ESSAYS IN PUBLIC ECONOMICS

INAUGURAL DISSERTATION

zur Erlangung des akademischen Grades
eines Doktors der Wirtschafts- und
Sozialwissenschaften (Dr. rer. pol.)
an der Wirtschafts- und Sozialwissenschaftlichen
Fakultät
der Universität Potsdam

vorgelegt von

Niklas Gohl-Greenaway

Berlin, 2023

Erstgutachter:
Prof. Dr. Rainald Borck.

Zweitgutachter:
Prof. Dr. Peter Haan.

Drittgutachter:
Prof. Dr. Arne Uhlendorff.

Eingereicht im Mai 2023.

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-60902
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-609026

Acknowledgements

I want to thank my two supervisors Rainald Borck and Peter Haan for their constant support, time, advice and valuable feedback throughout the writing process of this thesis. I benefited a lot from their friendly support and extensive mentorship. Further, both are co-authors and their encouragement and guidance to develop and apply my own ideas and research agenda greatly helped in writing this thesis. I am also very grateful to Arne Uhlendorff who throughout my research stay in Paris provided great support, advice and feedback, and for taking on the role as third reviewer. I would also like to thank Felix Weinhardt, Marco Caliendo and all other members of the PRSE, CEPA, DIW, BSE and CREST, who provided helpful comments, feedback and encouragement.

I would also like to thank my colleagues and PhD students in Potsdam and at the Berlin School of Economics. Specifically, I would like to thank Philipp Schrauth, Felix Degenhardt, Anne Schrenker, and Sebastian Becker who frequently discussed my research with me and shared their experiences, so that I did not feel so alone with the struggles, pressures, and challenges that come with being a PhD student. A special thanks to Philipp Schrauth with whom I co-authored a paper, frequently exchanged research ideas, help and encouragement. I greatly value his friendship that developed throughout this process.

I also want to commemorate Nedim Okan who left this world far too early and is greatly missed.

Last but not least, I want to thank my family. My parents and brother were a constant source of support that I could always rely on. Above all, I want to thank my wife Beth whose patience, love and support were key in achieving this. Thank you!

Collaborations and Publications with Co-authors

The five chapters of this dissertation are based on papers, which have been published or submitted elsewhere. The following list shows the co-authors of the respective paper as well as the place and type of publication and submission status.

`Chapter 2`:
`Fürstenau, E., Gohl, N., Haan, P., Weinhardt, F. (2023).` Working Life and Human Capital Investment: Causal Evidence from a Pension Reform.
Labour Economics, Volume 84, 102426, https://doi.org/10.1016/j.labeco.2023.102426

`Chapter 3`:
`Gohl, N (2023).` Working Longer, Working Stronger? The Forward-Looking Effects of Increasing the Retirement Age on (Un)employment Behaviour. *CEPA Discussion Papers, No. 63. Submitted to Journal of Human Resources.*

`Chapter 4`:
`Gohl, N., Haan, P., Michelsen, C., Weinhardt, F. (2022).` House Price Expectations. *DIW Discussion Papers 1994. Revise and Resubmit at Journal of Economic Behavior and Organization.*

`Chapter 5`:

`Borck, R., Gohl, N. (2021).` Gentrification and Affordable Housing Policies. *Center for Economic Policy Analysis, No. 39. Submitted to International Economic Review.*

`Chapter 6`:

`Gohl, N., Schrauth, P. (2022).` Ticket to paradise? The Effect of a Public Transport Subsidy on Air Quality. *CEPA Discussion Papers, No. 50. Revise and Resubmit at Journal of Urban Economics: Insights.*

Contents

Chapter 1

Introduction

Motivation

Understanding how public policies impact individual and household behaviour is at the core of *Public Economics* and a fundamental tenet for research-based policy-making. As former OECD Secretary-General Angel Gurría explained:

> *Improving the quality of our lives should be the ultimate target of public policies. But public policies can only deliver best fruit if they are based on reliable tools to measure the improvement they seek to produce in our lives.*
> —Angel Gurría[1]

To that end, this cumulative dissertation uses economic theory and state-of-the-art micro-econometric tools and evaluation methods to analyse public policies and their impact on welfare and individual behaviour. In particular, it focuses on policies in two distinct areas that represent fundamental societal challenges in the 21st century: the ageing of society and life in densely-populated urban agglomerations. Together, these areas shape important financial decisions in a person's life, impact welfare, and are driving forces behind many of the challenges OECD members currently face, as will be described in

[1] Introductory remarks by Angel Gurría, OECD Secretary-General, delivered at OECD Forum 2011, first Session on Measuring Progress, see (Gurría, 2011)

more detail below. Understanding and measuring the impact of public policies designed to address these challenges therefore, in the words of Angel Gurría, helps to *deliver best fruit* and improve the quality of lives. Each individual research area, the policies and topics analyzed are presented below.

Ageing Societies

Ageing societies pose a challenge to the long-term fiscal stability of many developed countries. More precisely, increases in life expectancy and the average age of the population in most OECD member countries have substantially raised the fiscal burden of public pension systems (OECD, 2019). In the 1980s, for example, there were two people older than 65 for every 10 people of working age in the OECD. That number is projected to increase by as much as approximately six people by 2060 in some member states (OECD, 2019).

As a response, many member states have implemented a range of pension reforms that aim to relieve their pay-as-you go pension systems, by reducing the number of recipients and increasing the number of taxpayers. In particular, the statutory retirement age has been raised substantially over the past three decades, meaning that people are facing longer working lives. This, in turn, may change the employment behaviour and human capital decisions of young and middle-aged people who are now confronted with a longer work horizon - an area of research relatively understudied in economic research on pension reforms. However, in order to promote employment, and for the increase in the statutory retirement age to actually take pressure off pay-as-you-go pension systems, it is crucial to understand how individuals adjust their behaviour in response to these reforms. The first two studies of this dissertation therefore analyse how individuals react to increases in the statutory retirement age by empirically assessing the impact of two pension reforms that raised the retirement age in Germany. Both papers focus on forward-looking responses to pension reforms, i.e., responses of individuals with a relatively long remaining work life at the time the pension reforms were announced.

Living in Urban Agglomerations

In Germany and across OECD member states, approximately 75% of all citizens live in cities with inhabitants of more than 50,000 (OECD, 2020). A large area of research, particularly in the field of *Urban Economics*, has focused on the advantages of living in cities, highlighting among others a positive correlation between population density, productivity and wages (Glaeser, 2011). However, living in large and densely populated metropolitan areas also implies potential societal challenges. This dissertation focuses on two of those: affordable housing and air pollution both of which will be introduced in more detail in the following.

Affordable Housing Within the housing market, sharp increases in housing prices and rents over the last decades have sparked widespread concern of residents and policymakers about the affordability of housing, which politicians have termed "the new social issue of our time" (Sagner et al., 2020). This is particularly problematic in large cities in developed countries, spurred by rising immigration and gentrification (Knoll et al., 2017). Policymakers have sought to address these patterns by introducing policies such as rent control, re-zoning measures and housing subsidies (Diamond et al., 2019; Autor et al., 2014).

While much of the previous research has focused primarily on the short- to medium-term effects of housing policies and the factors that influence housing decisions (Armona et al., 2019; Kholodilin et al., 2016, 2022), this dissertation focuses on long-term aspects of such changes in the housing sector. First, Chapter 4 quantifies short term and long term house price expectations of individuals, which are a key component for fully understanding housing and tenure decisions. In particular, house price expectations and subjective beliefs may be helpful to fully understand individual behaviour and explain observed patterns of rising housing prices, especially in urban agglomerations.

Second, Chapter 5 analyses the long term distributional welfare effects of gentrification patterns and a range of affordable housing policy measures by developing a spatial sorting model and calibrating it for the city of Berlin, Germany. Crucially, the study analyses the long run equilibrium and re-distributive effects of gentrification, rent control and other housing policies such as rezoning measures and housing subsidies.

Air Pollution Last, exposure to air pollution sparked by motorized traffic and congestion in particular in large metropolitan areas is a key challenge for most countries across the world, and has been shown to substantially impact individuals' health (Currie and Walker, 2011; Knittel et al., 2016; Margaryan, 2021). Governments across countries have increasingly sought to address this by, for example, creating low emission zones in large cities (Gehrsitz, 2017). Most recently, countries such as Germany or Spain have introduced public transport subsidies in the hope to provide affordable means of commuting and incentives for individuals to switch to less polluting means of transportation (The Guardian, 2022). Chapter 6 of this dissertation explicitly analyses the impact of such a large scale public transport subsidy, temporarily introduced in Germany, on air pollution, and analyses among others whether the effect varies between urban and rural areas.

Outline and Summary

Chapters 2 through 6 are self-contained studies, distinct from one another, each with its own research question and contribution. Table 1.1 gives an overview of each chapter and its main components. The remainder of this introduction presents each chapter, its research focus, methodology, and respective contribution in more detail. All papers are connected in the sense that they evaluate policy interventions or provide means to better understand individual behaviour in the context of policy decision making in the two areas described above. Further, all chapters use state-of-the-art empirical evaluation methods such as randomized controlled trials (RCT), quasi-experiments and quantitative modelling.

Chapter 2 - Working Life and Human Capital Investment: Causal Evidence from a Pension Reform Chapter 2 is co-authored with Peter Haan, Elisabeth Kurz and Felix Weinhardt and focuses on the link between human capital theory and pension

Table 1.1: Overview of Chapters

	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
Title	Working Life and Human Capital Investment: Causal Evidence from a Pension Reform	Working Longer, Working Stronger? The Forward-Looking Effects of Increasing the Retirement Age on (Un)employment Behaviour	House Price Expectations	Gentrification and Affordable Housing Policies	Ticket to Paradise? The Effect of a Public Transport Subsidy on Air Quality
Co-Author(s)	Peter Haan, Elizabeth Kurz, Felix Weinhardt		Peter Haan, Claus Michelsen, Felix Weinhardt	Rainald Borck	Philipp Schrauth
Primary outcome(s)	On-the-job training	Employment, unemployment	House price expectations	House prices, dwelling size, welfare	Air pollution
Data	Cross-sectional household survey data with information on on-the-job training (German Microcensus)	Administrative panel data on employment biographies (SIAB data), survey data on retirement expectations (SAVE survey) and job applications/offers (IZA Evaluation Panel)	Household survey data (German Socioeconomic Panel Innovation Sample)	Granular spatial data on household income groups, geo-coded data on amenities, postcal code specific housing prices	Granular spatial data on air pollution, weather, gasoline prices.
Methodology	Regression Discontinuity Design, Difference-in-Discontinuities	Difference-in-Differences with multi-valued treatment variable, Regression Discontinuity Design	Descriptives, OLS, Randomized Controlled Trial	Calibration of quantitative spatial sorting model plus counterfactual analysis, Instrumental Variable approach (IV)	Difference-in-Differences

reforms. Human capital theory, which goes back to Ben-Porath (1967) and Becker (1962), states that the longer the period in which individuals can earn returns on their investment, the more they should invest in their human capital and, for example, participate in training measures. We argue that raising the statutory retirement age extends the period in which individuals can earn returns on their human capital. Pension reforms that announce a postponement of the retirement age before affected individuals actually retire may therefore lead to an increase in training among those currently still employed, improving employability and the ability of individuals to work longer.

To investigate whether this is indeed the case, this paper uses a quasi-experiment triggered by a cohort-specific pension reform in Germany that increased the early retirement age for women born after 1951 by three years. More specifically, we use Regression Discontinuity Designs (RDD) and Difference-in-Discontinuity approaches to compare women born in the end of 1951 with women born at the beginning of 1952. In this way, we compare women who are highly similar in age and other characteristics apart from their legal retirement age, which allows us to isolate causal estimates.

Using data from the German Microcensus, our main finding is that an increase in the working life causally increases human capital investment: we show that the probability to participate in training increases by about 4.4 percentage points. Depending on the specification, the point estimates correspond to a relative increase of about 28.8%, which suggests that an increase in the working life has sizable effects on training participation. Crucially, we find that the effect is exclusively driven by women with a relatively high degree of education.

Our findings have important implications for the policy debate on pension reform. In particular, they show that highly educated women are able to adapt to a longer working life by investing more in training in advance to retirement. The debate usually abstracts from the dynamic human capital investment in the run-up to reaching the retirement threshold that we document, and focuses directly on employment outcomes between the old and new retirement threshold, showing that some individuals remain employed and others are in inactivity or unemployment (Geyer and Welteke, 2021). Our documented forward-looking training responses are important to fully analyse these patterns, as they may shape these employment responses and thus help to understand why some individuals are able to work longer while others are in unemployment or inactivity between the old and new retirement age. The question of why training is limited to highly educated women remains unanswered and may be a fruitful avenue for future research. To reduce retirement income inequality and the risk of poverty in old age, it is crucial to increase training also for the low educated.

Chapter 3 - Working Longer, Working Stronger? The Forward-Looking Effects of Increasing the Retirement Age on (Un)employment Behaviour This single-authored chapter is connected to the previous chapter, as both analyse the forward-looking responses of middle-aged individuals to pension reforms. Instead of training outcomes, however, it examines forward-looking employment responses of individuals to a pension reform that gradually raised the normal retirement age (NRA) from 65 to 67 in Germany. Forward-looking employment responses to pension reforms are relatively understudied in the economics literature, and the documented empirical evidence is mixed (Carta and de Philippis, 2021; Geyer and Welteke, 2021; Hairault et al., 2010). The paper contributes to the relatively sparse literature by providing new evidence for the existence of forward-looking employment effects. Most importantly, it provides new insights into the mechanisms that produce these effects.

