocorrelation

Table 3.10: Bias of direct impacts for different strengths of autocorrelation: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0, 0)^\intercal$, $\boldsymbol{\delta} = (0, 0)^\intercal$, $\boldsymbol{\theta} = (0.1, 0.8)^\intercal$

		OLS		SLX		SAR		SEM		SAC		SDM		SDEM	
ρ	λ	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$				
0.3	0.3	-0.0004	0.0022*	−0.0001*	0.0031	−0.0004	−0.0518	−0.0271	−0.1503	0.0050	−0.0142	0.0004	0.0068	−0.0017	−0.0055
0.5	0.3	-0.0002	0.0021*	0.0003	0.0030	−0.0000*	−0.0741	−0.0441	−0.2219	0.0055	−0.0464	0.0017	0.0119	−0.0065	−0.0300
0.7	0.3	-0.0000*	0.0014*	0.0007	0.0024	0.0009	−0.1104	−0.0711	−0.3317	0.0012	−0.1109	0.0048	0.0245	−0.0217	−0.0996
0.9	0.3	-0.0006*	-0.0037	0.0009	−0.0027*	−0.0080	−0.2325	−0.1360	−0.5993	−0.0270	−0.3120	0.0116	0.0494	−0.0761	−0.3347
0.3	0.5	0.0000*	0.0027*	0.0003	0.0036	0.0061	−0.0465	−0.0301	−0.1686	0.0088	−0.0334	0.0019	0.0152	−0.0026	−0.0111
0.5	0.5	0.0003*	0.0027*	0.0007	0.0036	0.0092	−0.0617	−0.0464	−0.2344	0.0075	−0.0718	0.0049	0.0280	−0.0086	−0.0410
0.7	0.5	0.0007	0.0024*	0.0013	0.0034	0.0142	−0.0871	−0.0725	−0.3389	−0.0005*	−0.1470	0.0119	0.0578	−0.0245	−0.1127
0.9	0.5	0.0005*	-0.0016	0.0018	−0.0006*	0.0039	−0.2090	−0.1364	−0.6015	−0.0356	−0.3655	0.0235	0.0998	−0.0773	−0.3407
0.3	0.7	0.0007*	0.0037*	0.0009	0.0045	0.0203	−0.0256	−0.0327	−0.1844	0.0166	−0.0452	0.0061	0.0381	−0.0037	−0.0174
0.5	0.7	0.0011*	0.0040*	0.0015	0.0049	0.0279	−0.0290	−0.0484	−0.2453	0.0129	−0.0899	0.0130	0.0680	−0.0106	−0.0510
0.7	0.7	0.0018	0.0043*	0.0023	0.0054	0.0363	−0.0438	−0.0737	−0.3452	0.0005*	−0.1747	0.0263	0.1230	−0.0265	−0.1226
0.9	0.7	0.0022*	0.0027*	0.0033	0.0037	0.0157	−0.1861	−0.1367	−0.6032	−0.0422	−0.4083	0.0360	0.1528	−0.0778	−0.3445
0.3	0.9	0.0021*	0.0069*	0.0021	0.0077	0.0599	0.0492	−0.0348	−0.1975	0.0430	−0.0172	0.0201	0.1160	−0.0047	−0.0232
0.5	0.9	0.0029*	0.0083*	0.0030	0.0092	0.0681	0.0507	−0.0499	−0.2540	0.0346	−0.0723	0.0328	0.1660	−0.0119	−0.0583
0.7	0.9	0.0042*	0.0110*	0.0044	0.0120	0.0654	0.0152	−0.0744	−0.3495	0.0148	−0.1717	0.0465	0.2142	−0.0274	−0.1284
0.9	0.9	0.0059*	0.0183*	0.0065	0.0191	0.0224	−0.1736	−0.1366	−0.6039	−0.0409	−0.4287	0.0449	0.1814	−0.0775	−0.3454
Average		0.0015*	0.0049*	0.0019	0.0056	0.0224	0.0904	0.0719	0.3393	0.0185	0.1568	0.0179	0.0845	0.0287	0.1291
(absolute bias)		0.0032*		0.0037		0.0564		0.2056		0.0877		0.0512		0.0789	

* Lowest bias for $\boldsymbol{x}_k$ within the parameter combination. Number of observations=900, repetitions=1000.

Table 3.11: Bias of indirect impacts for different strengths of autocorrelation: $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0, 0)^\intercal$, $\boldsymbol{\delta} = (0, 0)^\intercal$, $\boldsymbol{\theta} = (0.1, 0.8)^\intercal$

		SLX		SAR		SAC		SDM		SDEM	
ρ	λ	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$
0.3	0.3	-0.0260*	-0.1971	0.0890	−0.5762	0.2900	−0.0351*	0.0686	0.3528	−0.0420	−0.2831
0.5	0.3	-0.0978*	-0.5567	0.1863	−0.6482	0.3776	−0.1366*	0.1374	0.6562	−0.1525	−0.8284
0.7	0.3	-0.3245*	-1.6044	0.4664	−0.6929	0.4893	−0.6352*	0.3418	1.5363	−0.4851	−2.3461
0.9	0.3	-1.8563	-8.3520	0.7387	−3.6504*	−0.1147*	−6.0352	0.9216	3.9014	−2.3812	−10.6470
0.3	0.5	-0.0245*	-0.1936	0.2670	−0.1768	0.3690	0.0839*	0.1520	0.8308	−0.0477	−0.3147
0.5	0.5	-0.0958*	-0.5523	0.5008	0.0649*	0.4480	−0.0739	0.3238	1.6066	−0.1637	−0.8841
0.7	0.5	-0.3219*	-1.5977	1.0849	0.7313*	0.4992	−0.7792	0.8223	3.7944	−0.4983	−2.4074
0.9	0.5	-1.8524	-8.3378	1.4847	−1.9049*	−0.3411*	−6.8599	1.9225	8.3204	−2.3870	−10.6740
0.3	0.7	-0.0219*	-0.1870*	0.7086	0.8009	0.5559	0.4223	0.3612	2.0194	−0.0536	−0.3465
0.5	0.7	-0.0925*	-0.5435	1.2625	1.7761	0.6306	0.2161*	0.7878	3.9738	−0.1731	−0.9316
0.7	0.7	-0.3174*	-1.5845	2.2593	3.4158	0.6228	−0.6721*	1.7894	8.3603	−0.5078	−2.4529
0.9	0.7	-1.8458	-8.3083	2.2187	−0.2088*	−0.4630*	−7.3746	2.9508	12.8685	−2.3900	−10.6909
0.3	0.9	-0.0166*	-0.1651*	2.4503	4.6096	1.4205	2.1489	1.1525	6.6166	−0.0586	−0.3737
0.5	0.9	-0.0860*	-0.5139*	3.3502	6.4141	1.4331	1.7591	2.0378	10.4425	−0.1795	−0.9651
0.7	0.9	-0.3087*	-1.5387	3.9543	7.2462	1.2674	0.4890*	3.1944	15.0044	−0.5123	−2.4786
0.9	0.9	-1.8342	-8.2010	2.6037	0.6582*	−0.2434*	−7.1352	3.5207	15.1102	−2.3891	−10.6950
Average		0.5701*	2.6521	1.4766	2.0984*	0.5979	2.1785	1.2803	5.9622	0.7763	3.5824
(absolute bias)		1.6111		1.7875		1.3882*		3.6212		2.1794	

* Lowest bias for $\boldsymbol{x}_k$ within the parameter combination. Number of observations=900, repetitions=1000.

3.C Appendix Chapter 3 C: Lagrange multiplier test

Table 3.12: Rejection rates of H_0 (Lagrange multiplier test): $\boldsymbol{\beta} = (0.2, 0.5)^\intercal$, $\boldsymbol{\gamma} = (0, 0)^\intercal$

ρ	$\boldsymbol{\delta}$	λ	$\boldsymbol{\theta}$	LM_λ	LM_ρ	LM^*_λ	LM^*_ρ	$LM_{\lambda\rho}$
0.0	0, 0	0.0	0, 0	0.0410	0.0440	0.0520	0.0500	0.0580
0.4	0, 0	0.0	0, 0	1.0000	1.0000	0.0920	0.7750	1.0000
0.8	0, 0	0.0	0, 0	1.0000	1.0000	0.3880	0.9990	1.0000
0.0	0.4, 0.7	0.0	0, 0	0.0440	0.0520	0.0550	0.0590	0.0550
0.4	0.4, 0.7	0.0	0, 0	1.0000	1.0000	0.1340	0.9900	1.0000
0.8	0.4, 0.7	0.0	0, 0	1.0000	1.0000	0.9950	1.0000	1.0000
0.0	0, 0	0.4	0, 0	1.0000	1.0000	0.7190	0.1140	1.0000
0.4	0, 0	0.4	0, 0	1.0000	1.0000	0.8560	0.6830	1.0000
0.8	0, 0	0.4	0, 0	1.0000	1.0000	0.8380	0.9780	1.0000
0.0	0.4, 0.7	0.4	0, 0	1.0000	1.0000	0.9430	0.1100	1.0000
0.4	0.4, 0.7	0.4	0, 0	1.0000	1.0000	0.9970	0.9560	1.0000
0.8	0.4, 0.7	0.4	0, 0	1.0000	1.0000	1.0000	1.0000	1.0000
0.0	0, 0	0.8	0, 0	1.0000	1.0000	0.9990	0.2860	1.0000
0.4	0, 0	0.8	0, 0	1.0000	1.0000	0.9960	0.5190	1.0000
0.8	0, 0	0.8	0, 0	1.0000	1.0000	0.7450	0.7410	1.0000
0.0	0.4, 0.7	0.8	0, 0	1.0000	1.0000	1.0000	0.3000	1.0000
0.4	0.4, 0.7	0.8	0, 0	1.0000	1.0000	1.0000	0.7280	1.0000
0.8	0.4, 0.7	0.8	0, 0	1.0000	1.0000	1.0000	0.9410	1.0000
0.0	0, 0	0.0	0.1, 0.8	0.4240	0.9910	1.0000	1.0000	1.0000
0.4	0, 0	0.0	0.1, 0.8	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	0, 0	0.0	0.1, 0.8	1.0000	1.0000	1.0000	1.0000	1.0000
0.0	0.4, 0.7	0.0	0.1, 0.8	0.8320	1.0000	1.0000	1.0000	1.0000
0.4	0.4, 0.7	0.0	0.1, 0.8	1.0000	1.0000	0.9960	1.0000	1.0000
0.8	0.4, 0.7	0.0	0.1, 0.8	1.0000	1.0000	0.4970	1.0000	1.0000
0.0	0, 0	0.4	0.1, 0.8	1.0000	1.0000	0.9900	1.0000	1.0000
0.4	0, 0	0.4	0.1, 0.8	1.0000	1.0000	0.9530	1.0000	1.0000
0.8	0, 0	0.4	0.1, 0.8	1.0000	1.0000	0.8520	1.0000	1.0000
0.0	0.4, 0.7	0.4	0.1, 0.8	1.0000	1.0000	0.9560	1.0000	1.0000
0.4	0.4, 0.7	0.4	0.1, 0.8	1.0000	1.0000	0.0930	1.0000	1.0000
0.8	0.4, 0.7	0.4	0.1, 0.8	1.0000	1.0000	0.9600	1.0000	1.0000
0.0	0, 0	0.8	0.1, 0.8	1.0000	1.0000	0.2590	0.9930	1.0000
0.4	0, 0	0.8	0.1, 0.8	1.0000	1.0000	0.2370	0.9940	1.0000
0.8	0, 0	0.8	0.1, 0.8	1.0000	1.0000	0.1530	0.9940	1.0000
0.0	0.4, 0.7	0.8	0.1, 0.8	1.0000	1.0000	0.9880	1.0000	1.0000
0.4	0.4, 0.7	0.8	0.1, 0.8	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	0.4, 0.7	0.8	0.1, 0.8	1.0000	1.0000	1.0000	1.0000	1.0000

Number of observations=900, repetitions=1000. LM= Lagrange multiplier test, LM^*= Robust Lagrange multiplier test, each for H_0: $\lambda = 0$, H_0: $\rho = 0$, H_0: $\lambda, \rho = 0$.

Chapter 4

Neighbours Matter: A Nation-Wide Small-Area Assessment of Environmental Inequality in Germany

Abstract

This study investigates the presence of environmental inequality in Germany and analyses its spatial pattern on a very fine grained level. Using the 2011 German census and pollution measures of the E-PRTR, the study relies on nearly 100,000 one squared km census cells over Germany. SLX and community-fixed SLX models incorporate spatial spillover-effects into the analysis to account for the spatial distribution of socio-demographic characteristics. Results reveal that the share of minorities within a census cell indeed positively correlates with the exposure to industrial pollution. Furthermore, spatial spillover effects are highly relevant: the characteristics of the neighbouring spatial units matter in predicting the amount of pollution. Especially within urban areas, clusters of high minority neighbourhoods are affected by high levels of environmental pollution. This highlights the importance of spatial clustering processes in environmental inequality research.

This chapter has been published as:

Rüttenauer, T. (2018). Neighbours matter: A nation-wide small-area assessment of environmental inequality in Germany. *Social Science Research, 70*, 198-211. DOI: 10.1016/j.ssresearch.2017.11.009

4.1 Introduction

Environmental inequality research addresses the question of whether environmental pollution is unequally distributed across different groups of citizens. In the United States, a vast body of research has shown that ethnic minorities and socio-economically disadvantaged groups face a disproportionately high exposure to environmental pollution (e.g. Ash et al., 2013; Ash & Fetter, 2004; Banzhaf & Walsh, 2008; Been & Gupta, 1997; Crowder & Downey, 2010; Downey, Dubois, Hawkins & Walker, 2008; Downey & Hawkins, 2008; Mohai & Saha, 2007; Pais et al., 2014; Pastor et al., 2001; Szasz & Meuser, 2000). While the assessment of environmental inequality in some European countries has grown during the last years (e.g. Diekmann & Meyer, 2010; Funderburg & Laurian, 2015; Havard et al., 2009; Padilla et al., 2014; Richardson et al., 2010), empirical analyses in Germany are still scarce. Furthermore, the existing studies in Germany either use subjective measures of environmental pollution (Kohlhuber et al., 2006; Mielck, 2004) or focus on single regional areas (Kabisch & Haase, 2014; Raddatz & Mennis, 2013). Although this research has contributed greatly to the understanding of environmental inequality, Germany lacks a nation-wide assessment of environmental inequality that uses objective indicators for environmental pollution.

The present study addresses this gap in the literature by analysing the connection between the foreign-minority population and objectively measured industrial pollution. By connecting 2011 German census data with industrial facilities of the European Pollutant Release and Transfer Register (E-PRTR), the analyses build on an original dataset of 93,777 (1 km^2 sized) grid cells over Germany. Thus, the connection between the socio-demographic distribution and environmental pollution is evaluated on a very fine grained spatial level. The study uses the proportionate toxicity-weighted pollution from surrounding facilities as well as the proximity to the nearest facility as indicators for environmental pollution. Additionally, it incorporates housing indicators to test whether the connection between minority groups and pollution is mediated by housing-related variables.

Furthermore, this paper extends the question regarding the presence of environmental inequality in Germany in two ways. First, spatial models (SLX) are used to test for the existence of spatial spillover effects. Using these models allows for separating the effects of the unit's characteristics itself and the effects stemming from neighbouring spatial units. Hence, this study explicitly models the spatial pattern of environmental inequality and tests whether neighbouring units matter in predicting environmental pollution. Second, it investigates whether the spatial spillover effects differ between rural and urban areas. Therefore, community-fixed effects SLX models are used to compare environmental inequality within urban and rural areas. This analysis contributes to the existing literature not only by assessing the presence of environmental inequality but also by investigating its spatial structure. The results highlight the importance of spatial spillover and clustering effects in environmental inequality research.

4.2 Theoretical background

Previous scholars have argued that environmental inequality may stem from two different processes: selective siting and selective migration. While the selective siting argument posits that hazardous facilities are disproportionately sited in minority neighbourhoods, the selective migration argument assumes that minority households may disproportionately migrate into polluted areas.[1]

The reason for selective siting may be twofold. First, areas holding a high minority share are expected to face lower housing costs (Downey, 2005; Farber, 1998; Saha & Mohai, 2005; Wolverton, 2009, 2012). Second, minority groups may lack the political power to prevent the siting of new facilities (Hamilton, 1995; Mohai & Saha, 2015a; Pastor et al., 2001). Both characteristics make high minority areas an attractive location for industrial facilities. However, the empirical results regarding the causal mechanisms are mixed. While some studies find a connection between minority share and the probability of receiving an industrial facility (Funderburg & Laurian, 2015; Mohai & Saha, 2015b; Pastor et al., 2001), others find none (Downey, 2005; Oakes et al., 1996) or only weak support (Been & Gupta, 1997; Shaikh & Loomis, 1999) for selective siting.

In contrast to the selective siting argument, the selective migration argument assumes that neighbourhood characteristics are not a predictor of facility siting. Instead, the argument works the other way round: neighbourhood pollution predicts neighbourhood characteristics. This means that minority households selectively move into polluted areas. Again, two mechanisms clarify this argument. First, lower housing prices in polluted areas make these areas more attractive for low income households and, thus, to minority households because of their lower socio-economic status (Been & Gupta, 1997; Campbell et al., 2010; Crowder et al., 2011; Downey, 2005; Pais et al., 2014). This is called the 'racial income-inequality thesis', which will be tested indirectly in the analysis. Second, minority households may face significant discriminatory barriers on the housing market when trying to enter high quality neighbourhoods (Crowder et al., 2011, 2012; Pais et al., 2014). While previous research on the macro level failed to identify selective migration as a cause of environmental inequality (Been & Gupta, 1997; Downey, 2005; Funderburg & Laurian, 2015; Mohai & Saha, 2015b; Oakes et

[1] Note that minorities in the United States are predominantly defined as ethnic minorities. In Germany, minorities stem mostly from recent migration processes, and are thus defined as foreign-minority and later operationalised as foreign population. Still, it is assumed that the same mechanism apply as discussed in the American literature.

al., 1996; Pastor et al., 2001; Shaikh & Loomis, 1999), research on the household level found clear evidence for selective migration of minority households into polluted areas (Crowder & Downey, 2010; Pais et al., 2014). Furthermore, those studies on the micro level conclude that socio-economic factors can only partly explain the racial gap in moving behaviour.

Although studies conducted in Germany did not aim to compare the causes of environmental inequality, they give important insights about the presence of environmental inequality. Kohlhuber et al. (2006) find non-German citizens to perceive a higher degree of exposure to noise and air pollution also when controlling for socio-economic indicators. In the same manner, Raddatz and Mennis (2013) find the percentage of foreigners to be negatively correlated with the distance to the next industrial facility in the city of Hamburg, even when controlling for the percentage of welfare recipients. For the city of Berlin, Kabisch and Haase (2014) identify a lower provision of urban green space in city districts with a higher minority share. In addition to the identification of general environmental inequality in Germany, the first two studies show that dissimilarities in the exposure to pollution between minority and majority groups cannot be explained solely by socio-economic variables. However, while the first study uses subjective measures of pollution, the latter concentrate on single German cities. Therefore, the present study is the first nation-wide assessment of environmental inequality that uses objective measures of pollution.

4.3 Neighbours matter

Following theoretical explanations as well as findings from previous research, we expect the percentage of minority members to be positively correlated with the exposure to environmental pollution and the proximity to the next industrial facility. However, this simple correlation does not take into account the spatial distribution of minority groups and environmental pollutants. Although the cross-sectional design of this study does not allow testing causal mechanisms that lead to environmental inequality, the theoretical explanations offer some predictions of the spatial patterns that are to be expected from these processes.

First, if selective siting causes the unequal distribution of environmental pollution over different households, we expect the broader neighbourhood to be a main driver of environmental pollution. In other words, when companies decide where to place a new facility, it is not a single neighbourhood that is decisive but rather whole clusters of neighbourhoods. If the theory of selective siting is correct, companies will place facilities where low-income or minority households cluster because lower housing prices and less political protest can be expected in those areas. Therefore, not only the minority share of a unit itself, but also the minority share of the broader neighbourhood is expected to be an important predictor of environmental pollution. Furthermore, existing industrial facilities may attract new industrial facilities due to attractive infrastructure opportunities (Ard, 2016; Elliott & Frickel, 2013). This would reinforce the clustering of industrial facilities and minority neighbourhoods.

Second, if selective migration causes environmental inequality, the broader neighbourhood may be important as well. The argument of selective migration, as outlined above, only considers environmental quality as a decisive preference for a household's relocation. However, as shown theoretically by Schelling (1978) and confirmed by recent research (W. A. V. Clark & Coulter, 2015; Crowder et al., 2011, 2012; Krysan, Couper, Farley & Forman, 2009; Rathelot & Safi, 2014; Sager, 2012), minority and majority groups tend to segregate spatially because of similarity preferences, which means households prefer to have members of their own group within their neighbourhood. Thus, if households of minority groups tend to move into polluted areas because of affordable housing opportunities, they will attract even more members of the minority group while staving off majority-group households. Kim et al. (2014) show formally that the similarity preferences of households help to explain the observed disproportionate exposure of minority groups. As in the previous case of selective siting, minority households are expected to cluster around high pollution areas.

Again, it is important to note that it is impossible to distinguish between those processes – selective siting and selective migration – in a cross-sectional design. However, it is possible to test whether pollution correlates especially with high minority clusters by incorporating spatial spillover effects. More precisely, the present study will test if the minority share of neighbouring spatial units will influence the exposure to environmental pollution or, in other words, if neighbours matter. Additionally, housing market characteristics of the census tract and housing market characteristics of the neighbourhood are incorporated to investigate whether these characteristics can explain the correlation between minority share and environmental pollution.

Additionally, there might also be significant variation in the patterns of environmental inequality between different types of communities. As previous research has shown, urban and rural areas may face differences in segregation processes (Catney, 2016; Crowder et al., 2012; Sager, 2012) as well as in pollution patterns (Ard, 2015; Briggs, Abellan & Fecht, 2008). Therefore, patterns of environmental inequality may differ between urban and rural areas in two ways. First, the direct correlation between minority share and pollution may be stronger within metropolitan areas because these areas generally face higher segregation levels (Sager, 2012) and pollution may be more concentrated in these segregated areas. In line with this argument, results by Ard (2016) indicate that segregation is associated with increasing overall health risks from industrial pollution but also with additional health risks for minority neighbourhoods. Second, the spatial spillovers may be stronger in metropolitan areas. Previous research has shown that segregation between central and suburban districts within metropolitan areas (macro segregation) increased during the last decades (Lichter et al., 2015). This could indicate that the expected clustering effects may be stronger within urban areas.

On the other hand, segregation could also reduce environmental inequality. As Downey (2007) notes, minority succession processes may lead to the concentration of minority groups in city centres (or in other parts of the city) absolutely independent from the concentration of pollution sources. In contrast to the first argument, this second argument would result in a lower correlation between minority share and pollution within urban areas.

In sum, the present paper tests 1) if there is environmental inequality in Germany when using objective measures, 2) if neighbours matter in explaining the level of pollution and 3) if the patterns of environmental inequality differ between urban and rural areas.

4.4 Data and methods

4.4.1 Data

To answer these questions two different data sources are combined: demographic data of the 2011 German census and pollution reports of the European Pollutant Release and Transfer Register (E-PRTR) for 2011.

The German census (Statistische Ämter des Bundes und der Länder, 2015) provides information about the population on the basis of a one squared km grid. This grid divides Germany into 361,478 equally sized and distributed grid cells, of which 214,633 are populated by at least one household. Excluding grid cells with missing data (due to low numbers of inhabitants and confidentiality issues) leads to a final dataset of 93,777 grid cells.[2] This final dataset still contains information of nearly 73 million inhabitants (90.94% of the total German population). The 93,777 grid cells are defined as neighbourhoods in the current study, though they define no administrative or real-life boundaries of neighbourhoods, but just result from a random grid. This is an advantage for the present research as the grid boundaries itself should not be related to siting or clean-up decisions (because they are not meaningful for administrative authorities). On average, these one squared km sized neighbourhoods contain 778 inhabitants (median: 303 inhabitants) and thus, offer a very fine grained spatial level to assess the presence and the spatial patterns of environmental inequality.

The E-PRTR (European Commission, 2006) contains information about industrial facilities falling under one or more of the 65 E-PRTR industrial activities (European

[2] When including all 201,101 non-missing units into the baseline model (containing only the minority share as explanatory variable), results similar to the presented baseline model M1 are obtained.

Commission, 2006, pp. 79-82), which exceed a pollutant specific threshold of emissions (European Commission, 2006, pp. 83-86). For 2011 the dataset contains a total of 4,971 facilities, of which 1,476 report industrial emissions to air, and 3,495 report waste management activities or emission to land and water.[3] In sum, these facilities report nearly 1b tonnes of emissions to air and nearly 90m tonnes of processed waste. These pollution data are matched with the grid cells of the German census by using the georeferenced location of the facilities.

4.4.2 Variables

To measure the exposure to environmental pollution, two different indicators were constructed. The main indicator is the amount of toxicity-weighted industrial air pollution, while the proximity to the nearest industrial facility is used as a complementary indicator to examine the robustness of the results.

To compute the air pollution for each census grid cell, a 2 km buffer around each facility location was constructed and the emissions were allocated to the census cells proportional to the overlap of the 2 km buffer and the census cell boundaries. This method is similar to the method used by Banzhaf and Walsh (2008) and addresses the problem that facilities may be sited at the edges of spatial units (Mohai & Saha, 2006, 2007). In a second step, the aggregated air pollution was calculated as the toxicity-weighted sum of all air pollutants, which were allocated to each census grid cell in the first step. For weighting, the inhalation toxicity weights of the United States Environmental Protection Agency's (EPA) Risk Screening Environmental Indicators (RSEI) database were used (Environmental Protection Agency, 2015). The resulting measure is similar to the hazard-based results in the RSEI data and provides an estimate of the amount of industrial emissions weighted by each pollutant's potential for long-term health risks. This includes pollutants like chlorinated substances, heavy metals, and other gases and organic substances (like ammonia, benzene, or fluorine), while excluding greenhouse gases (like CO_2, NO_x, PM_{10}, SO_x), which are not classified as causing chronic human health impacts (e.g. cancer).[4] In a final step, the natural logarithm of this proportionate amount of toxicity-weighted air pollution was computed to take the right skewness into account.

This measure of exposure to pollution (as used by Banzhaf & Walsh, 2008) is superior to the widely used unit-coincidence methods (for a discussion see e.g. Mohai & Saha, 2007) or hazard-proximity measures (e.g. Crowder & Downey, 2010), but less

[3] Note that the E-PRTR relies on self-reported emissions, which is prone to reporting biases. However, complementary analyses using the E-PRTR reports from 2010 to 2012 and 3-years average emissions yield nearly identical results.

[4] Additional analyses not using toxicity weights but aggregating over all emission (including greenhouse gases) lead to similar conclusions but less conservative point estimates (not shown).

sophisticated than the RSEI scores used in recent studies in the United States (e.g. Ard, 2015; Ash et al., 2013). In contrast to the United States RSEI, the European E-PRTR does not offer exposure estimates based on plume models, accounting for stack high, wind direction, wind speed, and resulting surrogate dose of emissions. Though the current study uses the same air pollutants and toxicity-weights as RSEI data, it can only use a less accurate approximation of actual exposure to pollution as provided by the RSEI.

The second measure, the proximity to the nearest facility, is calculated as the inverted Euclidean distance between the centre of each census grid cell and the facility located nearest to the grid cell. In contrast to the measure of air pollution (including 1,476 facilities), the second measure also includes facilities reporting waste management activities or emission to land and water (4,971 facilities). The proximity to industrial facilities does not aim for measuring any physical health impacts, but rather represents an indicator of reduced neighbourhood quality (due to potential noise, visibility of industrial facilities, or odour). Thus, the two measures, toxicity weighted air pollution and proximity to the nearest facility, offer the possibility to evaluate the presence of environmental inequality by using two distinct dimensions of exposure to environmental pollution. While the amount of toxicity weighted pollution provides a measure related to health risks, the proximity should mostly capture disadvantages due to visibility or odour nuisances of close facilities. However, both measures can only provide an approximation of exposure to environmental pollution (though it is the best possible approximation based on the available data in Europe).

The explanatory variables stem from the 2011 census data. Minority share is available as the percentage of foreigners, which is defined as the proportion of inhabitants without a German citizenship. Additional control variables are included to test if the correlation between minority share and pollution can be explained due to differences in socio-economic characteristics. Most importantly, the living space per inhabitant can be seen as a proxy of socio-economic status.[5] Second, the percentage of vacant housing approximates the housing market conditions, as a higher percentage of vacant housing should correlate with less demand for housing and, therefore, lower housing prices. To investigate the differences between urban and rural areas, a dichotomous variable is constructed that takes the value one for all grid cells that are located within a community with at least 100,000 inhabitants, and zero otherwise. Furthermore, the number of inhabitants (which equals population density because of similar census cell sizes) as well as the percentage of inhabitants aged 65 or older are included as controls.

[5] Note that this is the best available proxy for wealth at this fine-grained spatial level. Though wealthy people might move to small flats in inner cities (e.g. to reduce commuting distance), this problems seems less severe in the present study as the analysis simultaneously controls for population density.

4.4.3 Method

In a first step, a simple OLS model is used to investigate the connection between minority share and pollution. However, the simple OLS model is not able to take the spatial nature of the data into account. Moran's I test confirms that the non-spatial OLS model exhibits a highly significant spatial residual correlation, indicating that observations are not independent. As shown by Pace and LeSage (2010), spatial autocorrelation in the dependent variable $cov(\boldsymbol{Wy}, \boldsymbol{y}) \neq 0$ does not only lead to erroneous inferences, but also to biased point estimates in case of $cov(\boldsymbol{Wx}, \boldsymbol{x}) \neq 0$, where $\boldsymbol{W}$ specifies the $N \times N$ spatial weights matrix (all $w_{ij} > 0$ for neighbouring i and $j \neq i$, and $w_{ij} = 0$ otherwise), $\boldsymbol{y}$ the $N \times 1$ vector of the dependent variable and $\boldsymbol{x}$ the $N \times 1$ vector of the independent variable. In other words, if the explanatory variable x_i of unit i is correlated with the explanatory variable x_j and the outcome variable y_j of the neighbouring unit j, OLS estimates of β_x suffer from omitted variable bias.

Several model specifications exist that incorporate the spatial dependence of the data by including spatial lags of the dependent variable (SAR), a spatially weighted error term (SEM), or both (e.g. LeSage & Pace, 2009; Ward & Gleditsch, 2008). However, the SEM does not estimate spatial spillover effects, which are of interest for the research question at hand. Though the SAR does provide estimates of direct and spillover effects, those are global spillover processes as a change in the explanatory variable of one observation affects all other observations, which is not intuitively interpretable (for a detailed explanation see e.g. LeSage, 2014). In addition, the ratio between direct and spillover effects is fixed for all parameters of the covariates as a SAR specification estimates only one autoregressive parameter. Thus, the flexible SLX model is more suitable for the research question of this study (Elhorst, 2014; Halleck Vega & Elhorst, 2015). This model allows testing whether 'neighbours matter' or more precisely, whether the minority share of neighbouring units spills over and affects the level of pollution in the focal spatial unit. Furthermore, the SLX model is attractive due its simplicity as the coefficients can be interpreted directly (in contrast to the SAR coefficients).

The SLX model is specified as

$$\boldsymbol{y} = \alpha\boldsymbol{\iota} + \boldsymbol{X\beta} + \boldsymbol{WX\theta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{W}$ and $\boldsymbol{y}$ are defined as above, $\boldsymbol{\iota}$ is a $N \times 1$ vector of ones, α the model intercept, $\boldsymbol{X}$ a $N \times K$ matrix of K covariates, and $\boldsymbol{\epsilon}$ a $N \times 1$ vector of residuals (for $i = 1, \ldots, N$ units). $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are $K \times 1$ parameter vectors: while $\boldsymbol{\beta}$ contains the parameters of the direct effect estimates, $\boldsymbol{\theta}$ contains the parameters of the local spatial spillover effects between neighbouring units. These spillover effects can be interpreted as the average effects of the focal unit's neighbours. In the present study, the weights matrix is defined

Table 4.1: Descriptive statistics

	Rural		Urban		Overall	
Variable	Mean	SD	Mean	SD	Mean	SD
Air pollution (ln kg)	0.54	(2.59)	1.79	(5.01)	0.66	(2.93)
Distance to nearest facility (km)	7.82	(4.57)	3.21	(1.81)	7.37	(4.59)
% Foreigners	3.02	(4.66)	9.00	(8.43)	3.59	(5.44)
Population	577.58	(835.71)	2649.91	(2887.97)	777.82	(1345.95)
% 65 and older	19.56	(8.74)	20.57	(7.44)	19.65	(8.63)
% Vacant housing	3.46	(4.12)	3.50	(3.54)	3.46	(4.07)
Living space	44.93	(6.66)	41.74	(5.95)	44.63	(6.66)
N	84716		9061		93777	

as contiguity (Queens) neighbours matrix, which includes all units as neighbours sharing at least one common border. As this would drop all observations without a direct neighbour (otherwise $w_{ij}x_j$ would spuriously equal zero), the nearest neighbouring unit was imputed for all units without a direct neighbour. In the final sample this affects 6.8% of the units. The outlined modelling strategy leads to an average number of 3.6 neighbours per unit (ranging from 1 to 8 neighbours). Alternative measures are considered in the discussion section.

Two different modelling strategies are used to investigate the correlation between minority share and pollution. The first strategy uses an overall SLX model (as defined above), thus investigating if the minority share in a census grid cell is correlated with pollution in a nation-wide comparison. This might, however, result from the fact that the average minority share is higher in cities or areas where also pollution is high (e.g. the mid-western region of Germany), thereby only capturing regional level-differences. Thus, the second strategy uses community-fixed effects SLX models. This strategy controls for the community-constant differences between 4,518 communities (e.g. the community's share of minorities or level of pollution) and, additionally, 'differences out' unobservable differences between the communities. Consequently, those fixed effects models investigate whether census grid cells with a high minority share are affected by a disproportionately high pollution within the communities (compared to census grid cells within the same community), independent of general regional differences. All models were estimated using *R*'s package *spdep* (Bivand & Piras, 2015).

4.5 Results

Table 4.1 shows the descriptive statistics of the used variables separated by rural and urban areas. The table includes the distance instead of the proximity to the nearest facility because it is intuitively interpretable (in contrast to its inverse). Industrial air pollution tends to be higher, and the average distance to the nearest industrial or waste

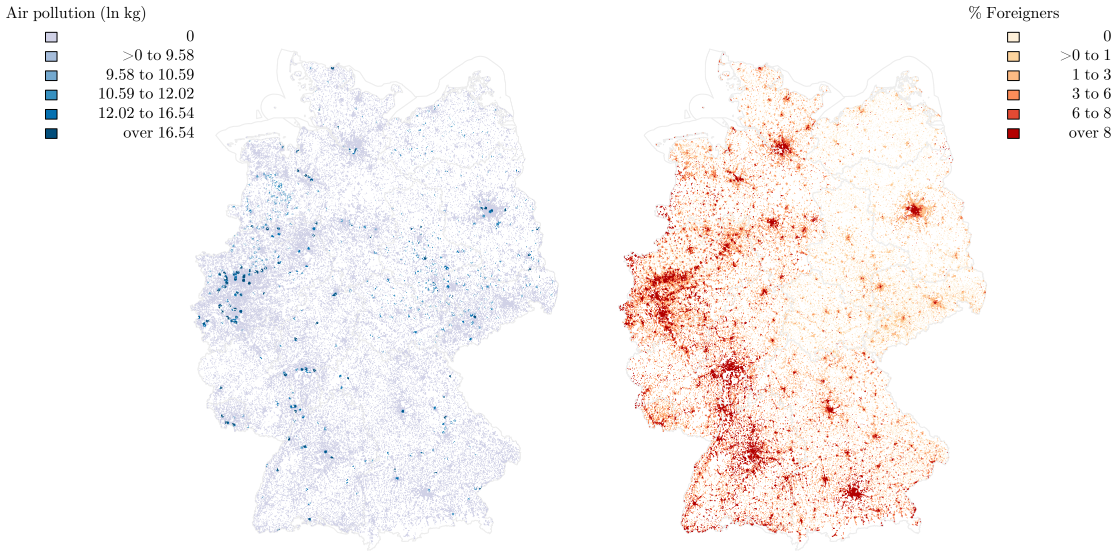

Figure 4.1: Spatial distribution of air pollution and minority share in Germany

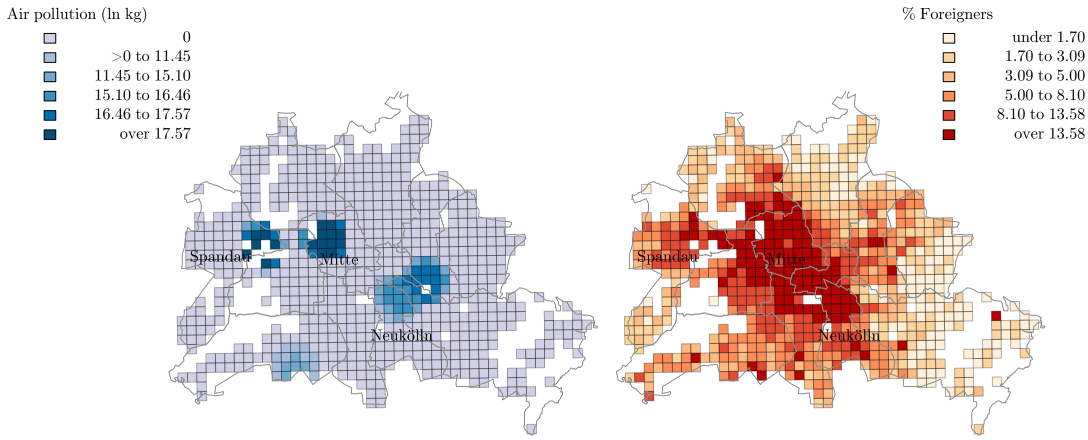

Figure 4.2: Spatial distribution of air pollution and minority share in Berlin

management facility tends to be lower in urban areas. This indicates a generally higher exposure to industrial pollution in urban areas. At the same time, the percentage of minority inhabitants living within each census cell is nearly three times higher in urban areas. These summary statistics clearly indicate that it is important to account for the differences between urban and rural areas when analysing the patterns of environmental inequality.

Figure 4.1 illustrates the spatial distribution of air pollution and minority share graphically. Analogue to the descriptive statistics, Figure 4.1 shows that highly polluted census cells as well as neighbourhoods with a high percentage of minority inhabitants cluster around urban areas. Particularly the mid-western part of Germany holds a high pollution and census cells with a high minority share. Regarding the remaining parts of Germany, high levels of minority inhabitants and pollution (to a lower extent) occur especially around the metropolitan areas. The spatial clustering of the variables of interest further supports the finding of high spatial correlation between the units of observations and illustrates graphically that the individual census cells cannot be handled as if they were independent of each other. To take the city of Berlin as an example (Figure 4.2), this finding does not only apply for the whole country but also for the distribution within cities. The Berlin district 'Mitte', for example, is characterised by a high percentage of minority inhabitants and is also affected by high pollution. Another interesting pattern can be found within the districts of 'Neukölln' and 'Spandau': within the districts, those cells with a high share of minorities are affected by the highest level of pollution. These geographical patterns offer a first sign for a positive correlation between minority share and pollution in Germany as a whole but also within metropolitan areas. Very similar patterns can be seen when looking at the proximity to the nearest industrial facility (as shown in Figures 4.3 and 4.4 in the appendix).

To test whether census cells with a higher minority share face a higher exposure to industrial pollution, Table 4.2 presents the results of the multiple regression models, regressing the industrial air pollution on the percentage of foreigners and further controls. All variables (except dichotomous variables) are standardised, allowing for the interpretation of the coefficients' magnitude in terms of standard deviations.

Model M1 represents the baseline model without any control variables or spatially lagged explanatory variables. The model confirms the impression of the previous figures: the percentage of minority inhabitants within a census cell is positively correlated with air pollution. Thus, model M1 confirms the presence of environmental inequality in Germany. Model M2 indicates, however, that the effect magnitude of model M1 is likely to be biased. When incorporating the spatially lagged minority share – meaning the average minority share in the neighbouring census cells – the model yields a much weaker direct correlation between minority share and air pollution. A big proportion of the

Table 4.2: OLS and SLX estimates. Dependent variable: industrial air pollution

	OLS	SLX		
	M1	M2	M3	M4
(Intercept)	0.000	−0.003	-0.027^{***}	-0.027^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
% Foreigners	0.120^{***}	0.054^{***}	0.039^{***}	0.012^{*}
	(0.003)	(0.005)	(0.005)	(0.005)
W [% Foreigners]		0.113^{***}	0.084^{***}	0.014^{*}
		(0.005)	(0.006)	(0.007)
Population			−0.006	0.009
			(0.005)	(0.007)
W [Population]			0.002	0.064^{***}
			(0.007)	(0.010)
% 65 and older			0.003	0.003
			(0.003)	(0.003)
W [% 65 and older]			0.006	0.009
			(0.005)	(0.005)
% Vacant housing			0.009^{**}	0.003
			(0.003)	(0.004)
W [% Vacant housing]			0.010^{*}	−0.003
			(0.005)	(0.005)
Living space			-0.034^{***}	-0.027^{***}
			(0.004)	(0.004)
W [Living space]			-0.075^{***}	-0.055^{***}
			(0.005)	(0.005)
Urban			0.245^{***}	0.117^{***}
			(0.013)	(0.016)
Urban × % Foreigners				0.055^{***}
				(0.013)
Urban × W [% Foreigners]				0.290^{***}
				(0.017)
R^2	0.014	0.019	0.032	0.044
Adj. R^2	0.014	0.019	0.031	0.044
N	93777	93777	93777	93777

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standardised coefficients. Standard errors in parentheses. W is specified as contiguity neighbours weights matrix. Model M4 interacts all covariates with the urban dummy (interaction with controls not shown).

effect in model M1 seems to be driven by the spatial correlation between neighbouring units. Furthermore, it confirms the hypothesis that neighbours matter. The level of pollution is not only correlated with the minority share of the focal unit but also with the minority share of neighbouring units.

When comparing the effect magnitudes of direct and spillover effects, it is important to keep in mind that the spatially lagged minority share represents the average minority share of the neighbours. This means that an increase in the minority share in all neighbouring units together (which are 3.6 neighbours on average) by one standard deviation is correlated with a 0.11 standard deviations higher pollution in the focal unit. Consequently, one could say that the characteristics of the surrounding area matter more than the characteristics of the census cell itself. However, this does not mean that a higher minority share in just one neighbouring unit has a stronger impact than the unit specific characteristics themselves.

The results presented so far support the hypothesis that high minority neighbourhoods experience a disproportionate high exposure to environmental pollution and that neighbours matter by influencing the pollution level. However, the models do not control if the correlation between minority share and pollution is a consequence of socio-economic or housing-related characteristics. Model M3 adds several control variables, all modelled as direct and spillover effects. Living space per inhabitant and the share of vacant housing show a significant and theoretically plausible effect on the amount of pollution. More importantly, these area- and housing-related control variables explain a considerable proportion of the minority effect. The magnitude of the direct as well as the spillover effect of the minority share is reduced by more than 25%. However, supplementary models (not shown) reveal that the housing-related variables do a rather poor job in explaining the disproportionate burden of neighbourhoods facing a high minority share. The reduction of the minority effect occurs mainly due to controlling for urban areas (or if excluded due to population density). Thus, a significant proportion of the correlation between minority share and pollution found in model M2 can be ascribed to the fact that minority households cluster in urban areas, which tend to suffer from higher levels of industrial air pollution. The argument that disproportionate pollution in high minority areas stems from cheaper housing opportunities cannot be supported by the analysis.

Model M4 further investigates if the patterns of environmental inequality significantly differ between urban and rural areas. Note that all covariates are interacted with the urban dummy in model M4, as all processes may differ in urban areas (control interactions are not shown). Both interaction terms are significant and show that the direct as well as the spillover correlations between minority share and pollution are higher in urban areas. While the main effect of the urban dummy shows that people in general experience higher levels of pollution in urban areas, the interaction confirms an additional burden for minorities in urban areas. Comparing main and interacted effect indicates that the magnitude of the disadvantage results mainly from disadvantages in urban areas. Another interesting finding is that the difference in spillover effects is much stronger than the difference in direct effects. While the direct effect in urban areas is approximately six times as strong as in rural areas, the spillover effect is more than twenty times higher in urban areas. This indicates that neighbours matter especially in urban areas. Particularly in urban areas, pollution is correlated with clusters of high minority areas. This holds true even when accounting for the higher average number of neighbours in urban (5.87) than in rural (3.29) areas.