Using Difference-in-Differences (DiD) methods with a multi-valued treatment variable, I compare cohorts with different treatment intensities before and after the reform and use high-quality administrative and survey data to show that individuals adjust their retirement expectations and (un)employment behaviour in response to an exogenous increase in their statutory work horizon.

Specifically, the paper documents four main findings. First, I show that even small differences in the statutory working life caused by pension reforms lead to a shift in the length of individuals' subjectively expected work horizon. Second, a longer working horizon significantly increases employment probabilities of middle-aged individuals and has a negative effect on the probability of being registered as unemployed during a given year. The positive employment effects are driven by higher inflows into employment as well as lower outflows out of employment. Third, the paper provides suggestive evidence that individuals with a longer working horizon have a higher job search intensity and receive more job offers by job agency officials. Fourth, exploiting the large sample size of the administrative data set used in this study, I show that the documented employment effects are driven by women and predominantly occur in jobs with lower physical intensity, i.e., jobs that can be exercised at older ages.

Overall, the results show that pension policy, job search and active labor market policy interact. This is crucial, as understanding this interdependence is key to boosting employment among middle-aged and older workers while taking pressure off pay-as-you-go pension systems. In particular, the heterogeneous results documented in this paper highlight that pension reforms affect individuals and occupations differently in advance to reaching the shifted retirement threshold. To prevent pension reforms and the extension of working lives from exacerbating inequality in old age, policymakers can actively take into account the different responses to rising retirement ages when promoting employment among middle aged individuals through labor market policies.

Chapter 4 - House Price Expectations Chapter 4 is co-authored with Peter Haan, Claus Michelsen and Felix Weinhardt. It is the first study in this dissertation that focuses on topics in *Urban Economics*, specifically the housing market. In comparison to all other chapters in this dissertation, it is also the only study to not explicitly analyse a policy intervention. However, it studies subjective expectations of house price developments, thereby contributing to a better understanding of the functioning of housing markets, which is a pre-requisite for designing effective housing policies and understanding their impact. More generally, housing is a key component of wealth and consumption. This makes housing critical to household decisions on consumption and saving, bank lending decisions and asset pricing. Since these factors are relevant for economic growth and financial stability, housing markets are important. As a result, understanding housing markets or determinants of tenure decisions has long been in the focus of economists and policy-makers.

The chapter examines short-, medium-, and long-run price expectations in housing markets using survey data, i.e., the *German SOEP Innovation Sample*. As the main contribution, we are able to examine house price expectations over a thirty-year time horizon, while previous research has focused on a much shorter time frame, see e.g. Armona et al. (2019).

We derive and test six hypotheses about the incidence, formation, and relevance of price expectations. To do so, we additionally use data from past sale offerings, satellites, and an information RCT (Randomized Controlled Trial). As novel finding, we show that price expectations show no evidence for momentum effects in the long-term.[2] We also do not find much evidence for behavioural biases in expectations related to individual housing tenure decisions. For example, we show that owner-occupiers do not have systematically different house price expectations to renters. Confirming existing findings, we find momentum effects in the short term and that individuals use aggregate national price information to update local expectations about house prices in their neighbourhood. Lastly, we corroborate existing evidence that expectations are relevant for portfolio choice, i.e., individuals with information about house price growth would allocate relatively more financial resources towards real estate in a hypothetical investment scenario.

Chapter 5 - Gentrification and Affordable Housing Policies This chapter is jointly authored with Rainald Borck. We develop a quantitative spatial economic model to assess the impact of gentrification and several policy measures to counteract the distributional effects of gentrification on the welfare of heterogeneous households. In the model, households differ by income and sort into locations based on housing prices and amenities. Our paper is calibrated to key economic variables in Berlin, Germany and our model has non-homothetic preferences over housing and (endogenous) amenities, which leads to movement of richer households into central city, high amenity neighbourhoods.

[2]Momentum effects here describe a situation in which house prices and expectations are likely to increase if past house prices have risen in previous periods. Armona et al. (2019) document house price expectations that are consistent with short term momentum effects.

We run counterfactual simulations to assess the distributional impacts of several affordable housing policies. One of these is based on a rental cap policy that was introduced (and later repealed following a ruling of the Constitutional Court) by the state government of Berlin. In essence, the policy set maximum rents for all existing and new rental contracts with the exception of newly built housing. Our simulation shows that this policy reduces welfare for all households. What is more, not only does the policy reduce welfare for all household types, but the welfare reduction is largest in relative terms for the poorest households. While rents fall for all households, they fall most for richer households who choose to live in expensive, high amenity locations mainly in the city center, and it is there that rent control is binding. Rent control also increases segregation by making these locations even more attractive for high income households. The policy is thus regressive in the sense that it hurts poor households most, even though richer households bear some of its cost in the form of reduced rental income.

We also consider alternative measures, namely a targeted housing subsidy for low income households and two rezoning policies, which essentially increase housing construction (in one part only or the entire city). We find that both housing subsidies and rezoning increase welfare most for the poorest households. The subsidy by design redistributes from rich to poor households, and these effects are partly amplified by a reduction in segregation. Rezoning increases supply and thus reduces rents for all households. The rezoning policies also lead to a slight reduction of segregation in the city and, depending on the spatial distribution of new housing, lead to a larger relative increase in welfare for poor households.

Overall, this suggests that rent control may not only be inefficient, as has long been suggested by economic research, but also may fail to help the poor households who suffer most from gentrification and rising housing costs. In part, the effects of rent control stem from households sorting across locations, which amplifies segregation and differences in endogenous amenities to the detriment of poor households.

Chapter 6 - Ticket to Paradise? The Effect of a Public Transport Subsidy on Air Quality Chapter 6 is co-authored with Philipp Schrauth. We use the temporary implementation of the 9-Euro-Ticket (9ET) from June-August 2022, which reduced monthly public transport prices by up to 90% in Germany as a source of exogenous variation. Thus, we can causally estimate the impact on an air quality index. In particular, cheaper public transport fares can create incentives for individuals to switch from private motorized transport, a major source of air pollution, to public transport. To the best of our knowledge, our study is the first to estimate the air quality impacts of a substantial public transportation fare reduction. Our study therefore contributes to the literature that addresses transportation externalities and how to address them, particularly in spatially constrained cities that do not allow for large infrastructural interventions.

As our empirical approach, we employ a time-shifted Difference-in-Differences (DiD) design. Our treatment group is the year 2022, in which the 9ET was passed. The control group is provided by the years 2018/19 in which the subsidy was not in place. The pre-treatment period is the month of May and the post-treatment period the month of June. Thus, the implicit assumption is that the air quality trend between June 2022 and June 2018/19 should have evolved parallelly in the absence of treatment.

We find an improvement in air quality, as suggested by the Air Quality Index (AQI), by about eight percent. The effect is largest in urban areas, during working days, and where the provision of public transport is well established. We also estimate that the effect diminishes slightly over time, with air pollution levels increasing again after the policy measure expires in September 2022. Based on our estimates and using prior findings about the relation between air quality and health, we document that a reduced fare price for public transport is likely to improve health-related outcomes. Monetizing those leads us to conclude that they have the potential to recoup the actual costs of the intervention.

Chapter 2

Working Life and Human Capital Investment: Causal Evidence from a Pension Reform[1]

Abstract

In this paper, we analyse if an increase in the working life leads to more human capital investment via on-the-job training. We obtain RDD-estimates from a sharp date-of-birth cut-off across which the Early Retirement Age (ERA) increased by three years. In our preferred specification, we find that this pension reform causally increased on-the-job training by 4.4 percentage points. We explore heterogeneity and additional outcomes and show that individuals with a high initial level of education explain this effect. Our results speak to a large class of human capital models as well as policies towards extending or shortening working life.

[1]Co-authored with Peter Haan, Elizabeth Kurz and Felix Weihnardt. This paper has been published in a slightly different version as *"Working Life and Human Capital Investment: Causal Evidence from a Pension Reform" IZA Discussion Paper, No. 12891*

2.1 Introduction

Human capital theory, starting with Ben-Porath (1967) and Becker (1962), predicts that the value of human capital investment increases with the payout period of the investment.[2] This important prediction is the basis for explaining the joint increases in life expectancy and educational investments witnessed in most countries starting in the twentieth century; see e.g. Soares (2005); Cervellati and Sunde (2013). At the same time, rising life expectancy pressures pension systems in many developed economies. The standard policy-response has been to increase retirement age in order to increase working life. This reduces the number of recipients and increases the number of tax payers, thus reducing the fiscal burden. While there exists a very rich and still growing literature on effects of pension policy on labour supply, the literature on effects of such policy on human capital investment is scarce.[3]

In this paper, we show that women affected by a public pension reform change their human capital investment by increasing on-the-job training before retirement. This effect is driven by highly educated individuals and not explained by other dynamic labour supply responses related to the reform. We discuss implications of these findings for pension policy in ageing societies. Moreover, we regard our analysis as providing an empirical test of the key prediction of the human capital theory that the length of the payoff period causally affects human capital investment decisions.

For the identification, we exploit a pension reform in Germany that abolished the earliest retirement option. In Germany, working life is largely determined by state pension rules and the reform that we study effectively removed the early retirement for women at the age of 60. State pension rules not only have economic incentive effects but as documented by Seibold (2021), also in the context of Germany, they set the legal framework and social norm for the duration of working life. We therefore can interpret this pension reform as generating quasi-random variation in the duration of the payoff period, which generates

[2]Learning-by-doing is an alternative explanation for educational investments over the life-course (Killingsworth, 1982; Foster and Rosenzweig, 1995).

[3]We are only aware of three studies: Montizaan et al. (2010), Bauer and Eichenberger (2017), and Brunello and Comi (2015).

an exogenous increase in the working life.

We present a simple theoretical human capital model in the Appendix to formally derive that individuals indeed have an incentive to increase on-the-job training when their working life is prolonged.[4] The central mechanism for this human capital effect is that the returns to training increase with the remaining working life of an individual. This channel is the same as in models of initial education investment decisions and life expectancy (e.g. Ben-Porath (1967) and Becker (1962)), where decisions on human capital and working life are taken jointly. This confirms that our setting speaks to human capital theory.

The pension reform we study has two features that make it particularly well suited to provide causal evidence on the effect of working life on human capital investment. First, the pension reform abolished an important early retirement program for women born after 1951 across a sharp cut-off. Women born in 1951 and before could enter retirement at the age of 60 through this pathway. In contrast, for women born in 1952 or later, this pathway was closed; these women can enter retirement only at the age of 63, or later. This means that not only does this reform provide a sharp cut-off, it also provides large variation at the cut-off. In the context of pension reforms, this is an unusual feature, as such reforms are generally phased-in on a (birth)month-to-month basis or provide only smaller variation. As a second key feature, the pension reform was already announced in 1998 and implemented in 1999. Thus, the affected women, i.e. women born in 1952 and aged 47, still had a long remaining working life to benefit from human capital investment.

Our analysis is based on the German Microcensus. This is a representative yearly household survey that covers 1% of all German households (about 370,000 households per year). The data includes information about job-related training, which we use to measure post-schooling human capital investment. The main variable that we use measures the incidence of job-related training in the past twelve month. Importantly, the sample

[4]There exist several reasons why individuals have a general motivation to invest in training. Most importantly, empirical evidence shows that training has a positive effect on wages and on employment, see e.g. Frazis and Loewenstein (2005) and Blundell et al. (2019). Moreover, training can improve the quality of work and can have positive effects on non-pecuniary outcomes (Ruhose et al., 2019).

size of this household survey is unusually large, allowing regression discontinuity design (RDD) estimation.

Our main finding is that an increase in the working life causally increases human capital investment: we show that the probability to participate in training increases by about 4.4 percentage points. Depending on the specification, the point estimates correspond to a relative increase of about 28.8%, which suggests that an increase in the working life has sizable effects on training participation. Note that we compute these average treatment effects on the treated (ATT) by quantifying our RDD-estimates in section 2.7.[5]

We examine other factors affected by the reform and if these can explain our findings. This additional analysis is important to shed light onto the underlying mechanisms of the training effect. This matters for policy conclusions as well as for the human-capital interpretation. In a recent study, Carta and de Philippis (2021) present evidence that a pension reform in Italy affected labour supply even before retirement age. In our context, we do find some evidence for such forward-looking behaviour in labour supply but no significant effect on net income.[6] In more detail, estimates on the probability of employment before the age of 60 are positive, between 2 and 3 percentage points but their statistical significance depends on the degree of the polynomial of the running variable that we include in our estimation. Estimates for unemployment are negative but never statistically significant at conventional levels. This is consistent with Chapter 3 of this dissertation, which among others studies forward-looking effects for the same reform but for younger individuals, and finds significant and positive forward-looking employment effects. Further, the results are also consistent with Carta and de Philippis (2021), who study a policy change that shifted normal retirement by seven years in Italy. Thus, it is reasonable that the larger reform in Italy also induced larger effects, as documented by Carta and de Philippis (2021), than the reform in Germany, which shifted the early retirement age by three years.

[5]We show estimates from a range of specifications, most of which are different to zero at the 1% level of statistical significance but of course still estimated with error. We discuss implications regarding (minimum) effect sizes that we can reject in Section 2.7.

[6]The next chapter of this dissertation explores forward-looking employment effects and underlying driving mechanisms in greater detail.

In the context of our analysis, accounting for forward-looking employment effects is of central importance. This is because the Microcensus is a repeated cross-sectional data set and changes in the composition of working women could generate spurious relations between working life and training.[7] It is possible though maybe not plausible, for example, that women who now continue being employed also participate less in training compared to women who remain in employment regardless of the policy. If this was the case, this could generate positive employment estimates in the absence of forward-looking human capital investments. To investigate this empirically, we conduct several additional analyses. First, we estimate effects on training for age-groups and education groups that do not show forward-looking employment effects, and find that the training effects persist and are especially high for college educated women who do not adjust their employment in response to the reform. Second, we show difference-in-discontinuities estimates for employment before retirement, finding no significant effect on employment overall. We demonstrate that while there is some limited evidence for forward-looking behavior for employment outcomes before retirement for subgroups in our setting, these do not interfere with the forward-looking human capital interpretation of our findings on training.

We also conduct a number of additional robustness tests of our main results. First, we use balancing and density checks and show that sorting or jumps in covariates around the cut-off do not drive our results. Second, we conduct placebo analyses to rule out that our results are driven by other aspects than the reform. Next, bandwidth variations around the optimal bandwidth of 12 months as well as donut regressions allow us rule out that the results are driven by bandwidth assumption, outliers, or few observations close to the cut-off. We also present difference-in-discontinuities estimates that estimate RDD effects for affected women against same-age men who were not affected by the reform. This analysis can account for potentially unobserved changes in economic conditions around the cut-off that affect working men and women similarly, and returns similar estimates to our main results.

[7] Note that even with panel data joint outcomes and composition effects remain difficult to disentangle with variation from a single policy.

Testing for heterogeneity, we find that the positive reform effect is driven by individuals who have higher initial education. The pension reform increases training for women with a college degree by more than thirteen percentage points. In contrast, the effect for women without college degree is not significant. Investigating further heterogeneity, we do not find evidence that this positive response is limited to specific firms or regions.

Our study is related to several strands of the literature: First and most directly, to the small existing literature on a link between retirement policy and training. Montizaan et al. (2010) exploit a one-year increase in pension age to estimate effects on training using data on Dutch public-sector workers that was collected one year after the reform. They find evidence for positive effects that is limited to larger firms. Brunello and Comi (2015) exploit the staggered implementation of a series of increases in the minimum retirement age in Italy to estimate training effects for the private sector. Their headline finding is that workers who were hit by a one-year increase in the pension policy in their fifties increased training participation by about seven to eleven percent, depending on specification.[8] In contrast to the previous studies we focus on the human capital effect for women rather than men. Since female labour market participation is in most countries considerably lower, it is of particular importance to understand how to increase female employment and human capital accumulation through pension reforms.

Second, we contribute to the literature on effects of pension policy on employment and income. Studies that estimate positive employment effects based on policy reforms include Duggan et al. (2007), Mastrobuoni (2009), Staubli and Zweimueller (2013), Atalay and Barrett (2015), Manoli and Weber (2016), Geyer and Welteke (2021), or based on structural retirement models Gustman and Steinmeier (1986), Rust and Phelan (1997), French (2005b), French and Jones (2011), Haan and Prowse (2014). Recently, Carta and de Philippis (2021) also estimate forward-looking employment effects before retirement. Crucially, this literature typically assumes an exogenous process of human capital investment, thus implying that individuals cannot adjust their human capital investment

[8] In ongoing work, Bauer and Eichenberger (2017) estimate effects on training from a one-year increase in a sample of Swiss women who were already age 56 when hit be the reform. They find evidence for positive effects similar to Brunello and Comi (2015).

through additional training in response to a pension reform. Notable exceptions are the structural analyses by Fan et al. (2017) and Blundell et al. (2019).[9]

Third, our paper is linked to the literature on human capital investment and life-expectancy, surveyed in e.g. Bloom et al. (2019). The causal empirical literature is small and focused on very particular settings. Oster et al. (2013) use variation in life expectancy driven by Huntington disease realizations across individuals, who have ex-ante similar risks. They find effects in line with human capital theory on college attendance and completion, health outcomes, as well as on job training for individuals between the ages of 17 and 35. In developing countries, Jayachandran and Lleras-Muney (2009) use a strong decline in maternal death rates in Sri Lanka and find positive effects on girls' educational investments measured in years of school education and literacy rates. In another important study, Baranov and Kohler (2018) exploit variation in mortality rates related to HIV medication in Malawi to study effects on savings and on children's educational investments. They find positive effects of an increase in life expectancy on both types of outcomes.[10]

2.2 Institutional Setting, Data and Descriptives

2.2.1 Pension Reform

Before turning to the empirical analysis, we start by summarizing the relevant aspects of the German pension system and the 1999 pension reform that induced exogenous variation in the working life.

The statutory public pension system is the central part of the pension system in Germany.

[9]There is empirical evidence that pension reforms have positive effects on life expectancy of men (Kuhn et al., 2010; Fitzpatrick and Moore, 2018), and Bertoni et al. (2018) find effects on healthy behaviour while working for Italian men. However, estimates for effects on life expectancy for women are small and not statistically significant at conventional levels (Fitzpatrick and Moore, 2018).

[10]Note that several studies also discuss the theory of human capital investment through training and provide empirical evidence about the effect on labor market outcomes in form of wages, job security, and employment probability, see, among others, Pischke (2001), Zweimueller and Winter-Ebmer (2000), Barrett and O'Connell (2001), Leuven (2005), Frazis and Loewenstein (2005), Picchio and van Ours (2012), and Ruhose et al. (2019).

It covers more than 80% of the workforce with the exceptions of groups that are not subject to compulsory pension insurance, most important civil servants, and the self-employed. It includes old-age pensions, disability pensions, and survivor's benefits. The system is financed by a pay-as-you-go (PAYG) scheme and has a strong contributory link: pension benefits depend on the entire working history. The pension system provides several pathways into early retirement, i.e. claiming retirement benefits before reaching the normal retirement age. In this analysis, we focus on the *pension for women*, which allows drawing benefits starting from age 60.[11]

The 1999 reform abolished the *pension for women* for cohorts born after 1951. Effectively, the reform raised the Early Retirement Age (ERA) for most women from age 60 to age 63, thus prolonging the working life.[12] Women born before 1952 could claim the *pension for women* if they fulfilled certain qualifying conditions. The eligibility criteria were: (i) at least 15 years of pension insurance contributions; and (ii) at least 10 years of pension insurance contributions after the age of 40. According to Geyer and Welteke (2021), about 60% of all women born in 1951 were eligible for the old-age pension for women. In our empirical analysis, we focus only on employed women who are neither self-employed nor civil servants; about 89% of these women fulfill the criteria and, therefore, are eligible for this pathway.[13] The pension reform was implemented when affected women born in 1952 were aged 47. Thus, these women had still a long horizon to benefit from human capital investments.

Geyer and Welteke (2021) and Geyer et al. (2020) have evaluated the labor market effects of the pension reform based on administrative data of the public pension insurance accounts and the Microcensus, respectively. The findings of these studies are relevant for the following empirical analysis. Most important the increase in the ERA has a sizable

[11]In addition early retirement is possible via : (1) the *invalidity pension*; (2) the *pension after unemployment or after old-age part-time work*; and (3) the *pension for the long-term insured*; for more details see Geyer et al. (2020). For a more general description on the German pension system, see Boersch-Supan and Wilke (2004).

[12]The *pension after unemployment or after old-age part-time work* was abolished at the same time as the *pension for women*. However, this does not affect our analysis, as the ERA for this pension type was already 63.

[13]We do not observe eligibility at the individual level and quantify what this implies for the interpretation of the reduced form estimates in section 2.7.

positive effect on the working life on individuals which is the necessary condition for an increase in human capital investment. In more detail, employment rates for eligible women aged between 60 and 62 increase by about 15 percentage points, the combined effect on inactivity and unemployment has with about 12 percentage points a similar size.[14]

2.2.2 The German Microcensus

The Microcensus is an annual, household-based survey with representative information about the population and the labor market in Germany. Participation in the survey is mandatory. It has a sampling fraction of one percent of the German population (about 370,000 households) and constitutes the largest annual household survey in Europe (RDC of the Federal Statistical Office and Statistical Offices of the Laender, 2015).

In the main analysis, we concentrate on employed[15] women younger than 60 years born in 1951 and 1952, who we observe from 2005 through 2012. For these years, the data include information about the month of birth and consistent information about participation in on-the-job training.[16] We observe around 1,250 individuals for each birth month in our sample. Thus, overall, the sample includes information on about 30,000 women born in the two cohorts of interest. The Microcensus includes important socio-demographic variables, such as age, education[17], marital status, household income, and firm size. We consider college education and the geographical "West"-dummy as predetermined and

[14]In a previous working paper version of this paper, see Gohl et al. (2021), we supplemented our main analysis with data from the German Socioeconomic Panel (SOEP). This longitudinal dataset allows us to observe whether a women who participated in training indeed worked longer due to the pension reform and to scale the training effects by employment effects. In particular, we used the pension reform as a "first stage" providing exogenous variation in employment to then in an instrumental variable approach see whether longer employment actually increases training in earlier periods. The results support the notion that this indeed is the case. Unfortunately, the sample size in the SOEP for the relevant cohorts is too low to produce robust evidence.

[15]Women working in "mini-jobs" are not counted as employed.

[16]Before 2005, the Microcensus only provides information about the birth year and the definition of training changes at several points in time. Therefore, the extension of the sample before 2005 would require additional assumptions. The Microcensus does not include information about training activities of unemployed, therefore we only consider the effects on employed individuals aged 53-60.

[17]Education is measured with ISCED 1997 levels: based on this information we can differentiate between women without college degree or with college degree or more. In total there are six ISCED levels in this categorization and anyone with a level of five or higher is classified as having a college degree.

include these as controls. The other variables are potentially endogenous, which we keep in mind when using these for balancing checks.

The Microcensus provides in addition information if an employed person has participated in on-the-job training during the twelve months prior to the survey. The training information includes specifically courses that are related to career development, e.g. to improve management, computer, or rhetoric skills.[18] A further question that is included relates to training that is not job-related. Examples of such training are classes in music, sport and health, cooking, or art that many individuals take in their free time, offered through a network of "Volkshochschulen."

On-the-job training in Germany can take place due to several reasons. According to survey evidence from the German Federal Ministry for Education and Training in 2007 more than 50 percent of all on-the-job training participation was self-motivated by the individual, only 32 percent based on firm requirements and the remainder suggested by a superior (Training Report, 2010: p.17). Unfortunately, our dataset does not allow to differentiate between these reasons and thus our training measure includes mandatory as well as voluntary training.[19]

2.2.3 Descriptive Analysis

In Table 2.1, we provide descriptive information about the key variables in our sample for women born in 1951 and 1952 who we observe over the 2005 to 2012 period in the Microcensus. In addition, we present in the first column the statistics for all employed women born in cohorts 1940-1985 aged 30 to 65. Finally, in the last column we show if the variables significantly differ between the two cohorts we use later for the analysis. Women born in 1952 have a significantly higher level of training participation, are

[18] The exact wording of the question reads: *Did you, in the last 12 months, take part in any form of vocational training? Examples of vocational training are occupational re-training, courses for career development, and general training courses in, for example, the fields computing, management, and public speaking.*

[19] Note that it is unlikely that there is any difference in the participation in mandatory training between cohorts 1951 and 1952 with regards to this mechanical form of training prior to reaching the age 60,i.e., when individuals in the cohort 1952 are more likely to work longer.

slightly more likely to have a high level of education and are less likely to live in West Germany. Crucially, in order to account for potential confounding effects stemming from these differences we later conduct a balancing test and control for covariates such as the educational level and whether the observed women lives in former West or East Germany.

Table 2.1: Summary Statistics: Employed Women

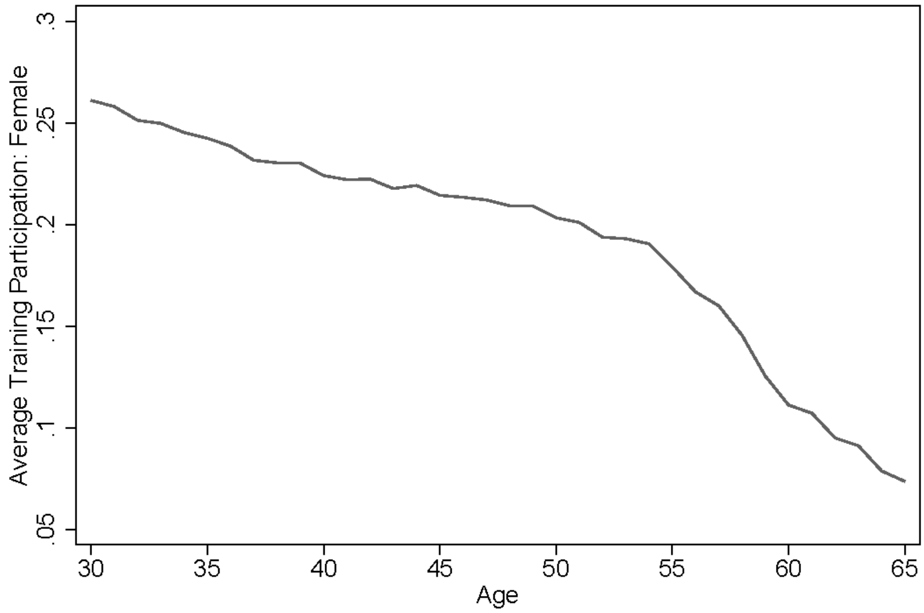
	All cohorts	Cohort 1951	Cohort 1952	Diff. 1951-52
On-the-Job training participation	0.205	0.1543	0.1717	-0.0174***
	(0.403)	(0.3613)	(0.3772)	(0.0044)
Large Corporation	0.504	0.5222	0.5247	0.0025
	(0.403)	(0.4995)	(0.4994)	(0.0060)
High level education	0.225	0.1887	0.1990	-0.0103**
	(0.417)	(0.3913)	(0.3993)	(0.0047)
Medium level education	0.651	0.6518	0.6560	-0.0042
	(0.477)	(0.4764)	(0.4750)	(0.0056)
High HH income	0.372	0.3296	0.3380	-0.0084
	(0.483)	(0.4701)	(0.4730)	(0.0056)
Married	0.631	0.7203	0.7165	0.0038
	(0.4001)	(0.4489)	(0.4507)	(0.0038)
West-Germany	0.810	0.7685	0.7582	0.0103**
	(0.0391)	(0.4218)	(0.4282)	(0.0050)
Observations	649,181	13,649	14,870	

Source: Microcensus 2005-2012, own calculations. Table 2.1 depicts average values of outcome variables and covariates for the female working population for women aged 30-65 (Cohorts 1940-1985) as well as by control (1951) and treatment cohort (1952). Standard deviations for average values are depicted in parentheses. Differences between cohort-specific mean values are tested by a two-sided t-test. Standard errors are depicted in parentheses and significance levels follow $*$ p 0.10, $**$ p 0.05, $***$ p 0.01.