However, the results presented so far do not control for regional level-differences over Germany. Community-specific unobservables may bias the results. This might be espe-

Table 4.3: Community-fixed effects estimates. Dependent variable: industrial air pollution

	Overall	Urban	Rural	Diff
	M5	M6	M7	(M6-M7)
% Foreigners	0.062***	0.076***	0.035***	0.041*
	(0.009)	(0.016)	(0.007)	(0.017)
W [% Foreigners]	0.163***	0.391***	0.077***	0.315***
	(0.025)	(0.070)	(0.015)	(0.071)
Population	−0.010	−0.011	−0.008	−0.002
	(0.007)	(0.009)	(0.006)	(0.011)
W [Population]	−0.039	−0.124***	−0.023	−0.100**
	(0.023)	(0.030)	(0.017)	(0.034)
% 65 and older	−0.005	−0.001	0.000	0.000
	(0.004)	(0.017)	(0.003)	(0.017)
W [% 65 and older]	−0.009	−0.101	0.004	−0.105
	(0.007)	(0.072)	(0.005)	(0.072)
% Vacant housing	0.014**	0.035	0.012**	0.023
	(0.005)	(0.027)	(0.004)	(0.027)
W [% Vacant housing]	0.021*	0.087	0.020**	0.068
	(0.008)	(0.092)	(0.007)	(0.093)
Living space	−0.012**	0.014	−0.008*	0.022
	(0.004)	(0.022)	(0.003)	(0.022)
W [Living space]	−0.036***	−0.161*	−0.022***	−0.139*
	(0.007)	(0.070)	(0.006)	(0.070)
R^2	0.018	0.076	0.005	
Adj. R^2	-0.031	0.067	-0.050	
N	93777	9061	84716	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standardised coefficients. Cluster robust standard errors in parentheses. W is specified as contiguity neighbours weights matrix. The differences and standard errors in column 4 are obtained by an overall fixed-effects model with interaction terms of the urban dummy and all covariates.

cially problematic when comparing rural and urban areas. Therefore, Table 4.3 presents results from community-fixed effects models. These models control for the differences between 4,518 German communities and use only the 'within-community variance' to estimate the effects, while conditioning for all community-constant characteristics. For ease of interpretation an overall model as well as two models separately for urban and rural areas are estimated. The fourth column of Table 4.3 additionally presents a test for differences between urban and rural areas. These differences were computed by interacting all covariates in the overall model (M5) with the urban dummy.

First of all, model M5 estimates the within-community environmental inequality for all communities taken together. It turns out that the extent of environmental inequality is even stronger within communities than it is in a country-wide comparison (comparing M3 and M5). This finding is in line with results from the American context, where environmental inequality is found to be stronger within than between cities (Ash & Fetter, 2004). Again, results support the hypothesis that neighbours matter. Also within communities, clusters of high minority neighbourhoods correlate significantly and positively with pollution. Additionally, the minority share is by far the strongest predictor of environmental pollution, though the models include controls for the share of vacant housing and living space per inhabitant, which should be more important if the socio-economic disadvantages of minority groups and resulting preferences for cheap housing are responsible for the higher exposure to environmental pollution.

Models M6 and M7 compare whether the patterns of environmental inequality differ between urban and rural areas. A first interesting finding is that population density of the surrounding area is a significant predictor of pollution only in urban areas, whereas the percentage of vacant housing correlates with pollution only in rural areas. In both urban and rural areas, pollution is significantly lower in areas with higher living space per inhabitant, though the direct effect is non-significant for urban areas. Regarding the minority share, the direct effect in urban areas is twice as strong as in rural areas. Furthermore, there are highly significant and strong differences in the spillover effects. The effect of the minority share spilling over from neighbouring units is more than four times higher within urban than within rural areas. Even when taking into account that observations in urban areas have on average 5.87 neighbours, the spatial spillover effect from a single neighbouring unit is nearly as strong in its magnitude as the effect of the focal unit's minority share itself. According to M6, a one-unit change in a single neighbouring cell has only a slightly lower impact (0.391/5.87=0.067) than a one-unit change of the focal unit's minority share in urban areas (0.076). If all neighbours increased their minority share simultaneously, this would lead to an effect that is more than four times higher than the effect of an increase in the unit's own minority share.

In sum, the analyses reveal: 1) the disadvantage of minority neighbourhoods is stronger within urban areas and 2) this stems mostly from the fact that neighbours matter especially in urban areas. In general, results confirm that neighbours matter: industrial air pollution tends to be especially high in areas where high proportions of foreign inhabitants agglomerate.

4.6 Discussion

Though the theoretical predictions are supported by the results, there remain three critical issues regarding the presented models. First, although the models account for the covariance between $\boldsymbol{y}$ and $\boldsymbol{WX}$, there still remains spatial autocorrelation in the residuals, which might result from spatially clustered unobservables that influence $\boldsymbol{y}$ and $\boldsymbol{X}$. Second, the conclusions, especially regarding the spatial spillover effects, might depend on the specification of the neighbourhoods weights matrix $\boldsymbol{W}$ and third, all analyses rely on toxicity-weighted air pollution from industrial facilities as an indicator for the exposure to environmental pollution. Therefore, these issues are addressed in the following paragraphs.

To reduce the remaining spatial autocorrelation of residuals, a spatial Durbin model was estimated. The Durbin model, additionally, includes the spatially lagged dependent variable $\boldsymbol{Wy}$ on the right-hand side of the equation (e.g. LeSage & Pace, 2009). This method reduces the spatial correlation of the residuals dramatically. Regarding the effect magnitudes, the spatial Durbin models yield similar direct but slightly higher spillover effects.[6] This applies to the overall as well as the fixed effects Durbin model. Additionally, overall and fixed-effects Durbin models confirm that direct and spillover effects are significantly stronger in urban areas. In sum, the spatial Durbin models offer very similar results to the SLX models, while reducing the remaining residual autocorrelation.

Regarding the specification of the neighbourhood weights matrix, additional models were estimated with alternative specifications. When excluding cases without direct neighbours instead of imputing these connections, nearly identical results are obtained. When replacing the weights matrix by a 10 nearest neighbours matrix (while weighting the neighbours by inverse distance), direct effect magnitudes decrease while the spillover effects increase. However, the main story does not change. Another model specification includes first as well as second order neighbours separately (two different weights matrices). These models produce nearly identical direct effects but slightly smaller spillover effects for the first order neighbours. Interestingly, also the second order neighbours produce effects comparable to first order neighbours with particularly strong impacts in urban areas. Combining first and second order effects, these models estimate an even stronger spillover effect from neighbouring units. This further supports the story that neighbours matter: even the neighbours of the neighbours play a role in determining the level of pollution. All estimated alternatives also support the findings from the community-fixed effects models and the different patterns between urban and rural communities.

To ensure that the conclusions do not depend on the measure of exposure to environmental pollution, Tables 4.4 and 4.5 in the Appendix present the analyses of the results section with an alternative measure, the proximity to the nearest facility. In contrast to the first indicator, this measure additionally includes waste management facilities (like landfill sites). These analyses basically confirm the previous results. Census grid cells with a high minority share are located closer to industrial facilities and, again, neighbours matter: a high minority share of neighbouring units is associated with closer industrial facilities. Though the relative effect magnitude of the spillover effects compared to the direct effects is lower when using the proximity, these analyses confirm that the total effect of neighbouring units is stronger. Furthermore, the

[6] Note that the impacts in a Durbin model depend on the weights matrix and differ from the model's coefficients. I refer to the average direct and indirect effects over all units here (for more information see e.g. Halleck Vega & Elhorst, 2015; LeSage & Pace, 2009).

results confirm the conclusion regarding the area differences. The correlation between minority share and proximity to industrial facilities is stronger in urban areas, especially regarding the spatial spillover effects. Hence, neighbourhoods with a high minority share and neighbourhoods lying in an area with a high minority share are not only exposed to a higher amount of air pollution but are also located closer to industrial and waste management facilities. However, it must be noted that both measures only use indicators of industrial pollution and do not account for other sources of pollution like traffic.

4.7 Conclusion

Although environmental inequality has gained increasing interest in Europe, this is the first nation-wide wide assessment of environmental inequality in Germany that uses objectively measured pollution. Furthermore, the present study includes the spatial nature of environmental inequality patterns by using innovative spatial methods and including the characteristics of neighbouring units. This seems to be important as the theoretical mechanisms of environmental inequality concern not only single units but rather spatial clusters. In sum, the analyses yield three important results.

First, there is a considerably high correlation between the minority share and exposure to environmental pollution. Not only in a nation-wide comparison, but also within communities, minorities experience a higher exposure to environmental pollution. This means that minorities live in more polluted areas in Germany and in more polluted neighbourhoods within communities. Housing-related control variables do a poor job in explaining the correlation between minority share and pollution, indicating that the 'racial income-inequality thesis' finds no support. It rather indicates that the high exposure of minorities does not mainly stem from their low socio-economic status and the resulting need for low-cost housing. However, to adequately confirm this finding, panel data or natural experiments are needed to account for other non-observable confounders. With certainty, it can only be stated that neighbourhoods and areas with a high level of minorities are exposed to disproportionately high levels of pollution.

Second, neighbours matter, which means that neighbourhoods that are located within clusters of high minority neighbourhoods face a higher level of pollution, independent of their own characteristics. Thus, processes of spatial clustering seem to play an important role in shaping patterns of environmental inequality. One explanation could be that regions where minorities agglomerate offer attractive sites for industrial facilities as the probability of resistance declines with distance rather than boundaries. Another explanation could be that segregation of minorities around industrial facilities contributes to their disproportionate exposure to pollution. In combination with further attraction of minority households due to similarity preferences, this could induce a self-reinforcing processes leading to a clustering of minorities around industrial facilities. This is especially interesting as results from the United Stated (Downey, 2007) rather indicate that segregation patterns do not play a crucial role in determining the extent of environmental inequality. If this also applies to Germany, then *why* do we observe these clustering processes? To answer this question empirically, further research should have a more detailed look at the spatial patterns of environmental inequality. Especially in a time-series framework this could lead to new insights regarding the causes or structural conditions generating the disproportionate exposure of minority households to industrial pollution.

Third, the correlation between the minority share and environmental pollution is stronger within urban than within rural areas. Especially the spatial spillover effects stemming from neighbouring units are much stronger within urban communities. Urban infrastructures, the level of segregation, or urban housing opportunities may foster the extent of environmental inequality. On the one hand, minorities might be less clustered or segregated in rural areas, which prevents companies from selectively siting facilities close to minorities. Spatial restriction in urban areas, in contrast, could 'force' companies to place facilities closer to inhabitants, which might then disproportionately affect minorities. On the other hand, rural areas exhibit a lower population density and more relaxed housing markets, which might provide better opportunities to 'escape' polluted areas. This is also supported by the finding that higher vacancy rates are correlated with pollution within rural areas (but not within urban areas): people seem to leave polluted neighbourhoods in rural areas. Especially minority households might, in turn, profit from those relaxed housing situations, leading to lower barriers regarding discriminating behaviour as well as price pressure.

In sum, the present study offers a first assessment of environmental inequality in Germany using objective measures of air pollution, but also highlights the importance of spatial processes associated with environmental inequality. However, the results also open new issues for further research. This study estimates the exposure to environmental pollution by proportionately allocating the toxicity-weighted emissions of facilities to surrounding neighbourhoods. Though this is an improvement over other studies in Germany using subjective measures, it is far behind the accuracy provided by EPA's RSEI in the United States. Further effort should be undertaken to improve the data quality in Europe, including advanced models based on estimated surrogate doses of exposure as it is done by the RSEI. This would not only improve the quality of environmental inequality research in Europe but also offer the opportunity to conduct comparative studies between different countries, thereby providing new insights for international research in this area.

Further research should also investigate the question of why neighbours matter. Though the discussion above offers some possible explanations, further research is needed to explain the patterns observed in this study. The strong difference between urban and rural areas indicates that structural characteristics might play an important role. Thus further research should try to enrich the analyses by other structural variables and investigate their influence on the clustering processes presented here. Geographic data on urban forms as provided by OpenStreetMap, for example, would offer an interesting extension of socio-demographic characteristics usually used in environmental inequality research. Moreover, combining the spatial patterns of environmental inequality with individual mobility data might offer important insights of the causal mechanisms leading to the disproportionate exposure of minorities to environmental pollution.

4.A Appendix Chapter 4

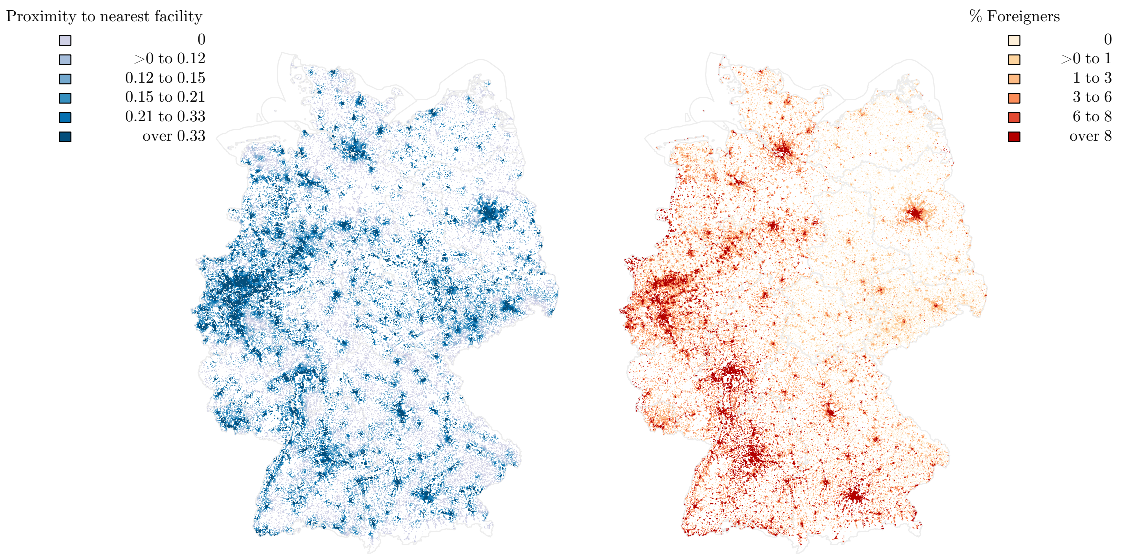

Figure 4.3: Spatial distribution of proximity to nearest facility and minority share in Germany

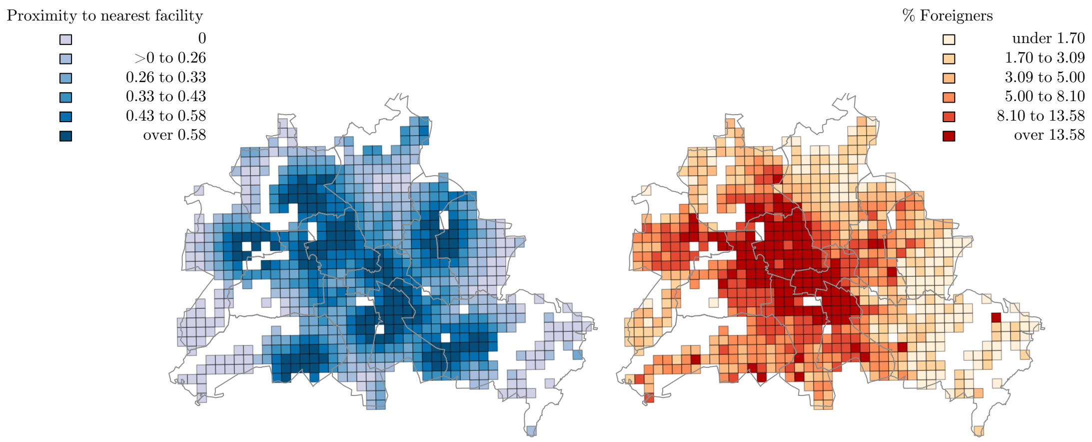

Figure 4.4: Spatial distribution of proximity to nearest facility and minority share in Berlin

Table 4.4: OLS and SLX estimates. Dependent variable: proximity to nearest facility

	OLS	SLX		
	M1	M2	M3	M4
(Intercept)	0.000	-0.007^{*}	-0.049^{***}	-0.025^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
% Foreigners	0.320^{***}	0.155^{***}	0.119^{***}	0.076^{***}
	(0.003)	(0.004)	(0.004)	(0.005)
W [% Foreigners]		0.282^{***}	0.165^{***}	0.089^{***}
		(0.005)	(0.006)	(0.006)
Population			-0.035^{***}	0.061^{***}
			(0.005)	(0.006)
W [Population]			0.180^{***}	0.372^{***}
			(0.007)	(0.009)
% 65 and older			0.014^{***}	0.005
			(0.003)	(0.003)
W [% 65 and older]			0.021^{***}	0.004
			(0.004)	(0.004)
% Vacant housing			0.016^{***}	0.006
			(0.003)	(0.003)
W [% Vacant housing]			0.004	-0.012^{**}
			(0.004)	(0.004)
Living space			-0.039^{***}	-0.025^{***}
			(0.003)	(0.003)
W [Living space]			-0.069^{***}	-0.046^{***}
			(0.004)	(0.004)
Urban			0.378^{***}	0.459^{***}
			(0.012)	(0.015)
Urban × % Foreigners				0.070^{***}
				(0.012)
Urban × W [% Foreigners]				0.142^{***}
				(0.016)
R^2	0.103	0.131	0.173	0.193
Adj. R^2	0.102	0.131	0.173	0.193
N	93777	93777	93777	93777

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standardised coefficients. Standard errors in parentheses. W is specified as contiguity neighbours weights matrix. Model M4 interacts all covariates with the urban dummy (interaction with controls not shown).

Table 4.5: Community-fixed effects estimates. Dependent variable: proximity to nearest facility

	Overall	Urban	Rural	Diff
	M5	M6	M7	(M6-M7)
% Foreigners	0.132^{***}	0.153^{***}	0.088^{***}	0.065^{*}
	(0.009)	(0.025)	(0.008)	(0.027)
W [% Foreigners]	0.216^{***}	0.357^{***}	0.122^{***}	0.236^{***}
	(0.018)	(0.036)	(0.013)	(0.038)
Population	-0.044^{***}	-0.086^{***}	0.017^{*}	-0.102^{***}
	(0.009)	(0.010)	(0.007)	(0.012)
W [Population]	0.079^{***}	-0.013	0.169^{***}	-0.182^{***}
	(0.013)	(0.024)	(0.016)	(0.029)
% 65 and older	0.005	-0.011	0.004	-0.015
	(0.002)	(0.012)	(0.002)	(0.012)
W [% 65 and older]	0.007	0.018	0.006	0.012
	(0.004)	(0.044)	(0.004)	(0.045)
% Vacant housing	0.017^{***}	0.035^{*}	0.011^{***}	0.024
	(0.003)	(0.017)	(0.003)	(0.017)
W [% Vacant housing]	0.009	0.036	0.003	0.033
	(0.005)	(0.053)	(0.005)	(0.054)
Living space	-0.035^{***}	-0.086^{***}	-0.018^{***}	-0.067^{**}
	(0.004)	(0.023)	(0.003)	(0.023)
W [Living space]	-0.063^{***}	-0.358^{***}	-0.032^{***}	-0.325^{***}
	(0.007)	(0.052)	(0.004)	(0.053)
R^2	0.069	0.167	0.044	
Adj. R^2	0.066	0.166	0.042	
N	93777	9061	84716	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standardised coefficients. Cluster robust standard errors in parentheses. W is specified as contiguity neighbours weights matrix. The differences and standard errors in column 4 are obtained by an overall fixed-effects model with interaction terms of the urban dummy and all covariates.

Chapter 5

Bringing Urban Space Back in: A Multi-Level Analysis of Environmental Inequality in Germany

Abstract

Various studies have shown that minorities bear a disproportionate exposure to environmental pollution. To understand the causes of this environmental inequality, it is important to analyse which structural conditions foster environmental inequality. This study uses an original dataset by combining the German 2011 census with georeferenced pollution data to analyse the variation of environmental inequality between German cities. While structural characteristics derived from standard theories of environmental inequality do a rather poor job of explaining regional differences, an overlooked indicator correlates strongly with environmental inequality: the geographic centrality of polluting facilities within the urban space. Including this structural measure into the city-fixed effects multilevel analysis accounts for more than 25% of the variation between cities. This highlights the importance of taking geographic conditions into account when analysing environmental inequality.

This chapter has been published as: Bringing urban space back in: A multilevel analysis of environmental inequality in Germany. *Urban Studies*. DOI: 10.1177/0042098018795786

5.1 Introduction

Scholars agree that environmental inequality exists. A huge body of research in the United States and also a growing body of research in Europe has shown that ethnic minorities are exposed to a disproportionately high amount of environmental pollution (e.g Ard, 2015; Ash & Fetter, 2004; Diekmann & Meyer, 2010; Downey & Hawkins, 2008; Laurian & Funderburg, 2014; Padilla et al., 2014; Pastor et al., 2005, 2001; Rüttenauer, 2018). However, some studies show that the level of environmental inequality varies considerably between different cities (Downey, 2007; Downey et al., 2008; Padilla et al., 2014), which offers an opportunity to analyse the structural conditions that foster environmental inequality. As has been argued by Schweitzer and Stephenson (2007), comparative research is needed to investigate the influence of urban conditions on the distribution of environmental hazards across social groups. Identifying decisive urban conditions helps to improve our understanding of the underlying causal mechanisms, and, in turn, may help urban planners to avoid the persisting disadvantage of minorities.

So far, systematic analyses of the varying inequality levels have been constrained to the United States. Still, it is unclear to which extent findings from the United States are transferable to Germany. First, environmental inequality in the United States is mostly concerned with ethnic minorities like Black Americans, Hispanics, or Asians. In contrast, minorities in Germany are mostly immigrant-minorities and stem from relatively recent immigration from other European countries or Turkey. Second, the United States exhibits much higher levels of residential segregation: while dissimilarity indices in the United States at the beginning of the century varied between 41 and 64 (depending on the ethnic group), Germany exhibits segregation indices around 20 (Musterd, 2005). Third, the urban structures differ between the United States and Europe, with a higher population density and more centrally organised cities in Europe (Huang et al., 2007). All those characteristics might play an important role for facility siting decisions or residential mobility, and therefore could on the one hand affect the extent of environmental inequality, but on the other hand also play an important role regarding the causes of environmental inequality. Thus, it is important to investigate the structural conditions that foster environmental inequality in a different context.