In Figure 2.1, we provide further general evidence about the age and cohort pattern of training for employed women. In Figure 2.1, we focus on all employed women born between 1940 and 1985. We find a declining pattern that is explained by cohort and age effects. Training rates are above 25% at the age of 30 and then monotonically decline to about 10% at the age of 60.

This descriptive pattern could already be interpreted as evidence in favor of human capital theory and the theoretical prediction that training should decrease towards the end of working life, but it is important to re-iterate the following: this descriptive figure mixes up cohort and age effects and we know that job-related training is increasingly important (Köller et al., 2017). One result of such unobserved time-trends could be that younger cohorts might have higher training levels throughout their working life and, thus, the age effect could be non-existent or even upward sloping to generate this overall pattern. Thus, descriptive evidence, like that shown in Figure 2.1, is relevant for documenting the incidence of training for different age groups but cannot inform causal questions.

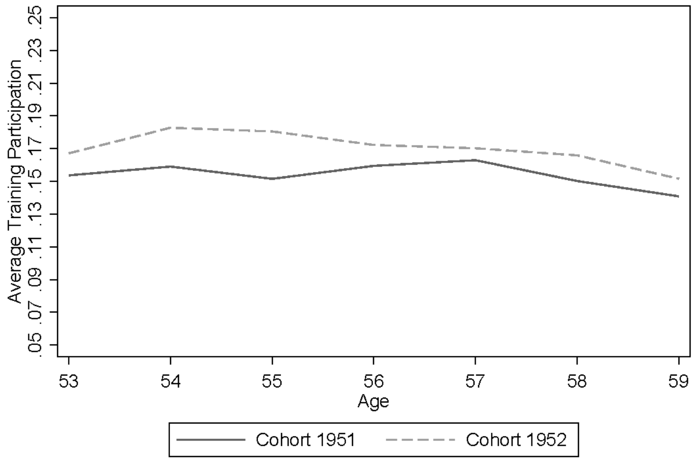
Figure 2.1: Average On-the-job Training Participation by Age



Note: Figure 2.1 plots the average on-the-job training participation by age for all employed women of the cohorts 1940-1985 using the Microcensus from 2005 onward.

To shed some more light on this, Figure 2.2 shows age-specific training participation rates for the two cohorts from the Microcensus that we use in the RDD to estimate the training effects. Two observations are of interest: First, the younger cohort of 1952 shows a higher incidence of training for all ages 53 to 59, when compared to the older cohort of 1951. In this Figure, this difference is a combination of the reform effect, since only the younger cohort was affected by the 1999 pension reform, as well as general differences that might occur between cohorts. In our analysis, we control for the latter using the RDD design. Second, the age-trend from 53 to 59 is negative, but only marginally so. This stands in stark contrast to the age-pattern shown in Figure 2.1 and underlines the necessity to separate out cohort from age-effects for a causal interpretation.

Figure 2.2: Average On-the-job Training Participation for Sample Group



Note: Figure 2.2 plots the average on-the-job training participation for all employed women of the cohorts 1951 and 1951 using Microcensus data from 2005 onward.

In Figure 2.A.1 in Appendix B, we turn to the training pattern by initial education. We find that the probability to train increases with the level of education. Specifically, employed women with no college degree (ISCED < Lower Tertiary) have training rates of about 11.7 percent. In contrast women with a college degree or higher tertiary education (ISCED ≥ Lower Tertiary) have training rates of 37.1 percent. Again, this figure is merely of descriptive nature, but the differences by educational level motivates our heterogeneity analysis along this dimension.[20]

2.3 Empirical Method: RDD

In the RDD-analysis, we exploit the 1999 pension reform to estimate the effect of an increase in working life on human capital investment. The reform leads to an arbitrary and distinct cut-off for women born before and after December 31, 1951, which determines assignments into the treatment and the control groups.

More formally, and similar to Geyer and Welteke (2021), in the empirical analysis the

[20]In Appendix Figures 2.A.3 and 2.A.2, we also show figures with/without college for the cohorts 1951 and 1952. The patterns are similar to the discussion of Figure 2.2 vs. 2.1 above.

standardized woman's month of birth is the running variable M that determines treatment D as one if she was born after December 31, 1951, and zero otherwise:

$$D_i = \begin{cases} 1, & if\ M_i \geq c \\ 0, & if\ M_i < c \end{cases} \tag{2.1}$$

For identification of a causal effect, it is important that no manipulation of the month of birth for women born in 1951 and 1952 and no selection into or out of treatment is possible. As a result, the treatment and control groups should be otherwise comparable around the cut-off. We provide supporting evidence based on balancing tests of important pre-policy covariates of the 1951 and 1952 birth cohorts as well as by moving the cut-off to hypothetical placebo dates. Moreover, as discussed e.g. in Geyer and Welteke (2021), no other relevant policy reform differently affected women born in 1951 and 1952.[21]

In the main specification, we implement the RDD in the following regression model:

$$y_i = \alpha + \beta D_i + \gamma_0(M_i - c) + \gamma_1 D_i(M_i - c) + X_i\delta + \sum_a \kappa_a A_{it} + \sum_y \lambda_y Y_y + \varepsilon_i \tag{2.2}$$

D_i is a dummy specifying treatment, that is equal to 1 if a woman is born 1.1.1952 or later and 0 otherwise. A woman's date of birth, measured in months, is described by M_i and c is the cut-off date for the increase in Early Retirement Age, ERA (January 1, 1952). The difference between a woman's birth month and the beginning of the ERA increase, $M_i - c$, gives the running variable. The running variable across all specifications is interacted with the treatment variable D_i to allow for different slopes before and after the cut-off. γ_0 is the coefficient of the running variable and γ_1 is the coefficient of the interaction term. In addition, we account for further explanatory variables X_i, including predetermined education and regional information. More precisely, we control for high/college education (ISCED 1997 $\geq$ Lower Tertiary), medium education (ISCED 1997 categories 3 and 4)

[21] Note that in 2007, there was another pension reform, which shifted the normal retirement for the 1952 cohort by an additional month compared with the 1951 cohort.

and an indicator variable for west Germany. We also include age fixed effects $\sum_a \kappa_a A_{it}$ and survey year fixed effects $\sum_y \lambda_y Y_y$ to control for potential difference in the survey years and differences in the macroeconomic environment.

In our main approach, we estimate this specification using local regressions and include polynomials up to the second degree in the running variable $M_i - c$ and its interaction with the treatment indicator $D(M_i - c)$.[22] Moreover, we estimate OLS regressions (linear and quadratic) and test for robustness for various bandwidth choices. The outcome variable Y in our analysis is on-the-job training, which is dichotomous i.e. taking on the value 1 if a women has participated in training in the last twelve months and 0 if she has not.[23]

For some robustness checks, we also implement a difference-in-discontinuities approach. The difference-in-discontinuities approach in the setting of this paper effectively estimates the difference between cohorts 1951 and 1952 for women and men and compares them to each other. Empirically, it is implemented by interacting the regression equation for the RDD with an indicator function, equal to one for the actually treated sample, i.e. in this case women. F_i thus is a dichotomous variable equaling one if the individual is female. The equation below details the estimation approach. Note that all variables follow the same notation as in the main RDD approach.

$$
\begin{aligned}
y_i = \alpha + \beta F_i D_i + \gamma_0 F_i (M_i - c) + \gamma_1 F_i D_i (M_i - c) + \eta_0 F_i + \eta_1 D_i \\
+ \eta_2 (M_i - c) + \eta_3 D_i (M_i - c) + X_i \delta + \sum_a \kappa_a A_{it} + \sum_y \lambda_y Y_y + \varepsilon_i
\end{aligned}
\tag{2.3}
$$

By implementing this strategy we explicitly control for any macro(economic) developments over the estimation period by using men as a control group. This is possible as the end of the *pension for women* did not impact men directly. Further, by including men as an additional control group we can explicitly rule out turn-of-the-year and seasonality effects, i.e. effects caused by the differences in birth years and/or months between control

[22] Local polynomials are estimated using the Stata package "rdrobust" (Calonico et al., 2018a).
[23] Estimation results based on a probit model (not reported) show very similar results.

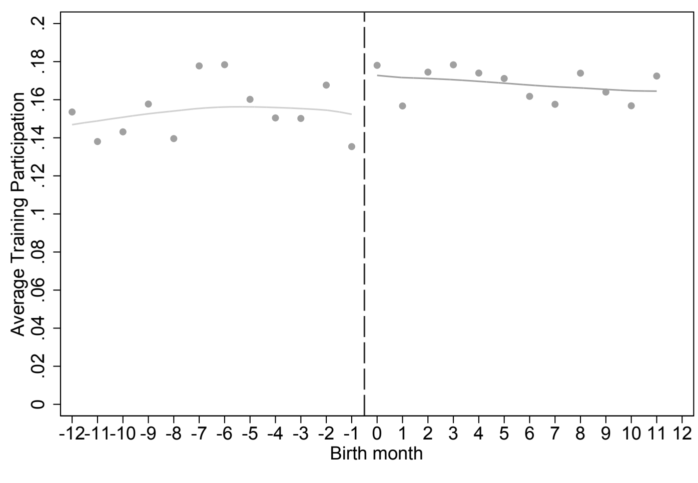
groups and treatment group.

2.4 The Reform Effect: Main Results

2.4.1 Training Effects: Graphical Analysis

Figure 2.3 shows participation rates in training by month of birth, 12 months before and after the cut-off birth date, 1.1.1952, for all employed women in their later working life, i.e., when they are aged between 53 and 60. The share of employed women participating in training is higher after the cut-off. Specifically, the average rate of participation among those born in the 12 months before the cut-off is approximately 15.4%. After the cut-off date, the graphs show a jump in the average rate of training participation for employed women under 60 to more than 16.5%.

Figure 2.3: On-the-job Training around the Cut-off Date



Notes: This figure plots average on-the-job training over the running variable, which is the distance to the cut-off date measured in birth month (x-axis). The fitted lines are local linear regressions using a first degree polynomial, a triangular kernel. In total, there are observations for 13,658 individuals below the threshold and 14,873 observations for individuals above the threshold. Source: Microcensus, own calculations.

Importantly, and in contrast to the descriptive evidence discussed in Section 2.2.2, women close to the cut-off are of almost identical age, thus cohort-effects are an unlikely explanation for this jump. While graphical RDD-evidence can be informative, eye-balling alone can be misleading. Thus, in the next section, we examine the robustness and significance of the graphical evidence using various choices in the RDD framework.

2.4.2 Training Effects: RDD Results

To quantify the effect of an increase in the working life on the investment into human capital, we use the RDD described in Section 2.3. In Table 2.2 we present the estimation results for different specifications with observations 12 months before and after the cut-off date (we discuss alternative bandwidth choices in section 2.5.4). The first two columns present results from regressions estimated using OLS with polynomials with linear and quadratic specifications of the running variable. The two remaining columns show the corresponding local estimates. Moreover, the table includes regressions without and with additional control variables. Age and year fixed effects are always included. Standard errors are clustered at the standardized birth month level and are reported in brackets. Our inference is robust to a specification without clustered standard errors using robust standard errors as suggested by Kolesar and Rothe (2018).

The results of these different specifications all point in the same direction, despite some differences in the magnitude of the point estimates: the increase in the Early Retirement Age has a positive and significant effect on the investment in training. Although positive, the linear OLS-specification in the top panel (without covariates) in Column 1 is not statistically significant. However, since this specification does not control for non-linear patterns around the cut-off, it is *a priori* not the preferred specification. In contrast, the local regressions in Table 2.2 consistently show positive and significant estimates in similar magnitude across all specifications. Following Cattaneo and Titiunik (2022), we focus on the local regressions as our main approach and only report these results in the remaining tables in the main text. In particular, we treat the quadratic local regression as our preferred specification, as this allows to control for potential non-linear patterns around the threshold.

Table 2.2: Regression Discontinuity: Training Effect

	OLS		Local Regression	
	(1)	(2)	(3)	(4)
Without Covariates				
Treatment Variable	0.0157	0.0352***	0.0235**	0.0418***
	(0.0147)	(0.0142)	(0.0120)	(0.0131)
With Covariates				
Treatment Variable	0.0158*	0.0360***	0.0272***	0.0395***
	(0.0077)	(0.0096)	(0.0079)	(0.0088)
Age FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Running Variable	Linear	Quadratic	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Pre-Policy Mean: 15.43 percent; Number of observations 28,519. Specifications including covariates control for a high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

The point estimate in the local regression approaches show that the probability to participate in training increases between 2.4-4.2 percentage points. This is a sizable effect that would translate into a relative increase of about 15.6-27.3% given the pre-reform share in training of 15.4%. Our preferred estimate in column (4), the local regression with quadratic polynomial and with controls, estimates a training effect of four percent, equivalent to a 26% increase in training. As discussed in Section 2.2, these estimates present intent-to-treat (ITT) estimates because not all women were affected by the reform. We return to the quantification of these ITT-estimates, when we derive estimates for the corresponding average treatment effects on the treated (ATT) in Section 2.7.

2.4.3 Training Effects: RDD Results, by Age

We now go into more detail and estimate the reform effects by age. We have already discussed the training-age profile descriptively in Section 2.2.3. Theoretically, we would expect that relatively younger women affected by the reform have a higher incentive to invest into more training. On the other hand, later retirement induced by the reform might be less salient for individuals further away from retirement age. To investigate age-heterogeneity empirically, we focus on the quadratic specification with covariates.

In Figure 2.A.5 in the Appendix we present the results. Overall, we do not find a clear age gradient. The point estimates for the different ages are quite similar. Given the smaller sample size in the age-specific regressions, the confidence intervals are relatively large, thus not all age-specific effects are significant.

One explanation for this finding is the salience of the reform effect. On the one hand, from a human capital perspective training should be most benefiting at younger ages, so the reform effect stronger initially. On the other hand, training in response to extension of the working horizon can be more salient the closer an individual is to potential retirement. Salience could therefore increase with age, thus contributing to the age-pattern documented here.

2.4.4 Forward-looking Employment and Income Effects

Employment We now test for direct effects on employment of the reform for pre-treatment ages, i.e. before the earliest retirement age of 60. This is of particular relevance because the Microcensus data set is a repeated cross-section and we base our analysis sample on women in employment before the age of 60. Thus, any effects of the reform on employment could induce sample selection and bias our estimates - or more precisely: the human-capital interpretation of our estimates.

The results of this additional analysis are presented in Table 2.3. Here, we show estimates similar to Specification 2.2 using linear and quadratic local regression specifications but for the population of all women in these age groups that responded to the Microcensus survey. In columns (1)-(4) employment and unemployment levels before the age of 60 are used as outcome variables, each coded with a dummy variable that equals one if the person is employed/unemployed at the time of the survey. The results show that there are no significant effects on unemployment, i.e., see columns (3)-(4). The effects on

Table 2.3: Employment and Income Effects

Outcome Variable	(1) Emp.	(2) Emp.	(3) Unemp.	(4) Unemp.	(5) Income	(6) Income
Without Covariates						
Treatment Variable	0.0225**	0.0143	-0.0104	0.0099	-1.9007	4.2213
	(0.0092)	(0.0189)	(0.0109)	(0.0121)	(17.9040)	(21.4920)
With Covariates						
Treatment Variable	0.0249***	0.0135	-0.0130	0.0103	-4.1830	-4.0053
	(0.0091)	(0.0110)	(0.01085)	(0.0103)	(0.0230)	(0.0307)
Age FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Running	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
N	54,087	54,087	54,087	54,087	50,035	50,035

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cut-off. Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Table 2.3 shows results from local linear and quadratic RDDs for a range of further outcome variables, which are denoted at the top of the table. The variable income measures the personal income denoted in Euro. All specifications control for a high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

employment are mixed. Employment effects, depicted in columns (1)-(2) are positive in the linear local specification. These significant effects in the employment outcome need to be examined more carefully. We therefore conduct additional tests in order to rule out that potential employment effects induced by the reform drive our training results.

First, we provide results by age groups in order to see which ages drive the employment estimates. Figure 2.A.6 in the Appendix presents the results for employment. The visual examination across Figures 2.A.5 (effects on training, discussed above) and Figures 2.A.6 (effects on employment) shows that training and employment effects take place at different ages. More precisely, the employment effects are driven by age 55 and to a smaller extent by age groups 58 and 59.[24] In contrast, the training effects are driven by age groups 53, 56, 57, 58. In a second step, we exclude all age groups for which we find significant employment effects and repeat our main estimation for the training outcome.[25] Table

[24] The point estimates are borderline significant at the ten percent level in the preferred quadratic specification. When using a linear specification of the running variable the results are not significantly different from zero.

[25] When using employment as an outcome and excluding the described age groups, we no longer find significant employment effects.

2.A.4 in the Appendix shows the results. The point estimates in both the linear and the quadratic specification are somewhat larger than in the estimation for the whole sample. Further, the training estimates remain highly significant. If employment effects were driving the training estimates, we would have expected smaller and insignificant effects.

Taken together, there is some, albeit not very strong, evidence for forward-looking employment effects in our setting. However, these occur largely for different age-groups of people and therefore do not mechanically drive our findings on training. This confirms our interpretation of the training estimates as evidence for forward-looking human capital investment.

Income: Another outcome that could be affected by the reform is income. This is because existing empirical evidence shows that training has a positive effect on wages (Frazis and Loewenstein, 2005; Blundell et al., 2019). Moreover, training can improve the quality of work and can have positive effects on non-pecuniary outcomes (Ruhose et al., 2019). The Microcensus does not include information on *gross* labor earnings but on *net* income on the individual and the household level. Net-income includes in addition to labor earnings information on transfer and tax payments. The results shown in Columns (5) and (6) of Table 2.3 show that there is no significant effect on income in either the linear or quadratic local specification.

In Table 2.A.1 in the Appendix we further repeat our estimations for a sample of employed women only. Panel A depicts the results for all employed women and Panel B for all women. Across all specifications, none of the coefficients is statistically significant at conventional levels. This null effect is not surprising. First, we do not find strong effects on employment behavior and second the net-income measure includes as well transfer income.[26]

[26] The SOEP, another German household level dataset, contains better measures of income but cannot be applied to the RDD-anaylsis due to smaller sample size. A DiD analysis based on the SOEP does not produce conclusive results and is included in a working-paper version of this article (Gohl et al., 2021).

2.5 Robustness

This section is dedicated to providing additional robustness tests, supporting our main results. First, we conduct balancing tests, which show that no other factor potential influencing training varies discontinuously around the cut-off. Second, we provide a density test to assess whether sorting patterns around the cut-off may undermine our identification strategy. Next, we conduct placebo tests shifting the cut-off date and and focusing on private training, which following human capital theory should not be affected by the reform. Further, we validate our RDD identification approach using difference-in-discontinuities regressions. Last, we provide results for different bandwidths and conduct donut regressions for our RDD approach. This allows to check the sensitivity of our approach to bandwidth choices and outliers.

2.5.1 Balancing and Density Tests

Balancing The assumption underlying the RD design is that other factors vary smoothly across the cut-off. We provide support for this assumption by using individual control variables as outcomes using the same specification as in Table 2.2. The resulting estimates are presented in Table 2.4.[27] The Microcensus does not offer many variables that can be safely considered as pre-determined with respect to the 1999-policy change. Therefore, in this first balancing analysis we restrict our attention to whether the individual has a college education and an indicator for "West" (Panel A). We aggregate the ISCED educational levels into two education groups of women with "college" and "non-college." The latter “West” indicator is a dummy variable that equals 1 if the individual was in West Germany in 1989. For both balancing variables and across all specifications, this analysis reveals no jump at the threshold. This supports the underlying RDD assumption of no other changes across the threshold. In addition, we present the balancing of further variables that potentially might be affected by the pension reform. In Panel B, we show that the reform had no significant effect on sorting into big companies, or marital

[27] Results also for OLS specifications are similarly balanced (not shown).

status.[28]

Next, in an additional analysis we seek to rule out more thoroughly that differences in the composition between the cohorts are driving our results. In particular, Figure 2.A.1 in the Appendix, clearly shows that initial education and on-the-job training are strongly correlated. We thus use each ISCED level as a separate outcome indicator variable, repeating our balancing exercise for these outcomes. The results from using the local quadratic polynomial specification are shown in Figure 2.A.4 in the Appendix. Overall, for most categories both cohorts are balanced. For categories *post-secondary* and *higher tertiary*, we find negative and significant point estimates, suggesting that individuals in cohorts 1952 are less likely to be in these categories. In order to rule out that such compositional differences drive our results, we therefore repeat our main estimations controlling for a more granular set of educational controls by including ISCED level fixed effects.

Table 2.A.2 in the Appendix shows that our main RDD estimate, regardless of the specification, does not change in response to the inclusion of additional education fixed effects. Columns (1)-(2), for reference, show the results for the main estimates using local regressions with linear and quadratic polynomials, i.e., the estimate contained in the paper so far. In columns (3)-(4) we replace our education controls[29] with ISCED 1997 level specific fixed effects, i.e., dummy variables for each ISCED Level with the first one serving as a baseline. The point estimates remain highly significant and do not change by much in both the linear and quadratic specification. In a next step, we control for even more granular fixed effects, including subcategories within each ISCED 1997 level as additional indicator variables.[30] Again the main estimates remain virtually unchanged, which leaves us confident that our estimations are robust to a more comprehensive set of educational controls.

[28]The indicator variable *Single* is significant at the 10 per cent level in the local linear regression, however, insignificant across all other specifications.

[29]So far we included a dummy for higher education (ISCED category lower tertiary and higher tertiary) and a medium education indicator (upper secondary and post secondary)

[30]For example, the ISCED level *tertiary education* can be split into vocational and academic education

Table 2.4: Balancing

Local Regressions	(1)	(2)
Panel A: Pre-determined covariates		
College Education		
Treatment Variable	0.0029	0.0191
	(0.0230)	(0.0307)
West		
Treatment Variable	0.01162	-0.0010
	(0.0233)	(0.0326)
Panel B: Further Variables		
Big Company		
Treatment Variable	-0.00900	0.0023
	(0.0080)	(0.0104)
Single		
Treatment Variable	-0.0089*	-0.0050
	(0.0051)	(0.0063)
Age FE	✓	✓
Year FE	✓	✓
Running Variable	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at the birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Outcome variables depicted in bolt. Number of observations in descending order of rows: 28,519, 28,519, 27,179, 28,519, 28,175, 28,519. Source: Microcensus 2005-2012, own calculations.

Density Next, we check for potential manipulation around the cut-off. To do so, we implement a density discontinuity test described in more detail by Calonico et al. (2018b). Manipulation of the birth date cut-off and consequent sorting in our setting is difficult and highly unlikely. However, as shown in the previous section there is some evidence for potential sorting into employment (before the age of 60) in response to the reform. This sorting would also show up in a density test of our main sample, as we focus exclusively on employed women in our main specification. As a result, in our setting the usual rule that violations of the density test invalidate the empirical analysis does not readily apply. Instead, this rule only holds for samples that do not show an employment response.

To test this, Table 2.A.3 in the Appendix depicts the test statistics and p-values of the density test using a bandwidth of 12 months as in our main specification and Figure 2.A.9 in the Appendix shows the corresponding plot. Column (1) shows the results for our whole sample. In line with the positive employment for subgroups effects the test statistic is significant. In column (2), we only include age groups for which we found insignificant employment effects to account for the mechanical shift in the density. As expected for this sample all p-values of the test now lie above the 10 percent level of statistical significance.

Next, we split the sample into those with a college degree and those without a college degree. Below in Section 2.6.1, we show that it is exclusively individuals with a college degree who drive the training results, while employment effects for this group are statistically insignificant at conventional levels across all specifications. The density test clearly shows that for individuals with a college degree, i.e. the group driving the training results, there is no significant jump in the density. In contrast, for individuals with no college degree, the density test statistic is highly significant. Overall, the density results show that for the sub-samples that drive the training results but not our employment results, there is no significant jump in the density around the cut-off.

2.5.2 Placebo Analysis - Different Cohorts and Private Training

We conduct placebo analyses that we present in Table 2.5. In the first placebo analysis, we artificially shift the cut-off date by one year to 1.1.1950 (Panel A) or to 1.1.1952 (Panel B). Importantly, the pension rules are identical before and after the chosen placebo cut-offs. The shift of one entire year (in either direction) is of particular relevance as this could capture seasonal effects related to the December to January timing of the reform introduction. The result from this additional analysis supports our identification strategy: the treatment effect is very close to zero and not significant in both placebo specifications, with and without additional explanatory variables. Moreover, these effects are precisely estimated and clearly differ from our main findings in Table 2.2.