The current study addresses this aim by analysing a novel dataset which combines pollution data from the European Pollutant Release and Transfer Register with socio-demographic data from 2011 German census on the level of 1 square km grid cells and information on the level of German cities. In a first step, city-fixed random-slope multilevel models show that the level of environmental inequality varies considerably between German cities. In the second step, I add cross-level interactions to investigate whether the effect-heterogeneity between the cities can be explained by structural differences between the cities. The findings show that structural characteristics derived from the standard strand of reasoning in environmental inequality literature – like residential segregation – do a relatively poor job of explaining inter-city differences. This is somewhat surprising, but in line with previous research from the United States. Still, further analyses emphasise the importance of taking the urban landscape into account. Including the average centrality of industrial facilities explains more than 25% of the inter-city variation in environmental inequality. This result challenges the importance of the standard theory in environmental inequality research, as residential segregation only plays a minor role in determining environmental inequality. Conversely, the analysis highlights the importance of the geographic location of polluters within the urban space.

Though many studies have shown that selective siting and selective migration processes contribute to environmental inequality outcomes (Banzhaf & Walsh, 2008; Crowder & Downey, 2010; Pastor et al., 2001; Saha & Mohai, 2005), it is less clear how important those processes are in producing the correlation between environmental pollution and minority households observed at the macro-level. Besides those social processes, only few studies have explicitly included characteristics of urban infrastructure and form when investigating environmental inequality (e.g. Baden & Coursey, 2002; Elliott & Frickel, 2015; Wolverton, 2012), showing for example that riversides, infrastructure, or agglomeration economies play an important role in determining facility locations. Similarly, the results of the present study show that a considerable proportion of environmental inequality relies on the centrality of industrial facilities. This demonstrates the importance of considering the urban space, where social processes like siting or migration take place. Doing so can enhance our knowledge about the causes of environmental inequality, but can also offer new insights into the side products of urban structures and urban planning.

5.2 Theoretical background

Previous literature on the causes of environmental inequality has mostly concentrated on two possible explanations for the disproportionate exposure of minority households to environmental pollution: selective siting and selective migration (for an overview see Mohai & Saha, 2015a).

The argument of selective siting assumes that polluting facilities temporally follow the settlement of minority or economically disadvantaged households. The reason that polluting facilities choose areas predominantly occupied by minorities may be threefold: taste-based discrimination, profit maximisation, and the avoidance of political protest (Campbell, Kim & Eckerd, 2015; Grant, Trautner, Downey & Thiebaud, 2010; Hamilton, 1995). First, decision makers may predominantly belong to the majority group and choose to locate unwanted industrial facilities close to minority households and far away from their own group. Second, housing and land-use prices may be lower around high-minority areas, as minority households exhibit generally lower incomes and consequently live in low-rent areas. If companies were economically rational, they would choose low-cost amenities and – as a side-effect – locate close to the minority population. Third, companies may seek the 'path of the least resistance' (Saha & Mohai, 2005) and, thus, choose locations where political protest is assumed to be low. If decision-makers within companies assume that minority groups are less likely to organise collective actions against facility siting ('not in my backyard') due to their limited resources and political efficacy, minority areas may provide an attractive siting location.

Several studies have used longitudinal data to test the argument of selective siting. By doing so on the spatially aggregated level, Funderburg and Laurian (2015); Mohai and Saha (2015b); Pastor et al. (2001); Saha and Mohai (2005); Shaikh and Loomis (1999) conclude that demographic disparities in facility-hosting areas already existed prior to the siting process, which supports the argument of selective siting. However, other longitudinal studies do not exhibit a consistent association between the demographic composition of an area and the probability of receiving a new industrial facility (Been & Gupta, 1997; Downey, 2005; Oakes et al., 1996), and still others rather emphasise infrastructural and historical patterns of facility siting (Baden & Coursey, 2002; Elliott & Frickel, 2013; Wolverton, 2012). Furthermore, agent-based models show that siting decisions alone cannot explain high levels of environmental inequality (Campbell et al., 2015). Thus, empirical results remain mixed regarding the role of selective siting.

In contrast to selective siting, the argument of selective migration assumes that the in-flow of minority residents temporally follows the occurrence of polluting facilities. Again, two alternative reasons for this causal pathway exist: the 'racial residential discrimination thesis' and the 'racial income-inequality thesis' (Best & Rüttenauer, 2018; Crowder & Downey, 2010; Pais et al., 2014). The 'racial residential discrimination thesis' posits that ethnic minorities are discriminated against in the housing market. Independent of socio-economic characteristics, ethnic minorities are steered into neighbourhoods with a lower environmental quality because housing agents or landlords may fear declining attractiveness due to minority in-migration, or spuriously anticipate lower preferences for environmental quality (e.g. Turner & Ross, 2005). The 'racial income-inequality thesis', in contrast, argues that selective migration of minority households is a result of economic disparities between ethnic groups. Assuming that households prefer a high environmental quality over a low environmental quality, it follows that demand as well as rents and housing prices are higher in clean neighbourhoods. At the same time, the willingness or ability to pay for environmental quality crucially depends on income. Thus, rich households sort into high-quality neighbourhoods, while poor households will end up in less desirable and more polluted neighbourhoods (e.g. Banzhaf & Walsh, 2008; Sieg et al., 2004). As minority households hold lower economic resources than their majority counterparts do, those households sort into low-quality neighbourhoods because they simply cannot afford rents and housing prices in high-quality neighbourhoods.

When analysing the aggregated socio-economic development of an area after hosting a new facility, most existing research does not support selective migration as a cause of environmental inequality (e.g. Downey, 2005; Funderburg & Laurian, 2015; Mohai & Saha, 2015b; Oakes et al., 1996; Pastor et al., 2005, 2001). However, studies using individual data and moving trajectories indeed find evidence in favour of the selective migration argument. Best and Rüttenauer (2018), Crowder and Downey (2010), and Pais et al. (2014) conclude that minority households selectively move into more polluted areas than majority households do. Even when controlling for income, these disparities in moving behaviour persist, which indicates that selective migration is not mainly driven by income. However, agent-based models of Campbell et al. (2015) indicate that the finding of minorities selectively moving into low-quality areas may be a result of similarity preferences rather than the tendency of choosing lower quality neighbourhoods. Thus, it is unclear how decisive those selective migration processes are for the observed extent of environmental inequality.

5.3 Macro-structural predictors

Regarding the extent of environmental inequality, previous studies have shown that some cities exhibit a high level of environmental inequality, while others show none or only low levels (Downey, 2007; Downey et al., 2008; Padilla et al., 2014). To explain these variations, this section first describes hypotheses that are derived from the standard strand of reasoning and subsequently develops a novel hypothesis that has not been tested by previous research.

The theory of selective siting highlights two factors that should explain the level of environmental inequality on the macro level: political efficacy and ethnic residential segregation. First, assuming that companies choose facility locations based on the level of expected political resistance, higher political efficacy of majority members should increase the 'externalization' of pollution onto minority groups. Second, independent from the actual motives of a company's decision, decision makers should only be able to place facilities disproportionately close to minority residents, if residential segregation exists. In cities where minority residents are evenly distributed in space (absolutely desegregated), there is no opportunity for companies to discriminate against minority households and to choose a facility location that is disproportionately close to minorities. Thus, higher residential segregation should increase the possibility of selective siting and consequently increase the level of environmental inequality.

The same argument applies for selective migration. An absolutely desegregated area makes selective migration implausible as an explanation for environmental inequality. In a desegregated city, there is no possibility that minority households would have moved selectively into polluted areas in the past, while majority households would not. If selective migration were at work, minorities would live segregated in areas closer to industrial facilities as a consequence of selective migration. Thus, selective migration processes should produce higher levels of segregation. Second, according to the 'racial income-inequality thesis', economic disparities should be the main reason for selective migration processes. Thus, higher levels of economic inequality within a city should lead to higher levels of environmental inequality.

In sum, the standard strand of reasoning in environmental inequality research leads to the following hypotheses: the level of environmental inequality increases with

H1) increasing levels of residential segregation,
H2) increasing political efficacy of the majority group,
H3) increasing economic inequality.

However, previous research produced rather inconclusive results regarding those hypotheses. Comparing the level of environmental inequality in U.S. metropolitan areas, Downey et al. (2008) find mixed results regarding the role of income inequality and spatial segregation. Though both factors have a significant effect in some models, the results vary with respect to the operationalisation of the dependent variable and the ethnic group under consideration. Similarly, Downey (2007) finds a significant association between residential segregation and the ethnic toxic concentration ratio, but the effect is small in magnitude and the explanatory power of the model only marginal. In contrast, a recent comparison of segregation measures by Ard (2016) shows that – for most segregation measures – increasing segregation leads to increasing health risks especially for minority residents. Still, it is difficult to interpret strength and explanatory power in this study. As noted by Downey (2007), it seems odd that segregation has no distinct impact on the level of environmental inequality given that it seems to be an important condition for the theories of selective siting as well as selective migration to work.

One explanation could be that the spatial proximity of industrial facilities is mainly driven by labour-market related characteristics. For example, Hersh (1995) finds that working class households are located closer to facilities because of attractive job opportunities. Similarly, Been and Gupta (1997), and Wolverton (2012) identify the percentage working in manufacturing as an important driver of facility siting. On the one hand, industrial workers offer an attractive labour force for companies; on the other hand, people employed in industry may trade-off environmental quality for attractive jobs (Wolverton, 2009). If minorities are, at the same time, overrepresented in the class of manufacturing or industrial workers, this will lead to a disproportionate burden of minorities. Thus, the extent of environmental inequality might also depend on the share of people working in the industrial sector. In line with this hypothesis, Krieg (1995) identifies a higher correlation between minority share and toxic waste sites in older industrialised towns, which leads to the hypothesis that the level of environmental inequality increases with

H4) an increasing share of people employed in the industrial sector.

Another explanation is that studies of environmental inequality 'fail to take the spatial distribution of environmental hazards within metropolitan areas into account' (Downey, 2007, p. 970). More precisely, the results of Downey (2005) in Detroit point out that ethnic minorities were somewhat separated from (new) polluting facilities because they lived – due to high residential segregation – in areas where industrial facilities decreased. While the Black population resided segregated in central areas of the city, new facilities emerged in suburban and predominately White areas. This process 'prevented' the Black population from living in areas of newly emerging facilities and, thus, led to the conclusion that facilities were not sited selectively in Black neighbourhoods between 1970 and 1990.

Elliott and Frickel (2015) support this implication in a historical analysis of hazardous industrial sites in four U.S. cities, finding a persisting geographical accumulation of industrial sites around the urban core. At the same time, they conclude that the extent of environmental inequality diminished over time due to the increasing churning of White middle-class households into those central areas, where minorities have been overrepresented so far. Also for NO_2 concentration levels (including also mobile sources), Padilla et al. (2014) find a positive correlation between deprivation and pollution only in those cities with high deprivation levels in the inner city.

It is well known from segregation research (e.g. Massey & Denton, 1989) that minorities generally tend to cluster around the central city districts. As has been stated above, this can also prevent minorities from being exposed to environmental pollution if this pollution occurs in more peripheral areas. Altogether, this suggests that, when taking the spatial distribution of pollution into account, especially the centrality of pollution is a crucial driver of environmental inequality. If facilities are located close to the city centre – where minority groups usually cluster – we expect to observe a high correlation between minority share and environmental pollution. This leads to the last (and so far overlooked) hypothesis that the level of environmental inequality increases with

H5) *increasing centrality of facilities within the urban space.*

Though previous results point towards the importance of the spatial distribution of pollution, this hypothesis has not been explicitly tested in previous research.

5.4 Analytical strategy

5.4.1 Data

The European Pollutant Transfer and Release Register (E-PRTR; European Commission, 2006) is a register capturing industrial activities and emission reports on all German facilities that fall under one or more of the 65 E-PRTR activities (European Commission, 2006, pp. 79-82) and exceed a pollutant specific threshold of emissions (European Commission, 2006, pp. 83-86). The regulations of this pollutant register are similar to the U.S. Toxics Release Inventory (TRI). All facilities exceeding a pollutant-specific threshold have to report their emissions and geo-locations.[1] In 2011, the dataset contains a total of 1,479 facilities reporting nearly 1b tonnes of emissions to air and nearly 90m tonnes of processed waste. Of those facilities, 367 (24.81%) are located within a 2 kilometre buffer around the 79 metropolitan areas and report emissions to air. Most of the air-polluting facilities in metropolitan areas operate in the energy and waste management sector (both 28%), metal production, chemical industry (both 10%), or the mineral industry (9%).

The 2011 German census (Statistische Ämter des Bundes und der Länder, 2015) provides information about the German population on the level of 361,478 equally distributed 1 kilometre grid cells. Reducing this dataset to non-missing observations (excluding cases with none or low number of inhabitants) and German metropolitan areas (more than 100,000 inhabitants) results in a *final dataset of 9,061 grid cells clustered within 79 cities*. In total these grid cells capture approx. 24 million inhabitants and contain on average 2,650 (median: 1,717) inhabitants.

In addition, the socio-demographic data of the 2011 German census are further enriched by characteristics of the 79 metropolitan cities provided by the 'Indikatoren und Karten zur Raum- und Stadtentwicklung' database (INKAR; BBSR, 2017). This is done by using not the level of municipalities but the level of districts ('Landkreise') as 66 of the 79 cities are districts by themselves ('Kreisfreie Stadt') and this level provides additional variables not available at the municipality level. Though for 13 of the cities, district-level characteristics also include sub-urban regions, additional analyses excluding these cities yield nearly identical results.

5.4.2 Variables

The main variable of interest is the amount of air pollution for each census grid cell. Therefore, I created a 2 kilometre buffer around the georeferenced location of each E-PRTR facility and allocated the reported emissions to air proportionate to the spatial

[1] Note that information in the E-PRTR is self-reported. To control for reporting biases, a model using 3-years average emissions was estimated, producing similar results.

overlap between the buffer and each census grid cell. This method is similar to the method Banzhaf and Walsh (2008) apply to the U.S. TRI data. This leads to a dataset of 9,061 census grid cells containing the proportionate amount of industrial emissions to air (including greenhouse gases, chlorinated substances, heavy metals, and other gases and organic substances), without weighting them by toxicity. The reason for not weighting the pollution by toxicity is that using EPA's RSEI toxicity weights reduces the effective number of facilities within the cities to 179 with only 49 cities holding at least one facility. However, additional analyses using toxicity weights lead to similar results with the same conclusions, no matter whether in- or excluding cities without any toxic facilities (see Figure 5.3 of the Appendix). To take the right skewness of the emissions into account, the natural logarithm of the emissions (+1) is calculated.

The main explanatory variable on the census grid cell level is the minority share. In the German census, this is available as the percentage of foreigners (not holding German citizenship) in each census cell. Additional control variables on the grid cell level include population (which equals population density because of same-sized spatial units), the percentage of people aged 65 or older, the percentage of vacant housing, and the living space per inhabitant (in m^2). The latter can be seen as a proxy for wealth especially within cities. Furthermore, I control for infrastructural characteristics by including the distance to highway junctions and railway lines (e.g. Wolverton, 2012). Those infrastructural characteristics are obtained from OpenStreetMap.

To investigate the differences between German cities, I use the following city-level indicators: residential segregation, economic inequality, political efficacy, the share of employees in industry, and facility centrality. As a measure for segregation, the spatial information theory index ($\tilde{H}$) as proposed by Reardon and O'Sullivan (2004) is computed for each city. The spatial version $\tilde{H}$ is calculated by taking the average proximity-weighted local neighbourhood of each point into account. In the main analysis, a kernel density bandwidth of 2,000m is used to compute the local environment. Economic inequality is operationalised by the German – foreign unemployment ratio (percentage of unemployed foreigners divided by the percentage of unemployed Germans). Political efficacy is measured as the total voter turnout of each city in the preceding federal election (2009), and the share of industrial workers as the percentage of employees in the industrial sector. In a last step, further controls on the city level are added. These controls include the average gross wages per capita in the industrial sector, the city size (area), and population growth rate per square kilometre within the last ten years.

To calculate the facility centrality in each city, I geo-coded each city's town hall as the city centre.[2] Subsequently, I computed the distance of each industrial facility within a 2 kilometre range of the city boundaries to the city centre. In the end, the

[2] Manual inspection of the 79 cities confirms this strategy to be adequate.

average of these distances was taken and divided by the maximum possible distance between the city centre and each point of the city boundary. Dividing the average distance by the maximum distance ensures that the centrality measure does not depend on the city's actual size but becomes a relative measure. For ease of interpretation, I calculated the inverse of this average relative distance to receive the average centrality of industrial facilities.[3] Formally, the facility centrality for each city $i = 1, ..., N$ can be written as

$$FC_i = \left(\frac{\frac{1}{M} \sum_{j=1}^{M} d_{ij}}{max(\tilde{\boldsymbol{d}}_i)} \right)^{-1},$$

where d_{ij} is the distance between each facility $j = 1, ..., M$ in the 2km surrounding of city i and the city's centre, and $\tilde{\boldsymbol{d}}_i$ a vector of the distances between the city centre and all coordinates of the city's boundary. The summary statistics of all variables are presented in Table 5.3 of the Appendix.

5.4.3 Method

The analyses rely on multilevel models, nesting the 9,061 census grid cells within the 79 cities (e.g. Hox & Wijngaards-de Meij, 2015). The main point of interest is the comparison of within-city environmental inequality between the cities. As in previous literature (e.g. Ash & Fetter, 2004; Diekmann & Meyer, 2010; Downey & Hawkins, 2008; Rüttenauer, 2018), the air pollution is regressed on the percentage of foreigners, and this regression coefficient is interpreted as the level of environmental inequality. To obtain a within-city effect, the first level (census grid cell) variables are group mean centered (for a detailed discussion on centering methods and implications see Enders & Tofighi, 2007). This is essentially similar to city-fixed effects estimators (including a dummy for each single city), which means that all city-wide characteristics are controlled for in the first-level estimates (e.g. generally higher pollution in larger cities). The model can be written as

$$pollution_{ij} = \beta_{0j} + \beta_1 forgn_{ij} + \beta_2 forgn_{ij} segr_j + ... + u_{1j} forgn_{ij} + \varepsilon_{ij},$$

for all $i = 1, ..., N$ observations and $j = 1, ..., J$ cities, where β_{0j} is a city-specific intercept, u_{1j} the random-slope parameter, and ε_{ij} the idiosyncratic disturbance.

The included cross-level interactions (β_2) provide an estimate of how city-level characteristics (e.g. segregation) influence the within-city correlation between the percentage of foreigners and industrial air pollution (β_1), i.e. the level of environmental inequality. Note that the main effects of the city-level variables are subsumed by the individual intercepts. Furthermore, the random slope parameter can be used to evaluate

[3] This ensures, furthermore, that cities without any facility receive a centrality value of zero, whereas the distance would be infinite.

the impact of the cross-level interactions on the between-city variation. All second level predictors (city level) were standardised around the grand mean and scaled by their standard deviation. Note that this approach does not claim to identify causal effects, as the cross-sectional design does not allow testing selective siting against selective migration processes (e.g. Baden & Coursey, 2002; Pastor et al., 2001). The aim is to analyse whether city-level characteristics are connected to the extent of environmental inequality. However, similar results are obtained when conditioning on facility location, or merging 2012 pollution reports with the 2011 census (see the Tables 5.4 and 5.5 in the Appendix).

5.5 Results

Model M1 in Table 5.1 presents the baseline model, only including the minority share as a predictor. The model confirms that, within German cities, there is a significant and relatively strong correlation between minority share and environmental pollution. Furthermore, this effect varies considerably and significantly between the cities (compared to a model without random slope: $\chi^2(1) = 184.01$, $p < 0.001$). More precisely, the coefficients for the minority share ranges from -0.1557 (Gelsenkirchen) to 0.8567 (Leverkusen).

This confirms that there is a large variation in the level of environmental inequality. While some cities have a negative coefficient, i.e. the majority is actually disproportionately more affected by environmental pollution, others have a very strong positive coefficient pointing towards a strong disadvantage of minority groups. For a first descriptive inspection, Figures 5.1 and 5.2 show the spatial distributions of air pollution and minorities for the most extreme cities (including the point locations of facilities and the city centre). While Figure 5.1 presents cases with the *highest* levels of environmental inequality, Figure 5.2 presents cases with the *lowest* levels of environmental inequality.

When looking at the distribution of the minority population (right panel for each city), minority households seem to experience relatively identical levels of residential segregation, but tend to live closer to industrial facilities in the four cities of Figure 5.1 than in the cities of Figure 5.2. Again in both figures, the minority share is higher around the city centre, meaning that minorities in all eight cases cluster around the city centre – a pattern similar to cities in the U.S. (Massey & Denton, 1989). However, there appears to be an interesting difference between both figures: pollution and facilities tend to be closer to the city centre in cities with high levels of environmental inequality (Figure 5.1). Facilities in cities with low levels of environmental inequality (Figure 5.2), in contrast, are located more distant from the city's centre. This seems to prevent minorities from being disproportionately affected by industrial pollution. *In those*

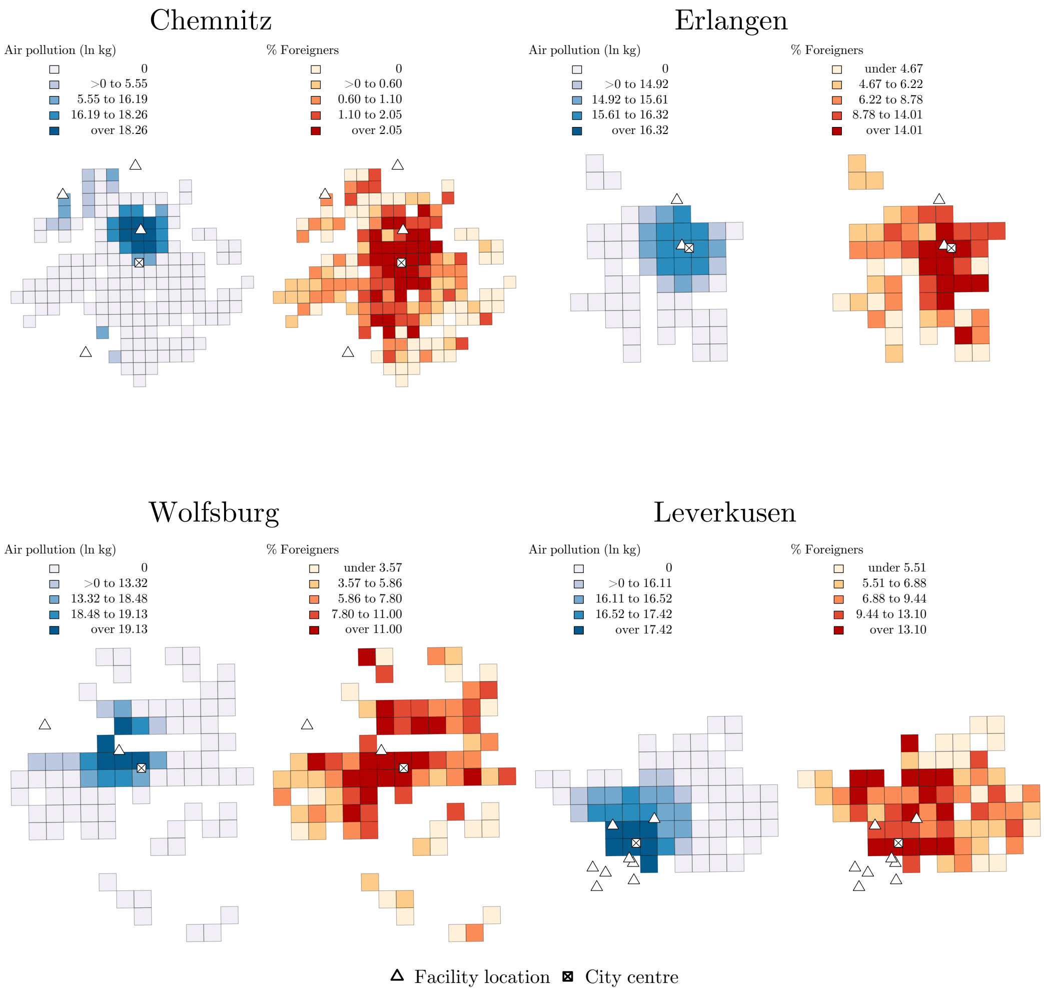

Figure 5.1: Distribution of pollution and minorities for cities with *high* level of environmental inequality

cities, where pollution happens to be strong in the peripheral areas, majority groups seem to be equally or even more affected by pollution than their minority counterparts.