Table 2.5: Placebo Analysis: Training Effects

Local Regressions	(1)	(2)
Panel A: Placebo 1950/51		
Without Covariates		
Treatment Variable	0.0017	0.0055
	(0.0102)	(0.0153)
With Covariates		
Treatment Variable	0.0044	0.0035
	(0.0207)	(0.0325)
Panel B: Placebo 1952/53		
Without Covariates		
Treatment Variable	0.0043	0.0035
	(0.0208)	(0.0326)
With Covariates		
Treatment Variable	-0.0009	-0.0029
	(0.0086)	(0.0110)
Panel C: Private Training		
Without Covariates		
Treatment Variable	-0.0020	-0.0007
	(0.0030)	(0.0038)
With Covariates		
Treatment Variable	-0.0021	-0.0009
	(0.0030)	(0.0038)
Panel D: Men		
Without Covariates		
Treatment Variable	0.0042	0.0063
	(0.0063)	(0.0086)
With Covariates		
Treatment Variable	0.0059	0.0097
	(0.0062)	(0.0073)
Age FE	✓	✓
Year FE	✓	✓
Running Variable	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Number of observations Panel A: 24,623 ; Number of observations Panel B: 34,150; Panel C: 28,519, Panel D: 31,557. Covariates: high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

Next, we exploit additional information on "private training" that is recorded by the Microcensus. In contrast to job-related training, such "private training" has benefits that also lasts beyond the working life and, thus, we do not expect to find differential take-up depending on the pension rule.[31] The results are presented in Panel C of Table 2.5. As expected, none of the estimates are statistically significant. Moreover, all estimates are very close to zero and negative, clearly different to the main results found for job-related training.

2.5.3 Difference-in-Discontinuities Analysis

Table 2.6 shows the estimates from the difference-in-discontinuities approach outlined in Section 2.3. The results support the main findings presented above. The results are a little bit smaller in size in comparison to the main results (i.e. 0.0266 vs. 0.0395, with quadratic polynomial and controls) but this difference is not significant at conventional levels. In contrast, all estimates remain distinct from zero and the null hypothesis of no effect is rejected at high levels of statistical significance. Overall, these results therefore confirm our previous estimates. Note we will refer back to this set of slightly smaller estimates when quantifying effects sizes in Section 2.7.

We also estimate forward-looking employment effects using the difference-in-discontinuities design (reported in Appendix Table 2.A.5). Here, we do not find any significant effects on employment induced by the reform. The fact that the results on training hold in the diff-in-disc setting, but not for the case of employment, supports the earlier conclusion that sample selection through effects on employment before formal retirement age do not drive our findings on training.

[31] In the Microcensus, private training is classified as general training measures with a predominant private focus to advance one's own skills and knowledge. Examples given for private training in the Microcensus questionnaire are training in the fields of music, sport and health, cooking, or art.

Table 2.6: Difference-in-Discontinuities: Training Effect

Local Regressions	(1)	(2)
Without Covariates		
Treatment Variable	0.0242**	0.0305***
	(0.0107)	(0.0111)
With Covariates		
Treatment Variable	0.0212***	0.0266***
	(0.0066)	(0.0065)
Age FE	✓	✓
Year FE	✓	✓
Running Variable	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; Number of observations: 60,076. Covariates: high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

2.5.4 Bandwidth and Donut RDD

Bandwidth choices can affect RD estimates, so we carefully examine if and how our main results are sensitive to these.

In Appendix Table 2.A.6 we show that our results also hold for additional bandwidth choices of 6, 9, and 12 months. The bandwidth of 12 months is the chosen bandwidth of the endogenous bandwidth selection routine "rdbwselect" using the mean squared error criterion, a triangular kernel and the preferred quadratic polynomial (Calonico et al., 2014). Other criteria and polynomials render bandwidths estimates ranging from 6 to 12. In all cases, our estimates remain in the same ballpark.

As additional robustness check, we examine if observations close to the cut-off drive our effects by estimating effects from donut-RD regressions. We estimate different specification of Equation 2.2 without the one or two birth month closest to the cut-off on both sides. Appendix Table 2.A.7 shows the resulting estimates for the linear and quadratic local polynomials as well as with and without individual control variables. In one specification (Panel A, column 2) the estimate is now only statistically significant at the ten percent level. All others remain highly significant, including from the more demanding two-month donut. Overall, this additional analysis confirms the main findings, with estimates of similar magnitudes throughout. Our results are not driven by outlier observations close to the cut-off.

2.6 Heterogeneity

2.6.1 Heterogeneity by Initial Education

We now extend the empirical analysis by focusing on effect heterogeneity along prior educational levels. We already showed descriptively that training participation positively correlates with prior educational levels in Section 2.2.2 and Figure 2.A.1. In this analysis, we test if the reform effect also varies by prior educational level. Since we only focus on women close to the cut-off, to alleviate issues of sample size, we again aggregate the ISCED educational levels into two education groups of women classified as "college" and "non-college."

Before turning to the results, note that Geyer et al. (2020) show that employment effects of the same pension reform for women aged 60-62 are of similar size for highly educated women (9.5%) and for women without higher education (8.2%). Thus, any differences in the reform effect along the education dimension are not related to differences in retirement decisions.

For training before the age of 60 we find very strong differences by education for the different specifications (see Table 2.7). Looking at our preferred specification including a quadratic specification of the running variable, women with college education or more increase training by nearly 13.8 percentage points, corresponding to a relative increase of about 43.4%.[32] The effect for women without college education is estimated to be close to zero and not significant at conventional levels.[33]

[32]We discuss the magnitude of this effect in detail in Section 2.7.

[33]Note that Figure 2.A.2 suggests that women aged 55 might be driving the strong results for women with a college degree. We therefore also estimate the heterogeneous results excluding age group 55 and the results remain unchanged.

Table 2.7: Training Effect Heterogeneity by Educational Level

Local Regressions	(1)	(2)
Panel A: Non-College		
Without Covariates		
Treatment Variable	0.0085	0.0158
	(0.0095)	(0.0133)
With Covariates		
Treatment Variable	0.0086	0.0160
	(0.0095)	(0.0133)
Panel B: College		
Without Covariates		
Treatment Variable	0.0844**	0.1383***
	(0.0354)	(0.0392)
With Covariates		
Treatment Variable	0.0831**	0.1377***
	(0.0353)	(0.0395)
Age FE	✓	✓
Year FE	✓	✓
Running Variable	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; Participation rates in training for cohort 1951 are 0.154, 0.317, 0.117 for all employed women, employed women with college and employed women without college, respectively. Number of observations: Panel A 22,984; Panel B: 5,535. Covariates: high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

We tested (and rejected) in the main results section above the hypothesis that forward-looking employment effects mechanically create the training effects that we observe due to the sampling of our data. The finding that the positive training estimates are entirely driven by large and statistically significant effects for the college-educated now allows one additional test. If composition effects were driving the training effects, we would expect to find forward-looking employment effects for the same groups. In Appendix Table 2.A.8 we present results for heterogeneity of employment effects. While the estimates are positive throughout, none of the estimates for college-educated workers is statistically significant. In contrast, two out of four estimates for workers without college are statistically significant at the 10% level or higher. This confirms our previous findings and interpretation.

2.6.2 Heterogeneity by Firm Size

Next, we examine if the training effect differs by the size of the company. For this, we estimate the RDD separately for women working in large vs small/medium-sized companies. We use the classification from the Microcensus, where companies with 50 or more employees are classified as large. Before turning to the interpretation of the results, we note again that company size is potentially an outcome in its own right, and can thus be considered endogenous. But as documented in Table 2.4, the pension reform did not have an affect on sorting into bigger companies.

Appendix Table 2.A.9 splits the sample by company size. For large companies, looking at the preferred local regression approach, the effects vary between 0.79 and 2.65 percentage points when including controls. For small-medium sized companies the effects fall in the range of 3.22 and 4.82 percentage points. Overall, out of the eight estimates provided, five are significant. The estimates thus suggest slightly larger and more significant effects in small and medium sized companies, but the confidence intervals of the estimates are overlapping. We therefore conclude that the main dimension of heterogeneity is along levels of initial education.

2.7 Quantifying Treatment Effects of the Reform

As discussed in Section 2.2.1, the pension reform only affected the working life of women who fulfill the eligibility criteria for the so-called *pension for women*. The Microcensus is a cross sectional data set without information about the employment history. Therefore, we can not directly determine the eligibility within this data. As a result, our estimates should be interpreted as intention-to-treat (ITT) effects, giving a lower bound of the true effect.

To gauge information about actual eligibility, we use information from the SOEP longitudinal data, according to which about 76% of all women employed before entering retirement were indeed eligible for this pathway into retirement. This rate increases to approximately 89% when excluding self-employed and civil servants, who are not, by definition, eligible. Further, SOEP data show that about 86% of employed women without a college degree and 94% of women with a college degree fulfill the eligibility criteria.

With this information and the estimated effects (ITT) presented in Tables 2.2 and 2.7, we can derive the average treatment effect on the treated (ATT) overall and for the different educational groups. More precisely, we divide the ITT estimates by the respective eligibility share, E_g, for each educational group g to obtain ATT estimates:

$$ATT_g = \frac{ITT_g}{E_g}$$

These are presented in Table 2.8. Overall, the pattern of the ATT effects is similar to the ITT effects, but the effects are slightly larger. The point estimates suggest that overall probability to participate in training increases by roughly 4.4 percentage points in our preferred quadratic specification including covariates (Panale A, column 2). For women with a college degree the increase is is substantially higher with 14.6 percentage points (Panel B, column 2), and the effect for women without college is close to zero (Panel C, column 2). These estimates imply a relative increase in the probability of training of

28.8% of all women and 46.2% for women with college degree.

These effects are sizeable so it is important to consider if they are plausible. We believe so, for three reasons. First, the effect sizes that we document are not dissimilar to the findings of the small existing empirical literature on training effects. In particular, Brunello and Comi (2015) estimate for private sector Italian workers effects on training of nine percent as a result of a policy that shifts the ERA by one year. In comparison, our estimates are larger but they are also estimated from a larger policy shift. Second, although our results represent sizable estimates, so far this discussion does not account for the fact that, while being distinct from zero at conventional levels of statistical significance, these coefficients are of course estimated with uncertainty. When we instead focus on the lower bounds of the corresponding 95%-confidence intervals, we obtain a value of 2.22 percentage points for the ITT using the local quadratic specification for the whole sample. This corresponds to an ATT of approximately 2.49 percentage points and a relative size of 16.17% overall. For college graduates the lower bound ITT estimate is 6.02 percentage points resulting in an ATT estimate of 6.40 percentage points, and a relative size of 20.18%, again using the quadratic specification. Last but not least, even lower hypothesised estimates cannot be refuted at conventional levels of statistical significance when using the results of the difference-in-discontinuities approach for the baseline.

Table 2.8: Average Treatment Effect on the Treated (ATT)

	(1)	(2)
Panel A: Whole Sample		
ATT	0.0305	0.0443
Relative Size ATT (in%)	19.81	28.77
Panel B: College		
ATT	0.0883	0.1463
Relative Size ATT (in%)	27.85	46.15
Panel C: Non-College		
ATT	0.0099	0.0185
Relative Size ATT (in%)	8.48	15.79
Running Variable	Linear	Quadratic

Notes: The ATT is derived by weighting the ITT effects presented in Tables 2.2 and 2.7. The eligibility rates are calculated from the SOEP data. Participation rates in training for cohort 1951 are 0.154, 0.317, 0.117 for all employed women, employed women with college and employed women without college, respectively. The corresponding eligibility rates are 89.14, 94.12, 86.58. Number of observation Panel A: 28,519, Panel: B 22,984, Panel C: 5,535

2.8 Conclusion

In this paper, we provide causal evidence for the theory of human capital accumulation, which has the key prediction that education investments depend on the length of the payout period. We exploit a sizable pension reform that sharply increased the Early Retirement Age for women between two adjacent cohorts from 60 to 63 years. The analysis is based on the German Microcensus using RDD estimation approaches.

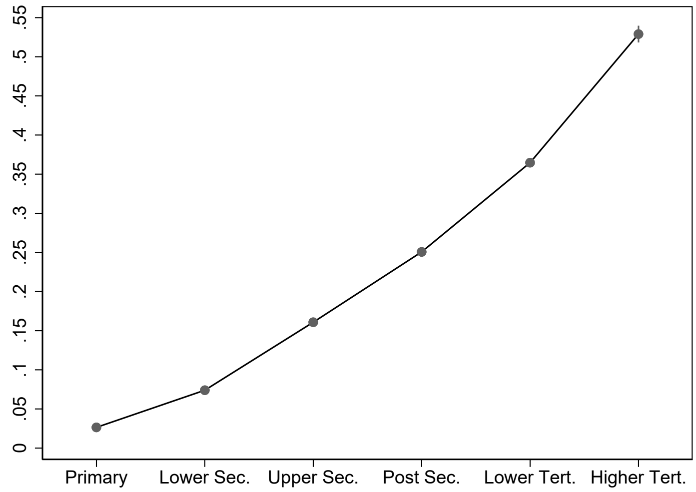
The empirical analysis offers support for the key prediction of human capital theory that the duration of the payoff matters for educational investment decisions. In more detail, our empirical results show that the increase in the ERA has a sizable effect on the human capital accumulation of employed women: in our preferred specification, the probability of training increases by about 4.4 percentage points, which corresponds to an increase of 28.8% for these age groups, and at least 16.2% considering lower bound of the 95% confidence interval. This finding is robust to changes in the bandwidth and for different specifications of the running variable in the RDD and is supported by placebo tests. Investigating heterogeneity, we show that the pension reform increases training for women with a college degree or more by more than fourteen percentage points, which corresponds to a relative increase of more than 46.2%, with a lower bound of 20.2%. The effect for women without college degree is not significant.

Our findings have important implications for the policy debate on pension reform and help to understand why an increase in the retirement age has positive employment effects but also leads to an increase in unemployment and inactivity (see, e.g. Geyer and Welteke (2021)). The debate usually abstracts from the dynamic human capital investment that we document. Our results show that retirement policy need not be thought of as independent of training decisions. Policy makers should consider positive effects on training demand that result from expansions of working life. Moreover, it remains an open question of why the training response is limited to highly-educated women. One explanation is that lowly-educated workers find it harder to increase their working life, thus benefiting less from potential training. A better understanding of the underlying reasons for the documented heterogeneity is key to increase training further, as well for the low educated to reduce inequality of old age income and the risk of old age poverty. From a policy perspective, in ageing societies increasing productivity or just the labor market attachment of older workers is becoming more critical. We believe, future work should examine the role of individual workers and firms in initiating the positive training effects that we document, adding to the still relatively underdeveloped literature on educational investments beyond initial schooling.

Appendices

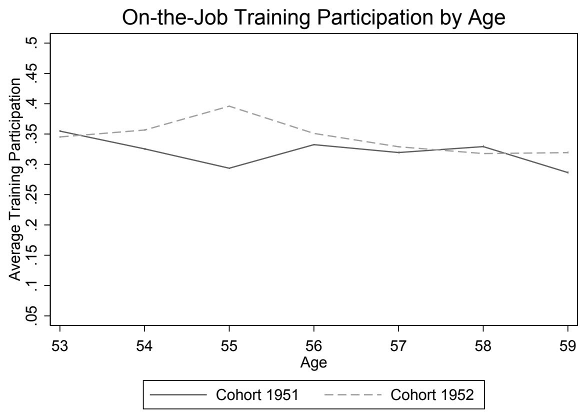
2.A Appendix A: Additional Tables and Figures

Figure 2.A.1: Average On-the-job Training Participation by ISCED Groups



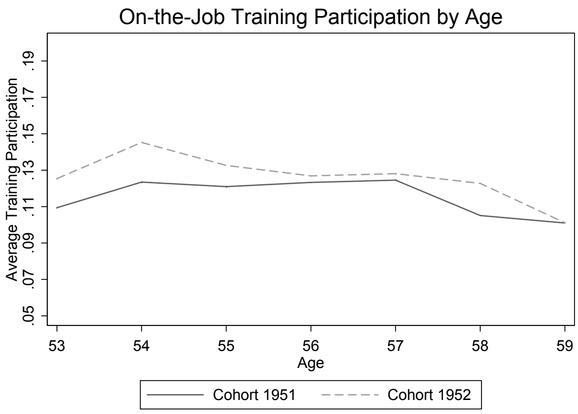
Note: Figure 2.A.1 plots the average on-the-job training participation and corresponding 95 % confidence intervals by ISCED 1997 groups for all employed women (Cohorts 1940-1985) using Microcensus data from 2005 onward. In total there are six ISCED 97 levels. Level 1 corresponds to primary education, level 2 to lower secondary, level 3 to higher secondary, level 4 to post-secondary education, level 5 to first stage tertiary and level 6 to second stage tertiary education.

Figure 2.A.2: Average On-the-job Training Participation for Sub-Sample with College Degree



Notes: The figure plots the average on-the-job training participation for all employed women with a college degree of the cohorts 1951 and 1952 from 2005 onward.

Figure 2.A.3: Average On-the-job Training Participation for Sub-Sample with No-College Degree



Notes: Figure 2.A.3 plots the average on-the-job training participation for all employed women with no college degree of the cohorts 1951 and 1952 from 2005 onward.

Figure 2.A.4: Balancing with ISCED Levels as Outcome

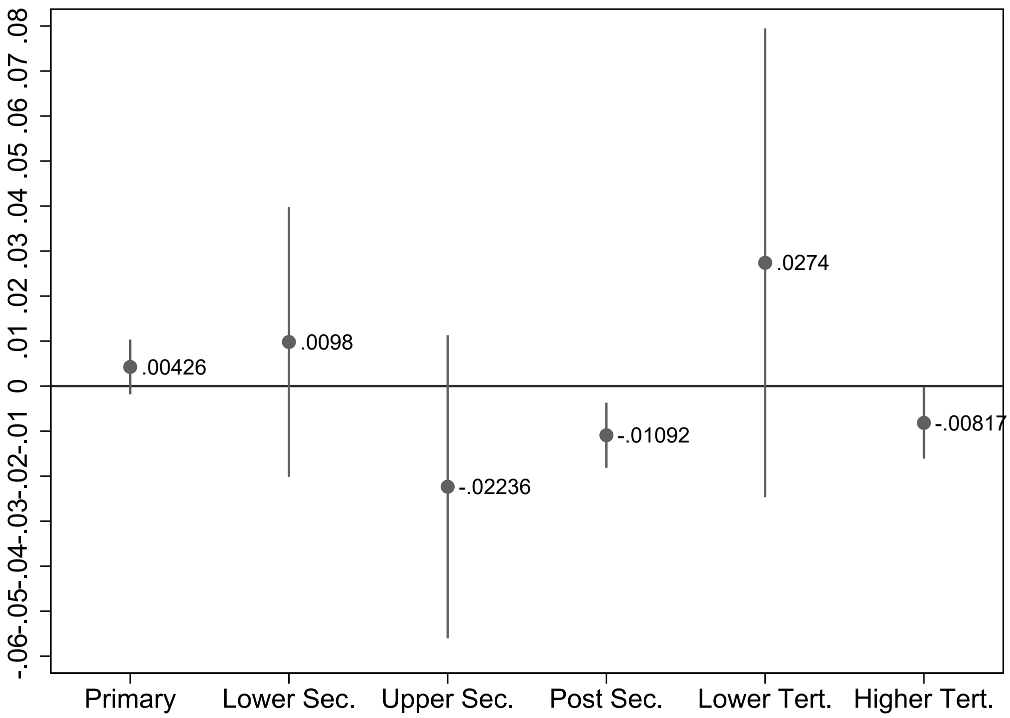


Figure 2.A.4 plots point estimates obtained from a quadratic local RDD specification using our main sample and an indicator variable that is equal to one if the observed individuals' education is classified in the given ISCED level and zero otherwise as outcome variable. Further, it shows the corresponding 95 % confidence intervals for each point estimate. In total there are six ISCED 97 levels. Level 1 corresponds to primary education, level 2 to lower secondary, level 3 to higher secondary, level 4 to post-secondary education, level 5 to first stage tertiary and level 6 to second stage tertiary education.

Figure 2.A.5: Training Effect by Age

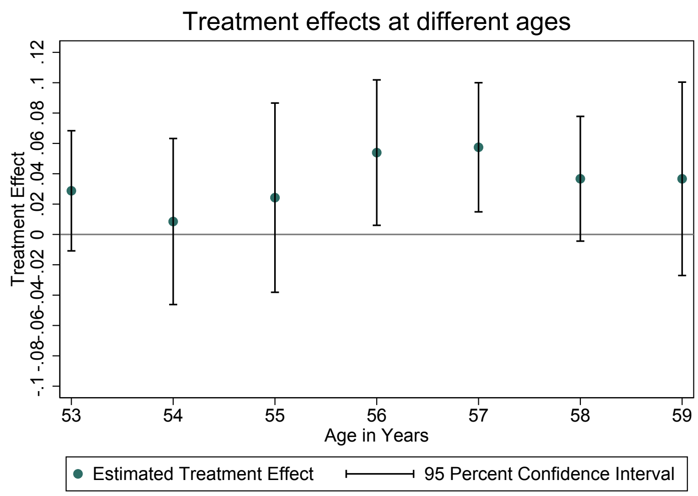
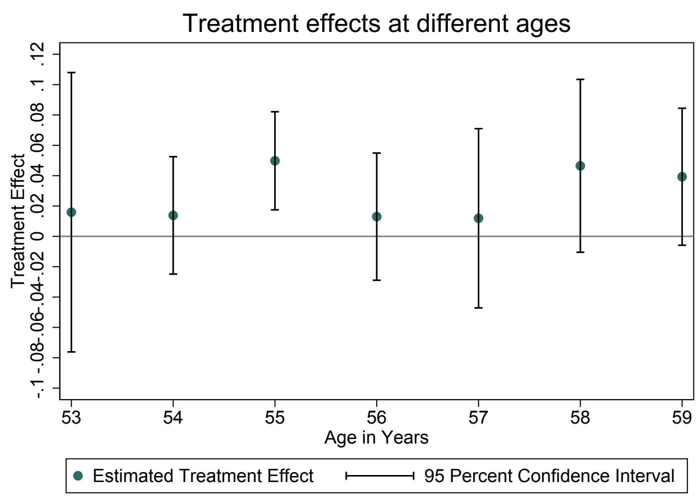


Figure 2.A.6: Employment Effect by Age



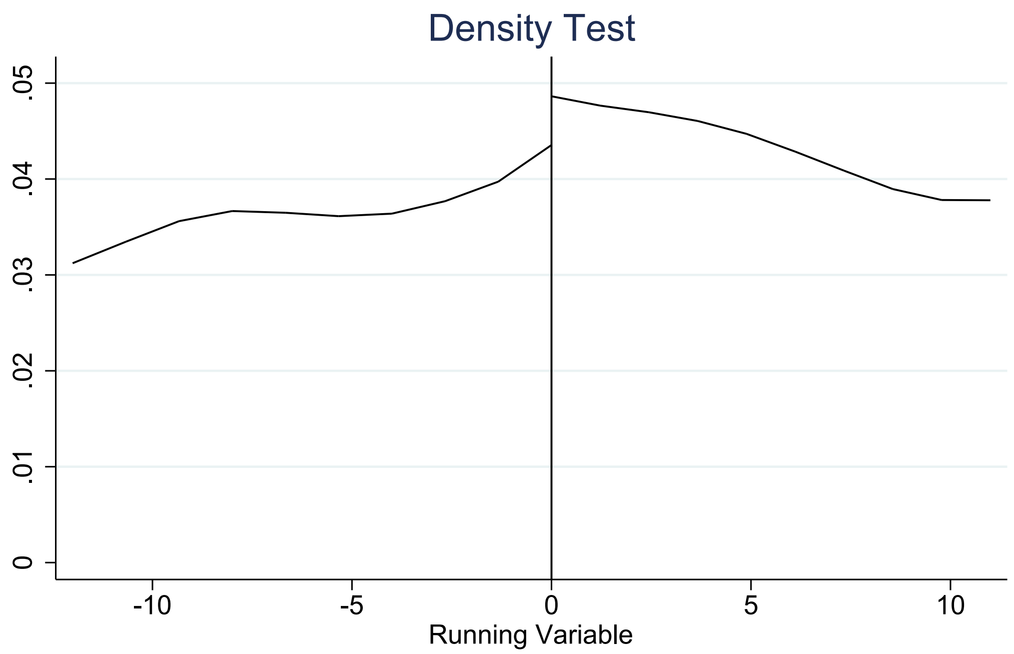
Note: Source: Microcensus 2005-2012; both graphs depict treatment estimates using the preferred quadratic RDD specification for the training outcome and the linear specification for the employment outcome controlling for covariates and running regressions for each age group separately.

Figure 2.A.7: Employment Effect before 60

Figure 2.A.8: Unemployment Effect before 60

Notes: Figure 2.A.7 and 2.A.8 plot the average employment and unemployment rates (y-axes) for all women born in 1951 and 1952 from 2005 onward by running variable, i.e. birth month (x-axis). The running variable denotes the distance to the cut-off measured in birth months.

Figure 2.A.9: Density Plot



The above figure plots the density on the y-axis and the running variable on the x-axis using a local polynomial approach of degree 2 following Calonico et al. (2018b). In our setting any jump in density around the cut-off relates to selection (employment effects) rather than manipulation. See section 2.5.1 for a discussion and Table 2.A.3 for density tests for sub-groups with no employment selection.

Table 2.A.1: The Effect on Personal Income (measured in Euro)

Local Regressions	(1)	(2)
Panel A: Employed Women		
Without Covariates		
Treatment Variable	-17.2570	15.5850
	(14.4350)	(18.2310)
With Covariates		
Treatment Variable	-19.2280	8.4177
	(18.6510)	(25.2550)
Panel B: All Women		
Without Covariates		
Treatment Variable	-1.9007	4.2213
	(17.9040)	(21.4920)
With Covariates		
Treatment Variable	-4.1830	-4.0053
	(15.2000)	(16.8490)
Age FE	✓	✓
Year FE	✓	✓
Running Variable	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at the birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Number of observations: Panel A: 27,179; Panel B: 50,035. The outcome variable measures the personal income denoted in Euro. Covariates: high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

Table 2.A.2: Training Effect: The Role of Educational Control Variables

	(1)	(2)	(3)	(4)	(5)	(6)
RD-Estimate	0.0272***	0.0395***	0.0248***	0.0408***	0.0244***	0.0405***
	(0.0077)	(0.0096)	(0.0077)	(0.0086)	(0.0078)	(0.0087)
Covariates	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Age FE	✓	✓	✓	✓	✓	✓
ISCED FE			✓	✓		
ISCED Sub FE					✓	✓
Running Variable	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Table 2.A.2 shows results from local linear and quadratic RDDs including education controls as in the main specification (1)-(2), ISCED fixed effects (3)-(4) and ISCED sub-category fixed effect (5)-(6). Source: Microcensus 2005-2012, own calculations.