To test the influence of the city-level characteristics statistically, models M2 - M5 of Table 5.1 present the results of the multilevel models including census-level controls and cross-level interactions. Model M2 adds the census-level control variables. Though the coefficient for the minority share is lowered in magnitude by including the controls, it remains highly significant and strong. As expected, the living space per inhabitant – which can be seen as a proxy for wealth in urban areas – is negatively associated with the level of environmental pollution. Furthermore, the distance to highway junctions and railway lines has the expected effect on air pollution. In line with results of Wolverton (2012), pollution is higher in areas with more favourable infrastructural opportunities. Yet, a high level of environmental inequality net of the census-level controls remains.

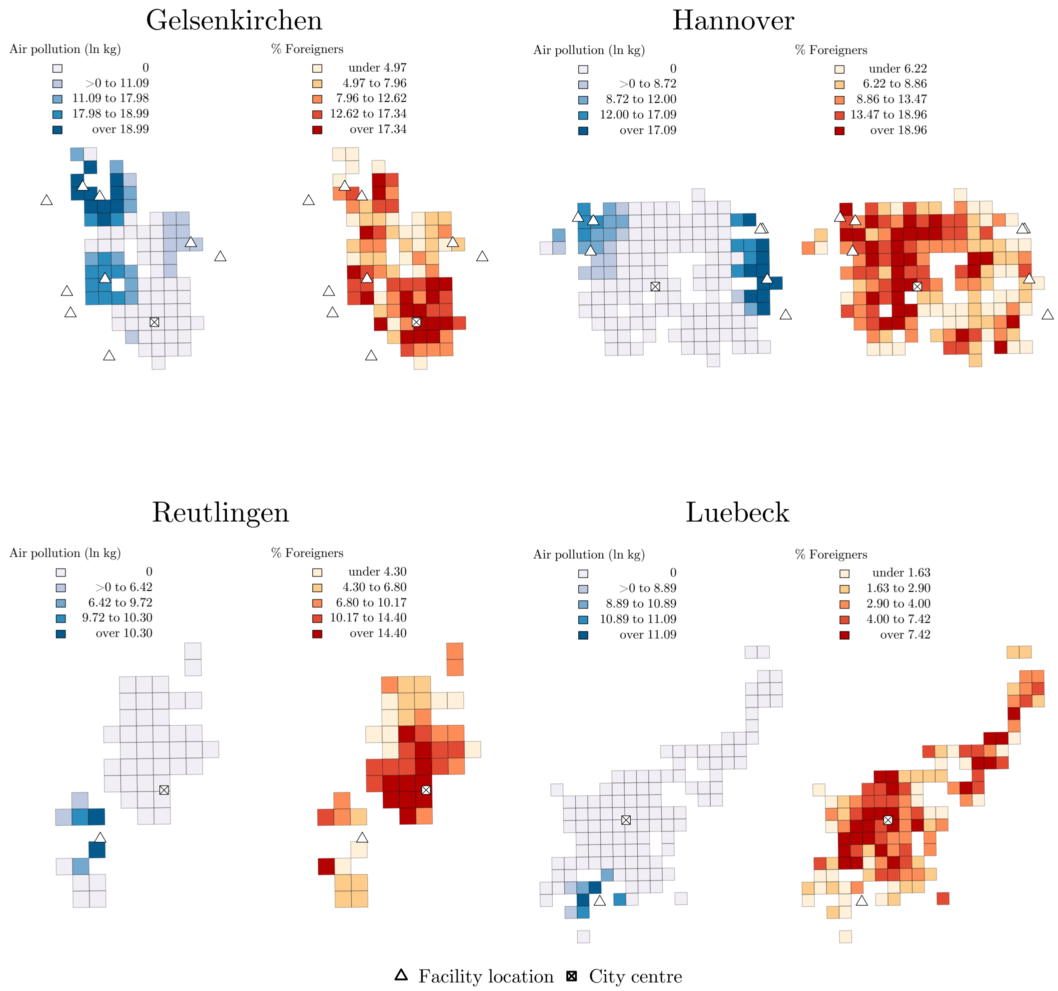

Figure 5.2: Distribution of pollution and minorities for cities with *low* level of environmental inequality

To answer the question of whether structural differences between the cities can explain the differing levels of environmental inequality, models M3 to M5 add the cross-level interactions between structural indicators and the minority share. Model M3, which includes the structural characteristics derived from the standard literature of environmental inequality, presents challenging findings. Neither economic inequality nor political efficacy has a significant influence on the level of environmental inequality within a city. The coefficient regarding economic inequality does not show the assumed direction: higher economic inequality is associated with lower levels of environmental inequality – though far from significant. Residential segregation, in contrast, shows the expected sign: higher residential segregation is associated with higher levels of environmental inequality. However, the effect is only significant at the 10% level and relatively weak in magnitude. Alternative segregation indices like the spatial dissimilarity index ($\tilde{D}$) or different radii for the local environment yield very similar

Table 5.1: City-fixed effects multilevel models. Dependent variable: ln air pollution

	M1	M2	M3	M4	M5
Census cell level					
% Foreigners	0.297***	0.187***	0.187***	0.187***	0.177***
	(0.033)	(0.033)	(0.034)	(0.030)	(0.034)
Population		−0.006	−0.006	−0.005	−0.005
		(0.011)	(0.011)	(0.011)	(0.011)
% 65 and older		−0.003	−0.004	−0.003	−0.003
		(0.010)	(0.010)	(0.010)	(0.010)
% Vacant housing		0.019†	0.020†	0.021*	0.022*
		(0.011)	(0.011)	(0.011)	(0.011)
Living space		−0.052***	−0.052***	−0.052***	−0.052***
		(0.013)	(0.013)	(0.013)	(0.013)
Distance to highway		−0.043***	−0.042***	−0.042***	−0.042***
		(0.011)	(0.011)	(0.011)	(0.011)
Distance to rail		−0.159***	−0.158***	−0.157***	−0.157***
		(0.011)	(0.011)	(0.011)	(0.011)
City level					
% Foreigners × $\tilde{H}_{2000}$			0.078†	0.085*	0.111*
			(0.043)	(0.037)	(0.048)
% Foreigners × Unemployment ratio			−0.003	−0.034	−0.034
			(0.029)	(0.027)	(0.029)
% Foreigners × Voter turnout			0.047	0.030	0.028
			(0.034)	(0.031)	(0.034)
% Foreigners × % Employed in industry			0.086**	0.066**	0.050
			(0.027)	(0.024)	(0.031)
% Foreigners × Facility centrality				0.113***	0.111***
				(0.025)	(0.025)
% Foreigners × Wages in production					0.020
					(0.036)
% Foreigners × Size					−0.051
					(0.059)
% Foreigners × Growth rate					0.023
					(0.032)
Fixed effects	yes	yes	yes	yes	yes
Random slope	yes	yes	yes	yes	yes
AIC	23656	23399	23452	23450	23496
N	9061	9061	9061	9061	9061
N cluster	79	79	79	79	79
σ^2 % Foreigners	0.065	0.059	0.052	0.038	0.039
σ^2 Residual	0.784	0.759	0.759	0.759	0.759

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$. Multilevel models with group centered first level variables. All variables are scaled by their standard deviation. Standard errors in parentheses.

results (not shown). Only the share of industrial workers yields a significant influence on the level of environmental inequality: the more people employed in industry, the higher the level of environmental inequality. This is in line with the hypothesis that minorities are more affected by pollution, because industrial facilities provide attractive employment opportunities for the working class. Still, adding those cross-level interactions reduces only a small proportion of the slope variance (around 12%). Note that segregation, unemployment ratio, and voter turnout alone would actually increase the variance. Thus, the structural characteristics derived from the standard strand of theoretical reasoning do not explain the varying levels of environmental inequality at all. Though surprising from a theoretical point of view, those findings reflect the results obtained by Downey (2007) in the U.S.: in line with the present study, economic inequality has an insignificant and in some models negative effect on environmental inequality, and residential segregation does an unexpectedly poor job of explaining environmental inequality.

Model M4 adds the measure of facility centrality as a cross-level interaction to test the influence of the geographic distribution of polluters. In line with the impressions from Figures 5.1 and 5.2, the centrality of facilities within a city shows a highly significant and positive association with environmental inequality. The higher the average centrality of facilities, the stronger the correlation between minority share and pollution. While a one standard deviation increase in minority share corresponds to a 0.187 standard deviation increase in pollution at the average value of facility centrality, this effect increases by 60% to 0.3 ($0.187 + 0.113$) when the facility centrality lies one standard deviation above the mean. Minorities are disproportionately affected by industrial pollution especially in cities with centrally located facilities. Another interesting finding of model M4 concerns the effect of spatial segregation. Once facility centrality is controlled for, the coefficient for residential segregation is significant at the 5% level. In addition, including the facility centrality explains a large proportion of the random slope variance. When comparing models M4 and M3, more than *one quarter of the slope variance* ($(0.052 - 0.038)/0.052 = 0.27$) can be explained due to the centrality of industrial facilities; more than due to all other variables combined. This highlights the importance of the facilities' location or, more precisely, centrality within the urban space for the extent of environmental inequality.

Model M5 adds further cross-level controls. Though none of those exhibit a significant effect, they have some influence on other effects of interest. First, the segregation effect increases in its magnitude and now equals the the facility centrality effects regarding its magnitude. This is mainly due to controlling for the city size. Second, the effect of the share employed in the industrial sector is lower in its magnitude and does not reach significance in model M5. This is mainly due to the collinearity between the share of industrial workers and wages in the industrial sector. Still even when adding further controls, the centrality of industrial facilities exhibits a highly significant effect on the extent of environmental inequality.

Note that the findings are robust against a variety of model specifications, like excluding cities with extreme values, restricting the data to observations with at least 200 inhabitants, using an absolute instead of a relative centrality index, estimating separate models for each hypothesis, using the proximity to air polluting facilities within the city as dependent variable, or including spatial spillover effects. Also when applying RSEI inhalation toxicity weights and thereby excluding greenhouse gases from the analysis, similar results are obtained regarding the centrality of facilities. Residential segregation and percentage employed in industry, in contrast, exhibit no significant effect on the level of environmental inequality when using toxicity-weighted pollution (see Figure 5.3 of the Appendix). Furthermore, additional spatial models (see Table 5.7 in the Appendix) reveal that the centrality of facilities does not only increase the general burden of minorities but also the clustering effect observed by Rüttenauer (2018).

5.6 Discussion

Does this mean that environmental inequality is not mainly driven by selective siting or selective migration processes, but rather a result of the urban structure? It is important to keep in mind that this is a cross-sectional finding, making it hard to interpret why the centrality of facilities is so important. In my opinion, there are two ways of interpreting the results: the centrality of facilities might either be a confounding or a mediating mechanism.

First, the impacts of facility centrality may be the result of two independent processes. Minorities cluster around the city centre because of reasons other than environmental quality (e.g. infrastructure, similarity preferences, etc.). At the same time, industrial air pollution occurs around the urban core because of infrastructural or historical reasons. Historical analyses of Portland by Elliott and Frickel (2013), for example, show a persisting clustering of industrial facilities close to waterways, independent of socio-demographic characteristics. In combination, those two independent processes may lead to the fact that minorities bear a disproportionately high level of environmental pollution. This would question the causal link between pollution and minority share.

Second, centrality of industrial facilities may be driven by some kind of selective siting, independent of the applied measures for political efficacy and economic inequality; or the clustering of minorities may be driven by high pollution within the urban core. In contrast to the first interpretation, this would suggest a causal link between pollution and minority share, mediated by the centrality of industrial facilities.

Yet, additional models (Table 5.2) do not provide an easy conclusion regarding the predictors of facility centrality. It seems that the location of the facilities is indeed independent of the centrality of the minority population within the city. Model M1 of Table 5.2 regresses the facility centrality on the relative minority centralization index (Massey & Denton, 1988), but does not exhibit a significant correlation. This means that facilities are not located more centrally if minorities live centrally. Similar conclusion apply to the level of segregation (M2). Interestingly, models M3 and M4 indicate

Table 5.2: Linear OLS models (city level). Dependent variable: facility centrality

	M1	M2	M3	M4	M5	M6	M7	M8
Centralization index[a]	0.020							0.186
	(0.114)							(0.171)
$\tilde{H}_{2000}$		-0.105						-0.198
		(0.113)						(0.155)
Unemployment ratio			0.295^{**}					$0.255^{\dagger}$
			(0.109)					(0.130)
Voter turnout				0.244^{*}				0.160
				(0.111)				(0.138)
% Employed in industry					0.118			0.147
					(0.113)			(0.122)
Cenrality of rivers						-0.071		-0.142
						(0.114)		(0.114)
Public transport							-0.084	0.021
							(0.114)	(0.118)
R^2	0.000	0.011	0.087	0.059	0.014	0.005	0.007	0.164
Adj. R^2	-0.013	-0.002	0.075	0.047	0.001	-0.008	-0.006	0.082
N	79	79	79	79	79	79	79	79

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$. All variables are centered around their mean and scaled by their standard deviation. Standard errors in parentheses.

[a] Relative Centralization Index as described by Massey and Denton (1988): proximity of the foreign population to the city centre relative to the proximity of the German population to the city centre.

that facilities are located more centrally in cities with higher economic inequality and higher political efficacy. While the first model rather supports the interpretation of the facility centrality as confounding mechanisms, models M3 and M4 rather point to its role as a mediator. Additionally, models M6 and M7 show that neither the centrality of rivers (measured as the minimal distance to the city centre), nor the extent of public transport opportunities (measured as the public transport stops per square kilometre) yield a significant effect on the centrality of industrial facilities. Though central facilities are often located next to rivers, there are also cities with central rivers but without centrally located facilities. Thus it remains somehow inconclusive why some cities exhibit centrally located facilities, while others do not. Further research using longitudinal data is needed to draw appropriate conclusions about the occurrence of centrally located facilities.

5.7 Conclusion

Though many studies investigate environmental inequality, there is surprisingly little consensus about the factors causing the uneven distribution of environmental burdens across different groups. While some empirical studies confirm that selective siting plays an important role, others rather emphasise the role of selective migration. The present study goes one step back and analyses under which circumstances we observe high levels of environmental inequality. Therefore, this study investigates varying levels of environmental inequality between German cities and tests whether different structural characteristics explain those variations.

The results of the city-fixed multilevel models reveal that residential segregation, economic inequality and political efficacy of the majority population do a rather poor job of explaining the level of environmental inequality. Though segregation has the expected positive effect, its significance depends on model specifications and its explanatory power is rather low. This is also what Downey (2007) concludes for the U.S., but nonetheless seems to be a surprising result, given the importance of residential segregation for the most common explanations of environmental inequality. Without a substantial level of spatial segregation, neither selective siting nor selective migration seem to be plausible explanations. Only the share of employees in the industrial sector exhibits a significant and positive correlation with the level of environmental inequality, pointing towards the importance of employment opportunities. However, a much stronger and more robust effect comes from the spatial distribution of pollution within the urban space. Including the spatial centrality of industrial facilities as a cross-level interaction explains more than 25% of the random slope variance. Though the measure of centrality applied in this study is rather simple (not considering the spatial distribution or density of facilities), it still does a very good job of explaining the varying levels of environmental inequality. Minorities are disproportionately affected if pollution occurs around the city centre – where minorities tend to cluster. These results point to the importance of incorporating the spatial structure of polluters within the urban space into the analysis of environmental inequality.

The results pose some questions for further research. First, the analysis is based on Germany which has low levels of spatial segregation and rather dense metropolitan areas. Thus, further research has to test whether the presented findings apply to other countries with other structural conditions. Second, the analysis relies only on industrial air pollution, while ignoring pollution coming from mobile sources like traffic. It is likely that results for mobile sources (like traffic) differ, as pollution may be generally more concentrated around the urban core. Third, this study is cross-sectional in nature. Hence, it is not possible to investigate the temporal order of facility siting and residential moving behaviour. Finally, further research should also aim to enrich the analyses by further data of the urban form of the cities. Other factors than the centrality of rivers or public transport could play an important role in determining the centrality of industrial facilities.

Nonetheless, the findings of this study challenge the standard reasoning of environmental inequality research. While most environmental inequality research focuses on individual decisions, only few studies have analysed the spatial context in which those individual processes occur. It seems, however, that structural constraints play an important role in determining the level of environmental inequality observed at the macro level. This is not to say that selective migration or siting does not occur, but only that the importance of these causal mechanisms may be overstated. Though companies may consider the minority share next to potential sites, other factors may be much more important (e.g. Elliott & Frickel, 2015; Wolverton, 2012). Similarly, environmental quality certainly plays a role for moving decisions, but other factors may be more important. Still, structural conditions may contribute to the (possibly unintended) consequence that minorities end up in more polluted neighbourhoods. This interpretation of the findings is also supported by the fact that agent-based models by Campbell et al. (2015) do not reach realistic levels of environmental inequality when just assuming selective siting and migration as causal mechanisms.

In sum, the results encourage further research not only to 'bring the polluters back in' (Grant et al., 2010), but also to bring the urban space back in, thereby considering the geographic location of polluters within urban areas. This might also help urban planners to develop strategies for avoiding or reducing the disproportionate exposure of minority households to environmental pollution.

5.A Appendix Chapter 5

Table 5.3: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
ln Air pollution	9061	4.04	6.65	0.00	20.93
% Foreigners	9061	9.00	8.43	0.00	87.10
Population	9061	2649.91	2887.97	3.00	23379.00
% 65 and older	9061	20.57	7.44	0.00	99.60
% Vacant housing	9061	3.50	3.54	0.00	60.00
Living space	9061	41.74	5.95	11.00	95.90
Distance to highway	9061	2683.66	2094.72	23.29	17239.26
Distance to rail	9061	1236.04	1240.91	0.02	10170.87
$\tilde{H}_{2000}$	79	0.03	0.01	0.01	0.08
$\tilde{D}_{2000}$	79	0.17	0.05	0.08	0.30
Unemployment ratio	79	2.34	0.37	1.28	3.43
Voter turnout	79	69.18	3.75	60.10	77.10
% Employed in industry	79	16.72	10.41	5.10	76.60
Facility centrality	79	2.78	2.17	0.00	17.78
Wages in production	79	3498.37	625.28	2422.90	5076.50
Size	79	17211.42	12977.61	4489.00	89170.00
Growth rate	79	0.80	4.20	−8.31	10.46

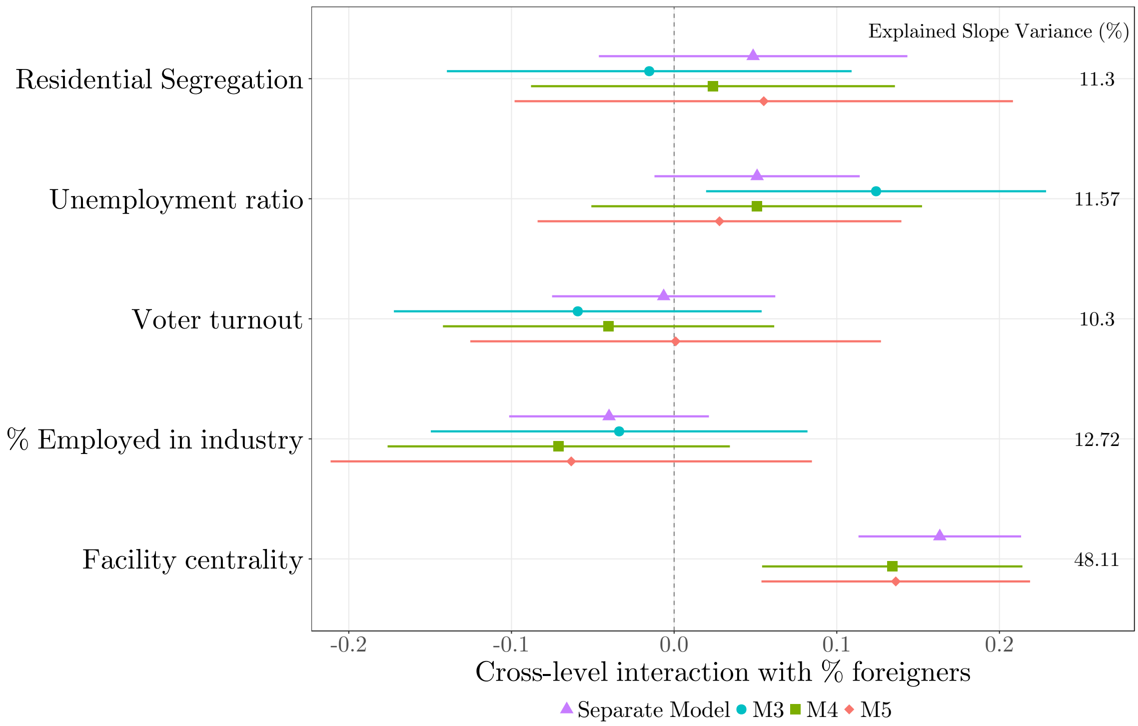

Figure 5.3: Coefficients of cross-level interactions with 95% confidence interval using toxicity-weighted air pollution, including explained slope variance for separate models

Table 5.4: City-fixed effects multilevel models conditioning on facility location. Dependent variable: ln air pollution

	M1	M2	M3	M4	M5
Census cell level					
% Foreigners	0.297***	0.150***	0.148***	0.147***	0.142***
	(0.033)	(0.033)	(0.034)	(0.029)	(0.032)
Population		0.007	0.008	0.008	0.008
		(0.011)	(0.011)	(0.011)	(0.011)
% 65 and older		0.001	0.000	0.000	0.000
		(0.010)	(0.010)	(0.010)	(0.010)
% Vacant housing		0.012	0.013	0.014	0.015
		(0.011)	(0.011)	(0.011)	(0.011)
Living space		−0.048***	−0.047***	−0.047***	−0.047***
		(0.012)	(0.012)	(0.012)	(0.012)
Predicted probability[a]		0.232***	0.232***	0.232***	0.232***
		(0.011)	(0.011)	(0.011)	(0.011)
City level					
% Foreigners × $\tilde{H}_{2000}$			0.076†	0.083*	0.103*
			(0.042)	(0.035)	(0.045)
% Foreigners × Unemployment ratio			−0.002	−0.036	−0.033
			(0.029)	(0.026)	(0.028)
% Foreigners × Voter turnout			0.054	0.036	0.032
			(0.034)	(0.029)	(0.033)
% Foreigners × % Employed in industry			0.091***	0.069**	0.059*
			(0.026)	(0.023)	(0.030)
% Foreigners × Facility centrality				0.123***	0.121***
				(0.024)	(0.024)
% Foreigners × Wages in production					0.011
					(0.034)
% Foreigners × Size					−0.036
					(0.056)
% Foreigners × Growth rate					0.029
					(0.030)
AIC	23656	23179	23230	23224	23270
N	9061	9061	9061	9061	9061
N cluster	79	79	79	79	79
σ^2 % Foreigners	0.065	0.058	0.051	0.033	0.035
σ^2 Residual	0.784	0.741	0.741	0.741	0.741

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$. Multilevel models with group centered first level variables. All variables are scaled by their standard deviation. Standard errors in parentheses.
[a] Predicted probability of probit model. Dep. var: Presence of at least one facility. Indep. vars: Distance to highway junctions, distance to railways, distance to rivers, number of facilities in neighbouring units.

Table 5.5: City-fixed effects multilevel models. Dependent variable: ln air pollution in 2012

	M1	M2	M3	M4	M5
Census cell level					
% Foreigners	0.314***	0.203***	0.196***	0.196***	0.188***
	(0.029)	(0.029)	(0.031)	(0.029)	(0.032)
Population		−0.001	−0.001	−0.000	−0.000
		(0.011)	(0.011)	(0.011)	(0.011)
% 65 and older		−0.005	−0.006	−0.006	−0.006
		(0.010)	(0.010)	(0.010)	(0.010)
% Vacant housing		0.026*	0.027*	0.028*	0.028**
		(0.011)	(0.011)	(0.011)	(0.011)
Living space		−0.047***	−0.047***	−0.047***	−0.047***
		(0.013)	(0.013)	(0.013)	(0.013)
Distance to highway		−0.048***	−0.047***	−0.047***	−0.047***
		(0.011)	(0.011)	(0.011)	(0.011)
Distance to rail		−0.155***	−0.154***	−0.153***	−0.153***
		(0.011)	(0.011)	(0.011)	(0.011)
City level					
% Foreigners × $\tilde{H}_{2000}$			0.048	0.054	0.078†
			(0.038)	(0.035)	(0.045)
% Foreigners × Unemployment ratio			0.021	−0.000	−0.001
			(0.027)	(0.026)	(0.028)
% Foreigners × Voter turnout			0.031	0.020	0.020
			(0.031)	(0.029)	(0.033)
% Foreigners × % Employed in industry			0.075**	0.061**	0.050†
			(0.024)	(0.023)	(0.030)
% Foreigners × Facility centrality				0.082***	0.080***
				(0.024)	(0.024)
% Foreigners × Wages in production					0.013
					(0.034)
% Foreigners × Size					−0.047
					(0.055)
% Foreigners × Growth rate					0.020
					(0.030)
AIC	23644	23390	23445	23451	23497
N	9061	9061	9061	9061	9061
N cluster	79	79	79	79	79
σ^2 % Foreigners	0.049	0.042	0.039	0.033	0.034
σ^2 Residual	0.785	0.759	0.759	0.759	0.760

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$. Multilevel models with group centered first level variables. All variables are scaled by their standard deviation. Standard errors in parentheses.

Table 5.6: Linear OLS models. Dependent variable: coefficient first stage[a]

	M1	M2	M3	M4	M5
$\tilde{H}_{2000}$	0.181		0.241^{*}	0.271^{**}	0.349^{**}
	(0.112)		(0.113)	(0.100)	(0.122)
Facility centrality		0.474^{***}		0.463^{***}	0.456^{***}
		(0.100)		(0.099)	(0.100)
Unemployment ratio			0.010	−0.112	−0.111
			(0.116)	(0.106)	(0.110)
Voter turnout			$0.222^{\dagger}$	0.162	0.137
			(0.123)	(0.110)	(0.121)
% Employed in industry			0.353^{**}	0.276^{**}	$0.205^{\dagger}$
			(0.106)	(0.095)	(0.122)
Wages in production					0.070
					(0.123)
Size					−0.132
					(0.118)
Growth rate					0.122
					(0.105)
R^2	0.033	0.224	0.191	0.376	0.397
Adj. R^2	0.020	0.214	0.147	0.333	0.328
N	79	79	79	79	79

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$. All variables are scaled by their standard deviation. Standard errors in parentheses.

[a] Interaction coefficient between % foreigners and city dummy, dep. var.: ln air pollution, control variables similar to first level controls in Table 5.1.

Table 5.7: City-fixed effects SLX multilevel models. Dependent variable: ln air pollution

	Model 1	Model 2	Model 3	Model 4	Model 5
Census cell level					
% Foreigners	0.089***	0.068***	0.064***	0.052***	0.060**
	(0.018)	(0.019)	(0.019)	(0.016)	(0.020)
W [% Foreigners]	0.421***	0.319***	0.314***	0.318***	0.296***
	(0.046)	(0.048)	(0.052)	(0.051)	(0.052)
City level					
% Foreigners × $\tilde{H}_{2000}$			0.021	0.007	0.026
			(0.021)	(0.016)	(0.027)
% Foreigners × Unemployment ratio			0.003	−0.006	−0.007
			(0.019)	(0.017)	(0.020)
% Foreigners × Voter turnout			0.029	0.018	0.020
			(0.021)	(0.018)	(0.024)
% Foreigners × % Empl. in industry			0.038*	0.033†	0.035
			(0.019)	(0.017)	(0.022)
% Foreigners × Facility centrality				0.037*	0.038†
				(0.018)	(0.020)
% Foreigners × Wages in production					−0.007
					(0.022)
% Foreigners × Size					−0.007
					(0.030)
% Foreigners × Growth rate					0.011
					(0.022)
W [% Foreigners] × $\tilde{H}_{2000}$			0.057	0.079	0.109
			(0.065)	(0.064)	(0.072)
W [% Foreigners] × Unemployment ratio			0.005	−0.044	−0.046
			(0.044)	(0.045)	(0.043)
W [% Foreigners] × Voter turnout			0.031	0.009	0.017
			(0.051)	(0.051)	(0.051)
W [% Foreigners] × % Empl. in industry			0.078†	0.047	0.047
			(0.040)	(0.040)	(0.046)
W [% Foreigners] × Facility centrality				0.160***	0.157***
				(0.041)	(0.038)
W [% Foreigners] × Wages in production					−0.014
					(0.054)
W [% Foreigners] × Size					−0.075
					(0.089)
W [% Foreigners] × Growth rate					0.031
					(0.048)
Fixed effects	yes	yes	yes	yes	yes
Random slope	yes	yes	yes	yes	yes
First level controls[a]	no	yes	yes	yes	yes
AIC	22996	22857	22943	22961	23015
N	9061	9061	9061	9061	9061
N cluster	79	79	79	79	79
σ^2 % Foreigners	0.008	0.007	0.007	0.000	0.005
σ^2 W [% Foreigners]	0.131	0.130	0.128	0.121	0.097
σ^2 Residual	0.724	0.706	0.707	0.708	0.707

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$. Multilevel models with group centered first level variables. All variables are scaled by their standard deviation. Standard errors in parentheses. The neighbours weights matrix W is specified a contiguity weights matrix.
[a] Similar first level controls as in Table 5.1. All control variables are additionally included as spatially lagged covariates.

Chapter 6

Final Discussion and Conclusion

6.1 Introduction

The topic of environmental inequality has received extensive attention in the United States. In Germany, by contrast, it has only played a minor role in sociological research, which seems to be a serious shortcoming given the importance of environmental conditions for other dimensions of social life. Thus, the main objective of this book is to resolve this shortcoming and to extend the empirical research on environmental inequality in Germany. Therefore, this book dealt with two main research questions:

1. Are minorities in Germany disproportionately exposed to environmental pollution?
2. What causes this disproportionate exposure of minorities to environmental pollution?

To answer these questions, I conducted several studies relying on different data sources and methods. By using different analytical strategies, this book assesses the main research questions from a diverse set of viewpoints, elaborating on several aspects of environmental inequality. However, the final question remains as to how those findings from different perspectives combine to a congruent overall picture of environmental inequality in Germany. In the following sections, I will first summarise the main findings of each study, subsequently combine these findings to an overarching conclusion, and finally discuss limitations and implications for further research.

6.2 Summary of results

6.2.1 Study 1

The first study in this book follows the standard method of empirical research in the social sciences by using household-level survey data of the German Socio-Economic Panel (SOEP). This study relies on the subjectively perceived impairment through air pollution and individual moving trajectories. In a first step, we replicate findings of previous research by showing that minorities and low-income households face a higher impairment through air pollution.

However, the main objective of this study is to investigate whether this higher impairment stems from selective migration patterns. Therefore, the study uses fixed-effects panel estimators to analyse whether income helps to increase the moving returns and whether minority households experience different environmental benefits from relocations. The results of the analyses reveal that income plays a rather minor role. Though increases in income lead to significantly higher improvements in environmental quality due to relocations, this effect is rather weak in its magnitude and depends on the model specification. Minorities, in contrast, experience a strong disadvantage when moving. First generation immigrants are not able to improve the environmental quality at their place of destination at all, while German households experience relatively strong improvements due to relocations. This finding confirms that selective migration patterns indeed contribute to the disproportionate exposure of minority households to environmental pollution.

The fact that this disadvantage of minority households is only marginally reduced when controlling for income contradicts the ‘racial income-inequality hypothesis’, assuming that minority disadvantages mainly stem from differences in the socio-economic status. Interestingly, the disadvantage completely vanishes in the second generation; second generation immigrants experience similar moving returns as native German households. To some extent, this finding contradicts the assumption that discriminatory barriers on the housing market are responsible for the differing moving returns, as many visible characteristics like names or looks are still present within the second generation. Though discrimination based on other characteristics cannot be ruled out, disadvantages may also result from different mechanisms. For instance, network ties to the mainstream could play an important role in finding adequate housing, or homogeneity preferences of (especially first generation) immigrant households might constrain their choice set of possible destinations. However, those explanations remain speculative at this stage of research.

It is important to keep in mind that this study has some limitations. First, it relies on subjective measures of air pollution. Though this should be less of a problem with fixed-effects panel estimators, changes in other characteristics than air pollution might still influence the time-varying perception of environmental quality. Second, we only test the mechanism of selective in-migration. The study investigates whether households experience different returns based on household characteristics if households move to another place of residence. We do not test whether increases in air pollution lead to the selective out-migration of socio-economically affluent or native German households, a question that should be investigated in further research.

6.2.2 Study 2

The second study differs from the remaining studies of this book, as it does not deal with the topic of environmental inequality per se, but rather investigates a methodological issue, which gains importance in the subsequent studies: statistical methods for the analysis of spatial data. At the outset, the study provides an overview of the most commonly used spatial regression models, and shows analytically that non-spatial regression models suffer from estimation bias in several constellations of spatial dependence, thereby giving a strong motivation to rely on spatial regression models when using spatially aggregated data. However, a variety of different model specifications exist, assuming different dependency structures within the data. Furthermore, specification tests are of little help in many situations, as all strategies suffer from severe drawbacks. Thus, I employ several Monte Carlo experiments to assess the performance of different specifications under the possibility of misspecification. The study extends previous simulations by evaluating the bias in impacts rather than the coefficients, as the regression impacts are the measures of interest in applied research.

The results reveal several interesting findings. First, in many situations the non-spatial OLS model exhibits unbiased estimates of the direct impacts (though it yields biased estimates of the coefficients). Nevertheless, several situations exist in which OLS suffers from bias. Second, the most popular spatial model specifications – the spatial autoregressive model (SAR), the spatial error model (SEM), and the spatial autoregressive combined model (SAC) – show severe drawbacks: these models are very sensitive to misspecification. The more flexible specifications – spatial lag of X (SLX), spatial Durbin (SDM), and spatial Durbin error model (SDEM) – all provide robust estimates of the direct impacts. Third, regarding the indirect impacts or 'spillover' effects, the relatively simple SLX model outperforms the more complex SDM and SDEM in a 'mixed world', i.e. in the presence of all three sources of spatial dependence (autocorrelation in the dependent variable, autocorrelation in the error, and local 'spillover' effects). Likewise, the results show that the SDM is more sensitive to non-spatial omitted variable bias compared to SLX and SDEM. This gives some motivation to rely on SLX or SDEM models if no prior information on the structure of the spatial dependence is present.

Note, however, that the conclusions of this study are based on the robustness of the models against misspecification. If the model correctly specifies the data generating process, all model specifications (SAC is an exception to some extent) yield unbiased estimates of the true impacts. A general conclusion that one should always use SLX is thus erroneous. As a first step, it should be evaluated whether theoretical considerations might guide the empirical model. If theoretical arguments give strong reasons for global or local 'spillovers', the specification best suiting this argument should be preferred. Yet, if several theoretical mechanisms might be at work, SLX seems to provide relatively robust estimates. This yields important implications for the following study, as environmental pollution as well as the minority share exhibit high levels of spatial correlation, which can result from various theoretical processes.

6.2.3 Study 3

The third study again deals with the main research questions of this book by investigating the presence and spatial pattern of environmental inequality in Germany. In contrast to the first study, study 3 uses objective measures of toxic air pollution and spatially aggregated socio-demographic data. Therefore, the study combines georeferenced pollution data of the European Pollution Release and Transfer Register (E-PRTR) and spatial data of the 2011 German census by allocating the toxicity-weighted amount of industrial air pollution to each census unit. By doing so, this study pursues two main objectives. First, it aims to assess whether minorities are affected by a disproportionate amount of toxic industrial emissions to air. Second, the study aims to draw a picture of the spatial patterns of environmental inequality. Theoretically, both mechanisms – selective migration and selective siting – predict that pollution and minority households spatially cluster together. Following the selective migration argument, a polluted area should attract minority households, which in turn attracts further minority households because of similarity preferences, thereby leading to an accumulation of minority households around industrial facilities. Similarly, if we follow the argument of selective siting, facilities should be placed predominantly in areas where minority households cluster, as it seems plausible that protest against siting declines with distance rather than strict boundaries.

The results of study 3 confirm results of study 1: the share of minorities within a spatial census unit positively correlates with the amount of toxic industrial air pollution. This effect is very robust against different model specifications and considerably high in magnitude. Even when considering the distance to the nearest industrial facility or the total amount of pollution without toxicity-weighting, minorities bear a disproportionately high burden. Furthermore, the analyses show that minorities live in more polluted areas in a German-wide comparison, but also live in more polluted areas within the communities, thereby offering the first nation-wide confirmation of the disproportionate exposure of minority households to environmental pollution. The disadvantage of minorities even persists when housing-related variables – the living space per inhabitant and the share of vacant housing – are controlled for. This, again, raises some doubt on the assumption that disadvantages of minorities stem from lower socio-economic resources and the need for affordable housing opportunities.

Regarding the spatial patterns, the results of the third study indeed confirm that the broader neighbourhood matters. The SLX models reveal that pollution tends to be high especially in those areas where a high share of minority households agglomerates. Moreover, the connection between the minority share and environmental pollution is significantly stronger within urban regions than within rural regions. Especially in urban areas, minority households spatially cluster around hazardous industrial facilities. This could be seen as an indication that residential segregation around industrial facilities plays a major role in shaping the extent of environmental inequality. One could assume that population pressure and tough housing market situations lead to higher levels of segregation in urban areas, which, in consequence, steers minorities into more polluted areas. However, the cross-sectional nature of the data does not allow to empirically test why neighbouring units matter in predicting the amount of pollution. As demonstrated in the preceding study, several causal pathways are possible, but cannot be separated empirically.

6.2.4 Study 4

The fourth study constitutes a follow-up study on the previous study. As in the previous empirical assessment, study 4 uses spatially aggregated data of the E-PRTR and the 2011 German census. However, in this follow-up study, the observations are restricted to the German metropolitan areas (i.e. all communities with at least 100,000 inhabitants), and further city-level characteristics derived from the INKAR database and OpenStreetMap are added to the data. The main objective of this study is to investigate the variation in the level of environmental inequality between German cities. While some cities exhibit a very high correlation of minority share and environmental pollution, others do not. To investigate this variation, the fourth study relies on city fixed-effects multi-level models, incorporating city-level characteristics as cross-level interactions. Note that this study, in contrast to the preceding study, uses the total amount of air pollution instead of the toxicity-weighted air pollution to deal with some methodological issues. Using toxicity-weights, however, leads to similar results with even stronger support for the conclusions.

Theoretically, several ways exist to explain the differences in environmental inequality between cities. First, according to the theories of selective migration and selective siting, and in line with the results of study 3 in this book, the level of residential segregation should play an important role. On the one hand, residential segregation constitutes a necessary condition for selective siting. On the other hand, it seems implausible that minority households moved selectively into specific regions in the past if we do not observe a substantial level of residential segregation. Second, following the 'racial income-inequality thesis' higher economic inequalities should lead to higher levels of environmental inequality. Third, according to the selective siting argument and its reasoning based on political opposition, environmental inequality should increase with increasing political efficacy of the majority group. Still, previous results in the United States point towards two alternative explanations: the extent of the industrial sector and the centrality of industrial facilities. On the one hand, industrial facilities may provide job opportunities, which are traded off against environmental quality. If minority residents are overrepresented the industrial sector, a high share of employees in the industrial sector would lead to a higher exposure of minorities. On the other hand, research on residential segregation has shown that minorities tend to cluster around central city districts because of infrastructural and network-related opportunities. Thus, the position of industrial facilities – and especially the centrality of these facilities – may be a crucial driver of environmental inequality.

Results of this last study are somewhat surprising. While economic inequality and political efficacy show no significant effect on the level of environmental inequality, higher levels of segregation are associated with higher levels of environmental inequality. However, segregation is only marginally significant and not robust across different model specifications. Furthermore, its explanatory power is very weak, indicating that residential segregation does a rather poor job of explaining environmental inequality. This seems somewhat odd given the importance of residential segregation for the mechanisms of selective siting and selective migration. The share of manufacturing workers, in contrast, exhibits a more robust and stable effect on environmental inequality, suggesting that minorities live disproportionately close to industrial facilities because these facilities provide attractive employment opportunities. Nonetheless, the results indicate that the centrality of industrial facilities is much more decisive for the level of environmental inequality: in cities with centrally located facilities minorities bear an especially high burden of environmental pollution. The centrality of industrial facilities does not only explain the direct correlation between minority share and environmental pollution, but also the clustering effects observed in the preceding study. Moreover, the centrality of facilities explains a large proportion of the variation between German cities. Interestingly, additional analyses reveal that the centrality of industrial facilities – at least in a cross-sectional assessment – seems to be independent of the centrality of the minority population, thereby casting some doubt on the causal link between minority share and environmental pollution.

In sum, study 4 provides results, which are challenging for the standard strand of theoretical reasoning in environmental inequality research. City-level characteristics based on the theories of selective migration and selective siting play a rather minor role in explaining varying levels of environmental inequality. The results emphasise the importance of infrastructural conditions for the formation of the disproportionate burden of minority households. Migration decisions may be driven by factors other than environmental quality, and siting decisions may be driven by factors other than the minority share within a region. Still, under specific structural conditions like centrally located industries, minorities tend to end up in closer proximity to industrial facilities.

6.3 Discussion and concluding remarks

In the end, the crucial question is how the results of this book help to answer the two main questions of 1) whether minorities are affected by a disproportionate amount of environmental pollution in Germany, and 2) what causes this disproportionate exposure.

6.3.1 Existence of environmental inequality

The *first question* is relatively easy to answer. All studies presented in this book confirm a disproportionate exposure of minorities to environmental pollution. When using household-level data, we find that minorities report a higher impairment through air pollution, which replicates earlier findings by Kohlhuber et al. (2006). When using spatially aggregated data, we find that minorities experience a disproportionately high exposure to toxic industrial emissions and live closer to industrial facilities. This is in line with other studies on the spatially aggregated level, showing that minorities experience a lack of public green-space (Kabisch & Haase, 2014), or live closer to industrial facilities in the city of Hamburg (Raddatz & Mennis, 2013). However, the present book extends those previous studies by providing the first nationwide assessment using objective indicators of environmental pollution. In addition, this book employs a methodological approach analogue to studies in the Unites States (e.g. Ard, 2015; Ash et al., 2013; Ash & Fetter, 2004; Banzhaf & Walsh, 2008; Downey, 2006a) by using the amount of toxicity-weighted environmental pollution. This provides the first nationwide empirical evidence for the disproportionate exposure of minorities to harmful industrial substances in Germany.

The finding of a disproportionate burden of minorities applies to a country-wide comparison, but also to a within-community comparison. Similar to research in the United States (Ash & Fetter, 2004), we find a stronger correlation between the minority share and pollution within communities than across the whole country. This means that the disproportionate burden of minorities is not only the result of different pollution levels across the country, or an urban-rural divide, but also occurs on the small scale of single communities and cities. Minorities live in more polluted areas in Germany but also in more polluted parts of the communities. In sum, this book confirms the presence of environmental inequality in Germany based on different data sources, different methods, and different spatial scales.

In addition, the results of this book provide some insights into the spatial pattern of environmental inequality. First, environmental pollution is especially high in areas where minorities agglomerate, meaning that there is some kind of clustering process around industrial hazards. Second, minorities are especially disadvantaged within urban or metropolitan regions: in these areas minorities experience an additional disadvantage and the clustering processes seem to be especially pronounced within urban regions. This is concerning given that a high proportion of minority households resides within metropolitan regions. Third, there is considerable variation in the magnitude of environmental inequality across German cities, showing that there are even some cities in which the majority group bears a higher burden of environmental pollution. Gelsenkirchen and Hannover, for example, seem to be outstanding cases in this regard, a fact that should receive more attention in further research. Notwithstanding, the vast majority of German cities exhibit a positive correlation between the minority share and environmental pollution.

6.3.2 Causes of environmental inequality

The *second question*, in contrast, is relatively hard to answer. At the beginning of this project it seemed clear to me that there are two possible causal mechanisms: selective siting and selective migration. Both mechanisms are intuitive and plausible. However, the studies in this book indicate that the causes of environmental inequality are more complicated than generally assumed by these theoretical explanations.

First, we could confirm that selective migration patterns contribute to the disproportionate exposure of minority households to environmental pollution. Similar to findings in the United States (Crowder & Downey, 2010; Pais et al., 2014), minority households in Germany indeed experience significantly lower improvements in environmental quality when moving to a new place of residence. Interestingly, this disadvantage cannot be explained by socio-economic differences. Neither do we find a noteworthy reduction of the minority effect in study 1 when controlling for income, nor can housing-related variables explain the disadvantage in study 3. This indicates that the assumptions of the 'racial income-inequality thesis' are not met by empirical evidence. According to the standard strand of reasoning, discrimination on the housing market remains as a possible explanation for the differing moving trajectories. However, the fact that second generation immigrant do not experience any disadvantages compared to native German households casts some doubt on this explanation. Though results are not fully conclusive in this regard, it seems implausible to assume that housing agents or landlords only discriminate against first-generation but not second-generation immigrants. Then why do we observe the lower moving returns of minorities? I will return to this question later in the current section.

Second, the clustering pattern of minority households around hazardous facilities, at a first glance, seems to conform to the theories of selective siting and selective migration. Both theories offer some support for the argument that minorities cluster around hazardous facilities. In combination with the finding that the general minority effect as well as the clustering effect are much stronger within urban regions, one could assume that residential segregation plays an important role in shaping environmental inequality. The assumption that minorities residentially segregate around more polluted areas is also in line with findings of the first study, showing the disadvantaged moving trajectories of minorities. Though results by Ard (2016) from the United States support this idea, others (Downey, 2007; Downey et al., 2008) indicate that segregation is only weakly associated with environmental inequality.

In line with Downey's (2007) conclusions, results of the last study point towards a different interpretation of the preceding findings. The comparison of German cities reveals that segregation plays only a minor role in shaping the disproportionate exposure of minorities to environmental pollution. Although residential segregation seems to be a necessary condition for environmental inequality – in a completely desegregated city, it would not be possible to observe a higher exposure of one group – it is apparently not a sufficient condition for the appearance of environmental inequality. In contrast, the results point towards two independent mechanisms. First, minorities tend to cluster around the urban core in nearly all cities. This observation has been made by scholars of residential segregation in the United States (e.g. Logan, Zhang & Alba, 2002; Massey & Denton, 1988) a long time ago, and is mostly attributed to other factors than the environmental quality. For instance, Alba et al. (1999) show that speaking the language of the host country increases the probability of immigrant households to move from the central city to suburban districts, as language skills facilitate the access to other networks and infrastructures. Similarly, Crowder et al. (2012) stress the importance of metropolitan-level characteristics like city size, the level of suburbanization, or the available housing stock in constraining the migration patterns of minorities. Second, some cities exhibit a considerably high pollution around the urban core. In combination, both mechanisms lead to a disproportionate burden of minority residents.

This conclusion also conforms to the results of the preceding studies. Study 1 shows the low moving returns of minority households, but casts doubt on the explanation that these disadvantages stem from socio-economic disparities or discrimination. An alternative explanation is that minorities constrain their choice set of possible destinations to more central parts of the city, thereby experiencing lower improvements in environmental quality due to relocations. This fits also to the vanishing effect in the second immigrant generation, if language skills indeed facilitate the access to integrated and suburban neighbourhoods (Alba et al., 1999; Logan et al., 2002). Moreover, study 3 shows that minorities tend to cluster around environmental hazards. Again, this may result from the tendency to reside in more central regions of the cities, producing agglomerations of high minority areas around the urban core. If pollution, at the same time, occurs around the urban core, we observe a clustering of minority inhabitants around polluted areas, which is, however, not a result of clustering around hazards, but rather a result of clustering around the urban core. This way of reasoning is also in line with simulation results by Campbell et al. (2015) and Kim et al. (2014). By using agent based models, they show that just assuming selective siting and selective migration processes does not lead to realistic levels of environmental inequality. Only after including homogeneity preferences as decision rules for relocations, the models exhibit a level of environmental inequality, which reflects the levels uncovered in observational studies.

In sum, the results of this book provide the following answer to the *second question*. Selective migration contributes to environmental inequality. However, the general argument of selective migration in environmental inequality research may be oversimplified. It assumes perfect information about the environmental quality of possible residences, low transaction costs, and flexible responses of households to the provision of public goods. Yet, migration decisions and the choice of destination are likely to be driven by other factors than environmental quality, like homogeneity preferences or the access to networks and other social infrastructures. Thus, selective migration into polluted areas may be a *side product* of other (non-monetary) factors involved in migration decisions. Similar conclusions can be drawn regarding the siting of industrial facilities: the demographic composition of nearby residents may play a role, but infrastructural and historical conditions are probably more important for siting decisions (see e.g. Baden & Coursey, 2002; Elliott & Frickel, 2013, 2015; Wolverton, 2012). What follows are two – possibly independent – processes: 1) urban infrastructure confines residential choices of minorities to the urban core, and 2) urban infrastructure facilitates centrally located industries. In combination, both processes lead to a disproportionate burden of minority households.

Note that this conclusion offers a congruent explanation for all findings in this book, but is not directly tested in the empirical studies. In order to test this overarching conclusion empirically, further research needs to combine household-level panel data with longitudinal neighbourhood-level data and pollution estimates. Investigating the interaction between micro-mechanisms and macro-conditions would allow to develop a more integrated picture of environmental inequality processes, thereby bolstering our knowledge of the causal mechanisms.

6.4 Implications for further research

Based on the conclusions, this book offers several implications for further research. On the one hand, further research should try to overcome some of the limitations of this book. On the other hand, further research should elaborate on some results of this book in more detail.

Regarding the limitations of this book, several aspects need to be mentioned. First and foremost, this book does not offer any direct tests on the mechanism of selective siting, owed mostly to the limited availability of suitable data. Though E-PRTR provides annual data on polluting facilities from 2007 on, till now we lack a source of annual socio-demographic data on a fine-grained spatial level. A solution for this problem could be to use less fine-grained data (e.g. community level data). However, this increases the unobserved variance within spatial units and bears the risk of masking decisive processes. Furthermore, many of the facilities included in E-PRTR were already in operation prior to 2007. Thus, to comprehensively test the selective siting mechanism, further research needs to collect appropriate information on siting dates, and ideally merge this information with socio-demographic data at the time of siting. Second, this book is mostly constrained to industrial pollution. Although the first study uses a subjective evaluation, not separating between the sources of pollution, the remaining studies are restricted to pollution stemming from industrial activities, thereby ignoring pollution from other sources like traffic. Conclusions may differ when using other sources of pollution, as traffic pollution, for instance, is likely to be generally more concentrated in central urban regions. Third, the two studies using E-PRTR data apply a relatively imprecise measure of exposure by allocating aerial emissions proportionately to the surrounding census cells. In the United States, the Risk-Screening Environmental Indicators (RSEI) model provides a much more elaborate method to estimate the exposure of surrounding regions to the emissions of industrial facilities by using meteorological data, facility-specific information, and properties of particulates. Further research should aim to develop or apply similar methods to the E-PRTR in Germany and provide more accurate estimates of exposure to environmental emissions. Fourth and again owed to data restrictions, this book did not connect the micro-level data used in the first study to macro-level outcomes of the remaining studies. Though we observe that selective migration patterns take place, it is not possible to evaluate the impact on the macro-level outcome. Therefore, further research should aim to connect longitudinal micro-level moving trajectories to the macro-level data of the socio-demographic composition and the level of industrial pollution, allowing to test the interplay of micro-mechanisms and macro-outcomes.

Besides these points, the results of this book also provide several implications that should be considered in further research. First, further research should pay more attention to the causes of selective migration patterns. This books shows that the moving disadvantage of minority households is not a result of lower socio-economic resources and the vanishing disadvantage in the second immigrant-generation casts some doubt on the explanation resting on discrimination. Though this book offers an alternative explanation based on moving trajectories within central city areas, this should be investigated empirically in further research. Furthermore, it is certainly an interesting questions whether social network ties and homogeneity preferences help to explain the lower improvements of minorities due to residential mobility. Altogether, the theoretical argument of selective migration needs to be further enriched by including more elaborate fine-grained mechanisms rather than just assuming that households seek to improve environmental quality and are only constrained by economic resources or discriminatory barriers. Instead, environmental quality at the place of destination might be a side product of more complicated moving decisions (e.g. Alba et al., 1999;

Crowder et al., 2012). Thus, more attention needs to be paid to the fine-grained mechanisms of selective migration.

Second, research in the area of environmental inequality needs to account for the urban structures where migration and siting processes take place. For instance, previous research has already demonstrated how the inclusion of infrastructural characteristics can change the conclusions regarding selective siting mechanisms (e.g. Baden & Coursey, 2002; Elliott & Frickel, 2015; Wolverton, 2012). However, urban characteristics are also very likely to influence the moving behaviour. For instance, the availability of public transportation, the extent of urban sprawl, land use policies and other urban characteristics may contribute to residential choices and thus also influence the extent of environmental inequality. In a similar line, Schweitzer and Stephenson (2007) have already advocated a closer connection between environmental inequality research and urban studies. Importantly, all social processes take place within a spatial context and research on environmental inequality has only started to take some of many possible characteristics of the urban space into account. Further research in this area seems particularly important, as identifying decisive urban forms and characteristics offers the opportunity to develop promising policy interventions.

Third and related to the previous point, environmental inequality research should more carefully elaborate on the statistical methods applied to spatial data. As this book (along with an extensive amount of research on spatial econometrics) shows, applying non-spatial regression methods to spatial data may not only produce erroneous inference but also biased point estimates. Moreover, using spatial data but non-spatial methods actually neglects potentially important information inherent in the data. As results of this book demonstrate, using spatial regression models can help to produce new insights into the spatial patterns, which, in turn, can foster our knowledge about the causal mechanisms. For instance, the sorting model of Tiebout (1956) as formalised by Banzhaf and Walsh (2008, 2013) predicts that moving decisions are a function of the environmental quality at the place of residence, but also of the quality in neighbouring regions, as these regions constitute the available choice set. With longitudinal data, this theoretical model could be explicitly specified and tested by applying spatial regression models, including spatially lagged covariates. Hence, spatial regression models should receive more attention in the field of environmental inequality research, as environmental inequality constitutes an explicitly spatial research question.

Environmental inequality constitutes a serious dimension of social inequality, as environmental pollution affects many dimensions of social life, like physical and mental health, social activity, or subjective well-being. Thus, it is vital to identify the causal forces leading to the disproportionate exposure of minorities to hazardous environmental pollution. Only when understanding the causes, we can develop promising strategies to prevent the persisting disadvantage of minorities. To achieve this goal, further research on environmental inequality has to elaborate on various theoretical and empirical aspects, and find explanations for the current inconsistencies in the field. Especially the combination of georeferenced micro-level moving trajectories and spatial macro-level data seems to be a promising way of bolstering our knowledge of the causal processes leading to environmental inequality.

References

Alba, R. D., Logan, J. R., Stults, B. J., Marzan, G. & Zhang, W. (1999). Immigrant groups in the suburbs: A reexamination of suburbanization and spatial assimilation. *American Sociological Review*, *64*(3), 446–460.

Allison, P. D. (2009). *Fixed Effects Regression Models* (Vol. 160). Los Angeles: Sage.

Ambrey, C. L., Fleming, C. M. & Chan, A. Y.-C. (2014). Estimating the cost of air pollution in South East Queensland: An application of the life satisfaction non-market valuation approach. *Ecological Economics*, *97*, 172–181.

An, R., Zhang, S., Ji, M. & Guan, C. (2018). Impact of ambient air pollution on physical activity among adults: A systematic review and meta-analysis. *Perspectives in Public Health*, *138*(2), 111–121.

Anderton, D. L., Anderson, A. B., Oakes, J. M. & Fraser, M. R. (1994). Environmental equity: The demographics of dumping. *Demography*, *31*(2), 229–248.

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton Univ. Press.

Angrist, J. D. & Pischke, J.-S. (2015). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton: Princeton Univ. Press.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, *26*(2), 153–166.

Anselin, L. & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. E. A. Giles (Eds.), *Handbook of Applied Economic Statistics* (pp. 237–289). New York: Dekker.

Anselin, L., Bera, A. K., Florax, R. & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, *26*(1), 77–104.

Ard, K. (2015). Trends in exposure to industrial air toxins for different racial and socioeconomic groups: A spatial and temporal examination of environmental inequality in the U.S. from 1995 to 2004. *Social Science Research*, *53*, 375–390.

Ard, K. (2016). By all measures: An examination of the relationship between segregation and health risk from air pollution. *Population and Environment*, *38*(1), 1–20.

Arora, S. & Cason, T. N. (1999). Do community characteristics influence environmental outcomes? Evidence from the Toxics Release Inventory. *Southern Economic Journal*, *65*(4), 691–716.

Ash, M., Boyce, J. K., Chang, G. & Scharber, H. (2013). Is environmental justice good for white folks? Industrial air toxics exposure in urban America. *Social Science Quarterly*, *94*(3), 616–636.

Ash, M. & Fetter, T. R. (2004). Who lives on the wrong side of the environmental tracks? Evidence from the EPA's Risk-Screening Environmental Indicators model. *Social Science Quarterly*, *85*(2), 441–462.

Auspurg, K., Hinz, T. & Schmid, L. (2017). Contexts and conditions of ethnic discrimination: Evidence from a field experiment in a German housing market. *Journal of Housing Economics*, *35*, 26–36.

Baden, B. M. & Coursey, D. L. (2002). The locality of waste sites within the city of Chicago: A demographic, social, and economic analysis. *Resource and Energy Economics*, *24*(1-2), 53–93.

Baden, B. M., Noonan, D. S. & Turaga, R. M. R. (2007). Scales of justice: Is there a geographic bias in environmental equity analysis? *Journal of Environmental Planning and Management*, *50*(2), 163–185.

Banzhaf, H. S. & McCormick, E. (2012). Moving beyond cleanup: Identifying the crucibles of environmental gentrification. In H. S. Banzhaf (Ed.), *The Political Economy of Environmental Justice* (pp. 23–51). Palo Alto: Stanford Univ. Press.

Banzhaf, H. S., Sidon, J. & Walsh, R. P. (2012). Environmental gentrification and discrimination. In H. S. Banzhaf (Ed.), *The Political Economy of Environmental Justice* (pp. 75–111). Palo Alto: Stanford Univ. Press.

Banzhaf, H. S. & Walsh, R. P. (2008). Do people vote with their feet? An empirical test of Tiebout's mechanism. *American Economic Review*, *98*(3), 843–863.

Banzhaf, H. S. & Walsh, R. P. (2013). Segregation and Tiebout sorting: The link between place-based investments and neighborhood tipping. *Journal of Urban Economics*, *74*, 83–98.

Bayer, P., Keohane, N. & Timmins, C. (2009). Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management*, *58*(1), 1–14.

BBSR. (2017). *Indikatoren und Karten zur Raum- und Stadtentwicklung (INKAR) 2017.* Retrieved 20.03.2017, from http://www.inkar.de/

Been, V. (1994). Locally undesirable land uses in minority neighborhoods: Disproportionate siting or market dynamics? *The Yale Law Journal*, *103*(6), 1383.

Been, V. & Gupta, F. (1997). Coming to the nuisance or going to the barrios - A longitudinal analysis of environmental justice claims. *Ecology Law Quarterly*, *24*(1), 1–56.

Best, H. & Rüttenauer, T. (2018). How selective migration shapes environmental inequality in Germany: Evidence from micro-level panel data. *European Sociological Review*, *34*(1), 52–63.

Betz, T., Cook, S. & Hollenbach, F. (2018). *Random? As if – Spatial Interdependence and Instrumental Variables.* SocArXiv. Retrieved 25.05.2018, from `https://osf.io/preprints/socarxiv/pgrcu/`

Bivand, R. & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, *63*(18), 1–36.

Bolte, G. & Mielck, A. (Eds.). (2004). *Umweltgerechtigkeit: Die soziale Verteilung von Umweltbelastungen.* Weinheim: Juventa.

Bowen, W. M., Salling, M. J., Haynes, K. E. & Cyran, E. J. (1995). Toward environmental justice: Spatial equity in Ohio and Cleveland. *Annals of the Association of American Geographers*, *85*(4), 641–663.

Bresnahan, B. W., Dickie, M. & Gerking, S. (1997). Averting behavior and urban air pollution. *Land Economics*, *73*(3), 340.

Briggs, D., Abellan, J. J. & Fecht, D. (2008). Environmental inequity in England: Small area associations between socio-economic status and environmental pollution. *Social Science & Medicine*, *67*(10), 1612–1629.

Brüderl, J. & Ludwig, V. (2015). Fixed-effects panel regression. In H. Best & C. Wolf (Eds.), *The Sage Handbook of Regression Analysis and Causal Inference* (pp. 327–357). London: Sage.

Bryant, B. & Mohai, P. (1992). *Race and the Incidence of Envrionmental Hazards: A Time for Discourse.* Boulder: Westview.

Buehler, R. (2011). Determinants of transport mode choice: A comparison of Germany and the USA. *Journal of Transport Geography*, *19*(4), 644–657.

Bullard, R. D. (1990). *Dumping in Dixie: Race, Class, and Environmental Quality.* Boulder: Westview.

Burniaux, J.-M., Dang, T.-T., Fore, D., Förster, M. F., Mira D'Ercole, M. & Oxley, H. (1998). Income distribution and poverty in selected OECD countries. *Economics Department Working Papers*, *189*. Retrieved from `http://www.oecd.org/eco/public-finance/1864447.pdf`

Campbell, H. E., Kim, Y. & Eckerd, A. (2015). *Rethinking Environmental Justice in Sustainable Cities: Insights from Agent-Based Modeling.* New York and London: Routledge.

Campbell, H. E., Peck, L. R. & Tschudi, M. K. (2010). Justice for all? A cross-time analysis of Toxics Release Inventory facility location. *Review of Policy Research*, *27*(1), 1–25.

Casey, J. A., James, P., Cushing, L., Jesdale, B. M. & Morello-Frosch, R. (2017). Race, ethnicity, income concentration and 10-year change in urban greenness in the United States. *International Journal of Environmental Research and Public Health*, *14*(12), 1546.

Catney, G. (2016). Exploring a decade of small area ethnic (de-)segregation in England and Wales. *Urban Studies*, *53*(8), 1691–1709.

Choi, S. J., Ondrich, J. & Yinger, J. (2005). Do rental agents discriminate against minority customers? Evidence from the 2000 Housing Discrimination Study. *Journal of Housing Economics*, *14*(1), 1–26.

Clark, L. P., Millet, D. B. & Marshall, J. D. (2014). National patterns in environmental injustice and inequality: Outdoor NO2 air pollution in the United States. *PloS one*, *9*(4), e94431.

Clark, W. A. V. & Coulter, R. (2015). Who wants to move? The role of neighbourhood change. *Environment and Planning A*, *47*(12), 2683–2709.

Cliff, A. D. & Ord, J. K. (1981). *Spatial Processes: Models & Applications*. London: Pion.

Cook, S. J., Hays, J. C. & Franzese, R. J. (2015). *Model Specification and Spatial Interdependence: Paper prepared for the 2015 Summer Methods Meeting*. Retrieved 09.07.2017, from `http://www.sas.rochester.edu/psc/polmeth/papers/Cook_Hays_Franzese.pdf`

Corrado, L. & Fingleton, B. (2012). Where is the economics in spatial econometrics? *Journal of Regional Science*, *52*(2), 210–239.

Crowder, K. & Downey, L. (2010). Inter-neighborhood migration, race, and environmental hazards: Modeling micro-level processes of environmental inequality. *American Journal of Sociology*, *115*(4), 1110–1149.

Crowder, K., Hall, M. & Tolnay, S. E. (2011). Neighborhood immigration and native out-migration. *American Sociological Review*, *76*(1), 25–47.

Crowder, K., Pais, J. & South, S. J. (2012). Neighborhood diversity, metropolitan constraints, and household migration. *American Sociological Review*, *77*(3), 325–353.

Crowder, K. & South, S. J. (2005). Race, class, and changing patterns of migration between poor and nonpoor neighborhoods. *American Journal of Sociology*, *110*(6), 1715–1763.

Crowder, K., South, S. J. & Chavez, E. (2006). Wealth, race, and inter-neighborhood migration. *American Sociological Review*, *71*(1), 72–94.

Daniels, G. & Friedman, S. (1999). Spatial inequality and the distribution of industrial toxic releases: Evidence from the 1990 TRI. *Social Science Quarterly*, *80*(2), 244–262.

Davidson, P. & Anderton, D. L. (2000). Demographics of dumping II: A national environmental equity survey and the distribution of hazardous materials handlers. *Demography*, *37*(4), 461–466.

Depro, B., Timmins, C. & O'Neil, M. (2015). White flight and coming to the nuisance: Can residential mobility explain environmental injustice? *Journal of the Association of Environmental and Resource Economists*, *2*(3), 439–468.

Diekmann, A. & Meyer, R. (2010). Demokratischer Smog? Eine empirische Untersuchung zum Zusammenhang zwischen Sozialschicht und Umweltbelastungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *62*(3), 437–457.

Dokshin, F. A. (2016). Whose backyard and what's at issue? Spatial and ideological dynamics of local opposition to fracking in New York State, 2010 to 2013. *American Sociological Review*, *81*(5), 921–948.

Downey, L. (1998). Environmental injustice: Is race or income a better predictor? *Social Science Quarterly*, *79*(4), 766–778.

Downey, L. (2003). Spatial measurement, geography, and urban racial inequality. *Social Forces*, *81*(3), 937–952.

Downey, L. (2005). The unintended significance of race: Environmental racial inequality in Detroit. *Social Forces*, *83*(3), 971–1007.

Downey, L. (2006a). Environmental racial inequality in Detroit. *Social Forces*, *85*(2), 771–796.

Downey, L. (2006b). Using geographic information systems to reconceptualize spatial relationships and ecological context. *American Journal of Sociology*, *112*(2), 567–612.

Downey, L. (2007). US metropolitan-area variation in environmental inequality outcomes. *Urban Studies*, *44*(5-6), 953–977.

Downey, L., Dubois, S., Hawkins, B. & Walker, M. (2008). Environmental inequality in metropolitan America. *Organization & Environment*, *21*(3), 270–294.

Downey, L. & Hawkins, B. (2008). Race, income, and environmental inequality in the United States. *Sociological Perspectives*, *51*(4), 759–781.

Drever, A. & Clark, W. (2002). Gaining access to housing in Germany: The foreign-minority experience. *Urban Studies*, *39*(13), 2439–2453.

Drukker, D. M., Egger, P. & Prucha, I. R. (2013). On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econometric Reviews*, *32*(5-6), 686–733.

Dubin, R. (2009). Spatial weights. In A. S. Fotheringham & P. Rogerson (Eds.), *The Sage Handbook of Spatial Analysis* (pp. 125–158). Los Angeles and London: Sage.

Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Berlin and Heidelberg: Springer.

Elliott, J. R. & Frickel, S. (2013). The historical nature of cities: A study of urbanization and hazardous waste accumulation. *American Sociological Review*, *78*(4), 521–543.

Elliott, J. R. & Frickel, S. (2015). Urbanization as socioenvironmental succession: The case of hazardous industrial site accumulation. *American Journal of Sociology*, *120*(6), 1736–1777.

Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138.

Environmental Protection Agency. (2015). *EPA's Risk-Screening Environmental Indicators (RSEI) Methodology. RSEI Version 2.3.4.* Washington, DC. Retrieved 11.02.2017, from https://www.epa.gov/sites/production/files/2015-12/documents/rsei_methodology_v2.3.4.pdf

European Commission. (2006). *Guidance Document for the Implementation of the European PRTR.* Retrieved 02.05.2016, from http://prtr.ec.europa.eu/docs/en_prtr.pdf

Farber, S. (1998). Undesirable facilities and property values: A summary of empirical studies. *Ecological Economics*, *24*(1), 1–14.

Flacke, J., Schüle, S. A., Köckler, H. & Bolte, G. (2016). Mapping environmental inequalities relevant for health for informing urban planning interventions - A case study in the city of Dortmund, Germany. *International Journal of Environmental Research and Public Health*, *13*(7), 711.

Florax, R., Folmer, H. & Rey, S. J. (2003). Specification searches in spatial econometrics: The relevance of Hendry's methodology. *Regional Science and Urban Economics*, *33*(5), 557–579.

Franzese, R. J. & Hays, J. C. (2007). Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. *Political Analysis*, *15*(2), 140–164.

Friedrichs, J., Galster, G. & Musterd, S. (2003). Neighbourhood effects on social opportunities: The European and American research and policy context. *Housing Studies*, *18*(6), 797–806.

Funderburg, R. & Laurian, L. (2015). Bolstering environmental (in)justice claims with a quasi-experimental research design. *Land Use Policy*, *49*, 511–526.

Gascon, M., Sánchez-Benavides, G., Dadvand, P., Martínez, D., Gramunt, N., Gotsens, X., ... Nieuwenhuijsen, M. (2018). Long-term exposure to residential green and blue spaces and anxiety and depression in adults: A cross-sectional study. *Environmental Research*, *162*, 231–239.

Gibbons, S. & Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, *52*(2), 172–191.

Gibbons, S., Overman, H. G. & Patacchini, E. (2015). Spatial methods. In G. Duranton, J. V. Henderson & W. C. Strange (Eds.), *Handbook of Regional and Urban Economics* (Vol. 5, pp. 115–168). Amsterdam: Elsevier.

Girard, D. A. (1989). A fast 'Monte-Carlo cross-validation' procedure for large least squares problems with noisy data. *Numerische Mathematik*, *56*(1), 1–23.

Grant, D., Trautner, M. N., Downey, L. & Thiebaud, L. (2010). Bringing the polluters back in: Environmental inequality and the organization of chemical production. *American Sociological Review*, *75*(4), 479–504.

Gresch, C. & Kristen, C. (2011). Staatsbürgerschaft oder Migrationshintergrund? Ein Vergleich unterschiedlicher Operationalisierungsweisen am Beispiel der Bildungsbeteiligung. *Zeitschrift für Soziologie*, *40*(3), 208–227.

Halleck Vega, S. & Elhorst, J. P. (2015). The SLX model. *Journal of Regional Science*, *55*(3), 339–363.

Hamilton, J. T. (1995). Testing for environmental racism: Prejudice, profits, political power? *Journal of Policy Analysis and Management*, *14*(1), 107–132.

Hanna, B. G. (2007). House values, incomes, and industrial pollution. *Journal of Environmental Economics and Management*, *54*(1), 100–112.

Havard, S., Deguen, S., Zmirou-Navier, D., Schillinger, C. & Bard, D. (2009). Traffic-related air pollution and socioeconomic status. *Epidemiology*, *20*(2), 223–230.

Hersh, R. (1995). *Race and Industrial Hazards: A Historical Geography of the Pittsburgh Region, 1900-1990.* Washington DC: Resources for the Future.

Hipp, J. R. & Lakon, C. M. (2010). Social disparities in health: Disproportionate toxicity proximity in minority communities over a decade. *Health & Place*, *16*(4), 674–683.

Hox, J. & Wijngaards-de Meij, L. (2015). The multilevel regression model. In H. Best & C. Wolf (Eds.), *The Sage Handbook of Regression Analysis and Causal Inference* (pp. 133–152). London: Sage.

Huang, J., Lu, X. X. & Sellers, J. M. (2007). A global comparative analysis of urban form: Applying spatial metrics and remote sensing. *Landscape and Urban Planning*, *82*(4), 184–197.

Hunter, L. M., White, M. J., Little, J. S. & Sutton, J. (2003). Environmental hazards, migration, and race. *Population and Environment*, *25*(1), 23–39.

Kabisch, N. & Haase, D. (2014). Green justice or just green? Provision of urban green spaces in Berlin, Germany. *Landscape and Urban Planning*, *122*, 129–139.

Kalter, F. & Granato, N. (2007). Educational hurdles on the way to structural assimilation in Germany. In A. F. Heath, S. Y. Cheung & S. N. Smith (Eds.), *Unequal Chances* (Vol. 137, pp. 271–319). Oxford: Oxford Univ. Press.

Kelejian, H. H. & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, *17*(1), 99–121.

Kelejian, H. H. & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, *157*(1), 53–67.

Kim, Y., Campbell, H. & Eckerd, A. (2014). Residential choice constraints and environmental justice. *Social Science Quarterly*, *95*(1), 40–56.

Kling, J. R., Liebman, J. B. & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), 83–119.

Kohlhuber, M., Mielck, A., Weiland, S. K. & Bolte, G. (2006). Social inequality in perceived environmental exposures in relation to housing conditions in Germany. *Environmental Research*, *101*(2), 246–255.

Krieg, E. J. (1995). A socio-historical interpretation of toxic waste sites. *American Journal of Economics and Sociology*, *54*(1), 1–14.

Krysan, M., Couper, M. P., Farley, R. & Forman, T. A. (2009). Does race matter in neighborhood preferences? Results from a video experiment. *American Journal of Sociology*, *115*(2), 527–559.

Landrigan, P. J., Fuller, R., Acosta, N. J. R., Adeyi, O., Arnold, R., Basu, N., … Zhong, M. (2018). The Lancet Commission on pollution and health. *The Lancet*, *391*(10119), 462–512.

Laurian, L. & Funderburg, R. (2014). Environmental justice in France? A spatio-temporal analysis of incinerator location. *Journal of Environmental Planning and Management*, *57*(3), 424–446.

Lee, L.-f. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, *72*(6), 1899–1925.

Legewie, J. & Schaeffer, M. (2016). Contested boundaries: Explaining where ethnoracial diversity provokes neighborhood conflict. *American Journal of Sociology*, *122*(1), 125–161.

LeSage, J. P. (2014). What regional scientists need to know about spatial econometrics. *The Review of Regional Studies*, *44*(1), 13–32.

LeSage, J. P. & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Boca Raton: CRC Press.

LeSage, J. P. & Pace, R. K. (2014). The biggest myth in spatial econometrics. *Econometrics*, *2*(4), 217–249.

LeSage, J. P. & Pace, R. K. (2017). Spatial econometric Monte Carlo studies: Raising the bar. *Empirical Economics*, in press.

Levinson, A. (2012). Valuing public goods using happiness data: The case of air quality. *Journal of Public Economics*, *96*(9-10), 869–880.

Lichter, D. T., Parisi, D. & Taquino, M. C. (2015). Toward a new macro-segregation? Decomposing segregation within and between metropolitan cities and suburbs. *American Sociological Review*, *80*(4), 843–873.

Liebe, U., Preisendörfer, P. & Meyerhoff, J. (2010). To pay or not to pay: Competing theories to explain individuals' willingness to pay for public environmental goods. *Environment and Behavior*, *43*(1), 106–130.

Logan, J. R. (2012). Making a place for space: Spatial thinking in social science. *Annual Review of Sociology*, *38*, 507–524.

Logan, J. R. & Alba, R. D. (1993). Locational returns to human capital: Minority access to suburban community resources. *Demography*, *30*(2), 243–268.

Logan, J. R., Zhang, W. & Alba, R. D. (2002). Immigrant enclaves and ethnic communities in New York and Los Angeles. *American Sociological Review*, *67*(2), 299–322.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.

Massey, D. S. (1990). Social structure, household strategies, and the cumulative causation of migration. *Population Index*, *56*(1), 3–26.

Massey, D. S. & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, *67*(2), 281–315.

Massey, D. S. & Denton, N. A. (1989). Hypersegregation in U.S. metropolitan areas: Black and hispanic segregation along five dimensions. *Demography*, *26*(3), 373.

Massey, D. S. & Denton, N. A. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, Mass: Harvard Univ. Press.

Meyer, R. & Liebe, U. (2010). Are the affluent prepared to pay for the planet? Explaining willingness to pay for public and quasi-private environmental goods in Switzerland. *Population and Environment*, *32*(1), 42–65.

Mielck, A. (2004). Unterschiede bei Lärmbelastung und Luftverschmutzung nach dem Haushaltseinkommen. In G. Bolte & A. Mielck (Eds.), *Umweltgerechtigkeit* (pp. 139–154). Weinheim: Juventa.

Mohai, P., Lantz, P. M., Morenoff, J., House, J. S. & Mero, R. P. (2009). Racial and socioeconomic disparities in residential proximity to polluting industrial facilities: Evidence from the Americans' Changing Lives study. *American Journal of Public Health*, *99*(S3), 649–656.

Mohai, P., Pellow, D. & Roberts, J. T. (2009). Environmental justice. *Annual Review of Environment and Resources*, *34*(1), 405–430.

Mohai, P. & Saha, R. (2006). Reassessing racial and socioeconomic disparities in environmental justice research. *Demography*, *43*(2), 383–399.

Mohai, P. & Saha, R. (2007). Racial inequality in the distribution of hazardous waste: A national-level reassessment. *Social Problems*, *54*(3), 343–370.

Mohai, P. & Saha, R. (2015a). Which came first, people or pollution? A review of theory and evidence from longitudinal environmental justice studies. *Environmental Research Letters*, *10*(12), 125011.

Mohai, P. & Saha, R. (2015b). Which came first, people or pollution? Assessing the disparate siting and post-siting demographic change hypotheses of environmental injustice. *Environmental Research Letters*, *10*(11), 115008.

Morello-Frosch, R. & Jesdale, B. M. (2005). Separate and unequal: Residential segregation and estimated cancer risks associated with ambient air toxics in U.S. metropolitan areas. *Environmental Health Perspectives*, *114*(3), 386–393.

Morello-Frosch, R., Pastor, M. & Sadd, J. L. (2001). Environmental justice and Southern California's "riskscape". *Urban Affairs Review*, *36*(4), 551–578.

Moreno-Jiménez, A., Cañada-Torrecilla, R., Vidal-Domínguez, M. J., Palacios-García, A. & Martínez-Suárez, P. (2016). Assessing environmental justice through potential exposure to air pollution: A socio-spatial analysis in Madrid and Barcelona, Spain. *Geoforum*, *69*, 117–131.

Morgan, S. L. & Winship, C. (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge Univ. Press.

Mur, J. & Angulo, A. (2009). Model selection strategies in a spatial setting: Some additional results. *Regional Science and Urban Economics*, *39*(2), 200–213.

Musterd, S. (2005). Social and ethnic segregation in Europe: Levels, causes, and effects. *Journal of Urban Affairs*, *27*(3), 331–348.

Neumayer, E. & Plümper, T. (2016). W. *Political Science Research and Methods*, *4*(01), 175–193.

Oakes, J. M., Anderton, D. L. & Anderson, A. B. (1996). A longitudinal analysis of environmental equity in communities with hazardous waste facilities. *Social Science Research*, *25*(2), 125–148.

Ondrich, J., Ross, S. & Yinger, J. (2003). Now you see it, now you don't: Why do real estate agents withhold available houses from black customers? *Review of Economics and Statistics*, *85*(4), 854–873.

Ord, J. K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, *70*(349), 120–126.

Pace, R. K. & LeSage, J. P. (2010). Omitted variable biases of OLS and spatial lag models. In A. Páez, J. Gallo, R. N. Buliung & S. Dall'erba (Eds.), *Progress in Spatial Analysis* (pp. 17–28). Berlin and Heidelberg: Springer.

Padilla, C. M., Kihal-Talantikite, W., Vieira, V. M., Rossello, P., Le Nir, G., Zmirou-Navier, D. & Deguen, S. (2014). Air quality and social deprivation in four French metropolitan areas: A localized spatio-temporal environmental inequality analysis. *Environmental Research*, *134*, 315–324.

Pager, D. & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, *34*(1), 181–209.

Pais, J., Crowder, K. & Downey, L. (2014). Unequal trajectories: Racial and class differences in residential exposure to industrial hazard. *Social Forces*, *92*(3), 1189–1215.

Pastor, M., Morello-Frosch, R. & Sadd, J. L. (2005). The air is always cleaner on the other side: Race, space, and ambient air toxics exposures in California. *Journal of Urban Affairs*, *27*(2), 127–148.

Pastor, M., Sadd, J. & Hipp, J. R. (2001). Which came first? Toxic facilities, minority move-in, and environmental justice. *Journal of Urban Affairs*, *23*(1), 1–21.

Pastor, M., Sadd, J. L. & Morello-Frosch, R. (2002). Who's minding the kids? Pollution, public schools, and environmental justice in Los Angeles. *Social Science Quarterly*, *83*(1), 263–280.

Pellow, D. N. & Nyseth Brehm, H. (2013). An environmental sociology for the twenty-first century. *Annual Review of Sociology*, *39*(1), 229–250.

Piketty, T. & Saez, E. (2014). Inequality in the long run. *Science*, *344*(6186), 838–843.

Pinkse, J. & Slade, M. E. (2010). The future of spatial econometrics. *Journal of Regional Science*, *50*(1), 103–117.

Preisendörfer, P. (2014). Umweltgerechtigkeit: Von sozial-räumlicher Ungleichheit hin zu postulierter Ungerechtigkeit lokaler Umweltbelastungen. *Soziale Welt*, *65*(1), 25–45.

Raddatz, L. & Mennis, J. (2013). Environmental justice in Hamburg, Germany. *The Professional Geographer*, *65*(3), 495–511.

Rathelot, R. & Safi, M. (2014). Local ethnic composition and natives' and immigrants' geographic mobility in France, 1982-1999. *American Sociological Review*, *79*(1), 43–64.

Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R. & Bischoff, K. (2008). The geographic scale of metropolitan racial segregation. *Demography*, *45*(3), 489–514.

Reardon, S. F. & O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology*, *34*(1), 121–162.

Richardson, E. A., Shorty, N. K. & Mitchell, R. J. (2010). The mechanism behind environmental inequality in Scotland: Which came first, the deprivation or the landfill? *Environment and Planning A*, *42*(1), 223–240.

Ringquist, E. J. (2005). Assessing evidence of environmental inequities: A meta-analysis. *Journal of Policy Analysis and Management*, *24*(2), 223–247.

Roberts, J. D., Voss, J. D. & Knight, B. (2014). The association of ambient air pollution and physical inactivity in the United States. *PloS one*, *9*(3), e90143.

Rüttenauer, T. (2018). Neighbours matter: A nation-wide small-area assessment of environmental inequality in Germany. *Social Science Research*, *70*, 198–211.

Sager, L. (2012). Residential segregation and socioeconomic neighbourhood sorting: Evidence at the micro-neighbourhood level for migrant groups in Germany. *Urban Studies*, *49*(12), 2617–2632.

Saha, R. & Mohai, P. (2005). Historical context and hazardous waste facility siting: Understanding temporal patterns in Michigan. *Social Problems*, *52*(4), 618–648.

Sampson, R. J., Morenoff, J. D. & Earls, F. (1999). Beyond social capital: Spatial dynamics of collective efficacy for children. *American Sociological Review*, *64*(5), 633–660.

Sampson, R. J., Morenoff, J. D. & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology*, *28*(1), 443–478.

Sass, V., Kravitz-Wirtz, N., Karceski, S. M., Hajat, A., Crowder, K. & Takeuchi, D. (2017). The effects of air pollution on individual psychological distress. *Health & Place*, *48*, 72–79.

Schelling, T. C. (1978). *Micromotives and Macrobehavior.* New York: Norton.

Schmitt, M. (2013). Subjective well-being and air quality in Germany. *Schmollers Jahrbuch*, *133*(2), 275–286.

Schwarz, N. (2010). Urban form revisited—Selecting indicators for characterising European cities. *Landscape and Urban Planning*, *96*(1), 29–47.

Schweitzer, L. & Stephenson, M. (2007). Right answers, wrong questions: Environmental justice as urban research. *Urban Studies*, *44*(2), 319–337.

Semyonov, M., Grodzeisky, A. & Glikman, A. (2012). Neighborhood ethnic composition and resident perceptions of safety in European countries. *Social Problems*, *59*(1), 117–135.

Shaikh, S. L. & Loomis, J. B. (1999). An investigation into the presence and causes of environmental inequity in Denver, Colorado. *The Social Science Journal*, *36*(1), 77–92.

Sieg, H., Smith, V. K., Banzhaf, H. S. & Walsh, R. P. (2004). Estimating the general equilibrium benefits of large changes in spatially delineated public goods. *International Economic Review*, *45*(4), 1047–1077.

Solon, G., Haider, S. J. & Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources*, *50*(2), 301–316.

South, S. J. & Crowder, K. (1997). Escaping distressed neighborhoods: Individual, community, and metropolitan influences. *American Journal of Sociology*, *102*(4), 1040–1084.

South, S. J., Huang, Y., Spring, A. & Crowder, K. (2016). Neighborhood attainment over the adult life course. *American Sociological Review*, *81*(6), 1276–1304.

Statistische Ämter des Bundes und der Länder. (2015). *Zensus 2011: Metoden und Berichte.* Wiesbaden. Retrieved 20.03.2017, from `https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaetze_Archiv/2015_06_MethodenUndVerfahren.pdf`

Szasz, A. & Meuser, M. (2000). Unintended, inexorable: The production of environmental inequalities in Santa Clara County, California. *American Behavioral Scientist*, *43*(4), 602–632.

Tiebout, C. M. (1956). A pure theory of local expenditures. *The Journal of Political Economy*, *64*(5), 416–424.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240.

Turner, M. A. & Ross, S. (2005). How racial discrimination affects the search for housing. In X. N. De Souza Briggs (Ed.), *The Geography of Opportunity* (pp. 81–100). Washington DC: Brookings Institution Press.

United Church of Christ - Commission for Racial Justice. (1987). *Toxic Wastes and Race in the United States: A National Report on the Racial and Socioeconomic Characteristics of of Communities with Hazardous Waste Sites.* New York: United Church of Christ.

Wagner, G. G., Frick, J. R. & Schupp, J. (2007). The German Socio-Economic Panel study (SOEP) - Evolution, scope and enhancements. *Schmollers Jahrbuch*, *127*, 139–169.

Ward, M. D. & Gleditsch, K. S. (2008). *Spatial Regression Models* (Vol. 155). Thousand Oaks: Sage.

Wilson, W. J. (1978). *The Declining Significance of Race: Blacks and Changing American Institutions*. Chicago: Univ. of Chicago Press.

Wilson, W. J. (1987). *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago: Univ. of Chicago Press.

Wolverton, A. (2009). Effects of socio-economic and input-related factors on polluting plants' location decisions. *The B.E. Journal of Economic Analysis & Policy*, *9*(1).

Wolverton, A. (2012). The role of demographic and cost-related factors in determining where plants locate: A tale of two Texas cities. In H. S. Banzhaf (Ed.), *The Political Economy of Environmental Justice* (pp. 199–222). Palo Alto: Stanford Univ. Press.

Wong, D. (2009). The modifiable areal unit problem (MAUP). In A. S. Fotheringham & P. Rogerson (Eds.), *The Sage Handbook of Spatial Analysis* (pp. 105–124). Los Angeles and London: Sage.

Yinger, J. (1986). Measuring racial discrimination with fair housing audits: Caught in the act. *The American Economic Review*, *76*(5), 881–893.

Zhang, X., Zhang, X. & Chen, X. (2017). Happiness in the air: How does a dirty sky affect mental health and subjective well-being? *Journal of Environmental Economics and Management*, *85*, 81–94.

Curriculum Vitae

CONTACT

Department of Social Sciences
TU Kaiserslautern
Erwin-Schroedinger-Str. 57
D-67663 Kaiserslautern
Phone: +49 (0) 631 205 5785
Email: ruettenauer@sowi.uni-kl.de

ACADEMIC POSITIONS

2015–present	Research Assistant, Chair of Sociology and Social Stratification (Prof. Dr. H. Best). TU Kaiserslautern
Apr.–Sep. 2015	Research Assistant, Chair of Quantitative Methods of Empirical Social Research (Prof. Dr. H. Best). JMU Würzburg

EDUCATION

Aug. 2016	Spatial Econometrics. Summer School in Social Science Data Analysis. University of Essex
2015	Master of Arts, Sociology (Minor: Statistics). LMU Munich
Aug.–Dec. 2013	Globalization Program. Venice International University
2012	Bachelor of Arts, Sociology (Minor: Economics). LMU Munich
2009	Abitur. Lise Meitner Gymnasium Unterhaching

PUBLICATIONS

Rüttenauer, T. (2018). Bringing urban space back in: A multilevel analysis of environmental inequality in Germany. *Urban Studies*, OnlineFirst.

Rüttenauer, T. (2018). Neighbours matter: A nation-wide small-area assessment of environmental inequality in Germany. *Social Science Research, 70*, 198-211.

Best, H., & Rüttenauer, T. (2018). How selective migration shapes environmental inequality in Germany: Evidence from micro-level panel data. *European Sociological Review, 34*(1), 52-63.