Table 2.A.3: Density Manipulation Test

Specification	(1)	(2)	(3)	(4)
Test Statistic	1.9910**	1.5344	2.2518**	-0.0552
P-Value	0.0465	0.1249	0.0243	0.9560
Sample	Whole	Age drop	Non-College	College
Bandwidth	12	12	12	12

Notes: Table 2.A.3 depicts the test statistics of a density test based on Calonico et al. (2018b) as well as the corresponding p-value, the sample and the bandwidth. Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01. Column (1) uses the whole sample of employed women in cohorts 1951 and 1952, column (2) drops the age groups that drive the employment effects in the paper. Columns (3) and (4) split the sample by education level.

Table 2.A.4: Regression Discontinuity: Traning results excluding ages 55, 58 and 59.

Local Regressions	(1)	(2)
Without Covariates		
Treatment Variable	0.0293***	0.0442***
	(0.0099)	(0.0110)
With Covariates		
Treatment Variable	0.0298***	0.04406***
	(0.0079)	(0.0102)
Age FE	✓	✓
Year FE	✓	✓
Running Variable	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; Number of observations: 16,903. Covariates: high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

Table 2.A.5: Difference-in-Discontinuities: Employment Outcome

Local Regressions	(1)	(2)
Without Covariates		
Treatment Variable	-0.0030	0.0127
	(0.0084)	(0.0109)
With Covariates		
Treatment Variable	-0.0074	0.0077
	(0.0077)	(0.0095)
Age FE	✓	✓
Year FE	✓	✓
Running Variable	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; Number of observations: 99,316. Covariates: high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

Table 2.A.6: Training Effect for different Bandwidths

Bandwidth	6m	9m	12m	6m	9m	12m
Without Covariates						
Treatment Variable	0.0307**	0.0307**	0.0235**	0.0559***	0.0403**	0.0418***
	(0.0152)	(0.0130)	(0.0120)	(0.0168)	(0.0145)	(0.0131)
With Covariates						
Treatment Variable	0.0292***	0.0307***	0.0237***	0.0419***	0.0357***	0.0395***
	(0.0097)	(0.0085)	(0.0079)	(0.0100)	(0.0098)	(0.0088)
Age FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Running Variable	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level . Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; the table shows local regression results for different bandwidths. Number of observations 6m bandwidths: 13,941; Number of observations 9m bandwidths: 20,938; Number of observations 12m bandwidths: 28,519. Covariates: high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

Table 2.A.7: Donut Regressions

Local Regressions	(1)	(2)
Panel A: One month donut		
Without Covariates		
Treatment Variable	0.0149**	0.0241*
	(0.0064)	(0.0130)
With Covariates		
Treatment Variable	0.0195***	0.0338***
	(0.0045)	(0.0096)
Panel B: Two month donut		
Without Covariates		
Treatment Variable	0.0197***	0.0407***
	(0.0071)	(0.0089)
With Covariates		
Treatment Variable	0.0199***	0.0410***
	(0.0067)	(0.0071)
Age FE	✓	✓
Year FE	✓	✓
Running Variable	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01. Number of observations Panel A: 25,991; Number of observations Panel B: 23,375. Covariates: high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

Table 2.A.8: Employment Outcome: Heterogeneity by Educational Level

Local Regressions	(1)	(2)
Panel A: Non-College		
Without Covariates		
Treatment Variable	0.0217***	0.0105
	(0.0077)	(0.0102)
With Covariates		
Treatment Variable	0.0235***	0.0116
	(0.0080)	(0.0101)
Panel B: College		
Without Covariates		
Treatment Variable	0.0244	0.0209
	(0.0355)	(0.0528)
With Covariates		
Treatment Variable	0.0329	0.0322
	(0.0341)	(0.0491)
Age FE	✓	✓
Year FE	✓	✓
Running Variable	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; Participation rates in training for cohort 1951 are 0.154, 0.317, 0.117 for all employed women, employed women with college and employed women without college, respectively. Number of observations Panel A: 45,762; Panel B: 8,325. Covariates: indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

Table 2.A.9: Heterogeneity by Company Size

Local Regressions	(1)	(2)
Panel A: Big Company		
Without Covariates		
Treatment Variable	0.0087	0.0217
	(0.0143)	(0.0169)
With Covariates		
Treatment Variable	0.0079	0.0265**
	(0.0120)	(0.0119)
Panel B: Small-Medium Sized Company		
Without Covariates		
Treatment Variable	0.0324**	0.0566***
	(0.0146)	(0.0143)
With Covariates		
Treatment Variable	0.0322***	0.0482***
	(0.0114)	(0.0120)
Age FE	✓	✓
Year FE	✓	✓
Running Variable	Linear	Quadratic

Notes: Standard Errors in parentheses; clustered at birth month level. Sample includes twelve month before and after reform cutoff. Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01. Big companies in the Microcensus are classified as companies with 50 or more employees. Number of observations Panel A: 14,750; Panel B: 13,769. Covariates: high initial level of education indicator (ISCED 97 Level 5-6), and a variable for medium education (ISCED 97 Level 3-4) as well as an indicator variable for west Germany. Source: Microcensus 2005-2012, own calculations.

2.B Appendix B: Theoretical Model

In this Appendix, we derive a stylized theoretical human capital model and show that individuals, *ceteris paribus*, have an incentive to increase on-the-job training when working life increases exogenously. The central mechanism for this human capital effect is that the returns to training increase with the remaining working life of an individual i, denoted by R_i.[34] [35] The theoretical model presented below illustrates the mechanism in a simplified and intuitive setting through a discrete time model consisting of three stylized periods. Note that Y_{ti} denotes an individual i's income in period t and C_{ti} denotes the level of consumption in period t. Each individual derives utility through consumption, $U(C_{ti})$, with the standard assumption of $U'(C_{ti}) > 0$, $U''(C_{ti}) < 0$. Further, each individual has earnings E_t, which are a composite of individual wages, employment security and quality of work. Earnings in period one depend on the initial level of schooling, i.e. $E_1(S_i)$, which is determined exogenously prior to period one. Earnings, ceteris paribus, are increasing with education, specifically we assume $E_1'(S_i) > 0$, $E_1''(S_i) < 0$. In the first period, the individual decides on his or her time investment in human capital, I_i, through participation in on-the-job training measures. Earnings in period two depend on the chosen level of training, specifically we assume $E_2'(I_i) > 0$, $E_2''(I_i) < 0$.

1. Period

Income in the first period consists of labor in income which varies with the level of initial schooling, S_i:

$$Y_{1i} = E_1(S_i)$$

2. Period

[34] Note that the remaining working life in our stylized theoretical setting is treated as exogenous.

[35] An analogous human capital effect can be generated in a model of firm's investment decision when the working life of workers increases. When workers are not perfectly mobile (Acemoglu and Pischke, 1998, 1999), the intuition is straight forward in our model: The longer the payout period for the investment of the firm, i.e. the longer the worker stays in the firm, the higher the investment in human capital.

Income in period two is given by:

$$Y_{2i} = E_2(I_i)R_i \tag{2.B.1}$$

where R_i is the duration of the remaining working life and I_i denotes the level of human capital investment.

3. Period

Income in period three is given by:

$$Y_{3i} = \alpha E_2(I_i)(T_i - R_i) \tag{2.B.2}$$

Period three is the period of retirement. The duration of period three is $T_i - R_i$, where T_i is the individual life expectancy. We assume that retirement is a discrete decision to exit the labor market completely. Income in the retirement period is covered by the state pension, which is a fraction α, with $\alpha < 1$, of labor income in period two.
Utility over all three periods then is given by:

$$U_G = U_{1i} + \beta U_{2i}R_i + \beta^2 U_{3i}(T_i - R_i)$$

Where:

$$U_{1i} = U(C_{1i} - a(S_i)I_i)$$

$$U_{2i} = U(C_{2i})$$

$$U_{3i} = U(C_{3i})$$

Following e.g. Blundell et al, (2019), we assume that individuals face utility cost of training in period 1, $a(S_i)$, which fall with schooling S_i, i.e. $a'(S_i) < 0$.
Individuals maximize U_g subject to the intertemporal budget constraint, which is given by:

$$y_{1i} + \beta y_{2i} R_i + \beta^2 y_{3i}(T_i - R_i) \geq C_{1i} + \beta C_{2i} R_i + \beta^2 C_{3i}(T_i - R_i)$$

Hence, the Lagrangian is:

$$L = U_G + \lambda \left[y_{1i} + \beta y_{2i} R_i + \beta^2 y_{3i}(T_i - R_i) - (C_{1i} + \beta C_{2i} R_i + \beta^2 C_{3i}(T_i - R_i)) \right]$$

and the set of First Order Conditions is given by:

$$\frac{\partial L}{\partial C_{1i}} = 0 \Rightarrow U'(C_{1i} - a(S_i)I_i) - \lambda = 0 \tag{2.B.3}$$

$$\frac{\partial L}{\partial C_{2i}} = 0 \Rightarrow \beta R(U_i'(C_{2i}) - \lambda) = 0 \tag{2.B.4}$$

$$\frac{\partial L}{\partial C_{3i}} = 0 \Rightarrow \beta(T - R)\beta^2[\lambda - U(C_{3i})] = 0 \tag{2.B.5}$$

$$\frac{\partial L}{\partial I_i} = 0 \Rightarrow \lambda[-a_1(S_i) + (R_i\beta + (T_i - R_i)\alpha\beta^2)E_2'(I_i)] = 0. \tag{2.B.6}$$

Based on the set of First Order Conditions and applying the implicit function theorem, we derive our results. More precisely, we can take the derivative of 2.B.6 with respect to R_i and I_i respectively. Dividing both terms with each other and multiplying the result by negative one, then according to the implicit function theorem gives $\dfrac{\partial I_i}{\partial R_i}$ denoted below in 2.B.7.

$$\frac{\partial I_i}{\partial R_i} = -\frac{\dfrac{\partial L}{\partial I_i \partial R_i}}{\dfrac{\partial^2 L}{\partial^2 I_i^2}} = \frac{[\alpha\beta - 1]\, E_2'\,(I_i)}{[R_i + (T_i - R_i)\alpha\beta]\, E_2''(I_i)} \tag{2.B.7}$$

An increase in the remaining working life, R_i implies a positive impact on training as long as $E_2' > 0$ and $E_2'' < 0$. Note that $\alpha\beta - 1$ is negative since $\beta < 1$ and $\alpha < 1$. Hence, we can derive the main proposition that motivates the empirical analysis of our paper.

Proposition I (Working life effect)

The effect of an increase in the working life on training is positive.

Secondly, the effect of training participation with respect to schooling is captured by Equation 2.B.8:

$$\frac{\partial I_i}{\partial S_i} = \frac{a'(S_i)}{\beta\,[R_i + (T_i - R_i)\alpha\beta]\, E_2''(I_i)} \tag{2.B.8}$$

As long as the utility cost of training falls with the initial level of schooling, i.e. $a'(S_i) < 0$, the effect will be positive, which implies Proposition II.

Proposition II (Initial education effect)

The level of time investment in training rises with initial schooling as long as utility cost of training are falling with the initial level of schooling.

Thus, the model captures the empirical finding that higher educated individuals participate in training more often, which is described as a dynamic complementary between initial education and later training, see e.g. Cunha and Heckmann (2007) and Jacobs (2009).

Chapter 3

Working Longer, Working Stronger? The Forward-Looking Effects of Increasing the Retirement Age on (Un)employment Behaviour.[1]

Abstract Leveraging two cohort-specific pension reforms, this paper estimates the forward-looking effects of an exogenous increase in the working horizon on (un)employment behaviour for individuals with a long remaining statutory working life. Using difference-in-differences and regression discontinuity approaches based on administrative and survey data, I show that a longer legal working horizon increases individuals' subjective expectations about the length of their work life, raises the probability of employment, decreases the probability of unemployment, and increases the intensity of job search among the unemployed. Heterogeneity analyses show that the demonstrated employment effects are strongest for women and in occupations with comparatively low physical intensity, i.e., occupations that can be performed at older ages.

[1]Single-authored.

3.1 Introduction

Rising life expectancy is putting pressure on pay-as-you-go pension systems in most developed economies. In response, many OECD countries have raised their statutory retirement ages over the past three decades to extend the working lives of their citizens and reduce the fiscal burden of an aging society. Many of these pension reforms were announced before the affected individuals actually reached retirement age, which can lead to shifts in expected work horizons and affect employment outcomes and job search behaviour among relatively young individuals with long remaining working lives.

This study explicitly examines these potential forward-looking effects of pension reforms. To this end, I use both difference-in-differences (DiD) and regression discontinuity (RDD) approaches, analyze high-quality administrative and survey data, and exploit cohort-specific pension reforms in Germany to estimate the effects of an exogenous extension of statutory working lives on individual pension expectations, anticipatory employment responses, and job search behaviour.

The study documents four main findings. First, I show that even small changes in the statutory working life caused by pension reforms lead to a shift in the length of individuals' subjectively expected working horizon. Second, a longer working horizon significantly increases employment probabilities of relatively young individuals and has a negative effect on the probability of being registered as unemployed during a given year. The positive employment effects are driven by inflows into employment as well as lower outflows out of employment. Third, the paper provides suggestive evidence that individuals with a longer working horizon have a higher job search intensity and receive more job offers by job agency officials. Fourth, exploiting the large sample size of the administrative data set used in this study, I show that the documented employment effects are driven by individuals in jobs with lower physical intensity, i.e., jobs that can be exercised at older ages.

The exogenous variation used to identify the described effects stems from two pension reforms in Germany.[2] First and foremost, I study the 2007 pension reform, which step by step shifted the normal retirement age (NRA) from 65 to 67 for cohorts 1947-1964. In order to isolate causal estimates I employ a difference-in-differences inspired approach, exploiting the variation in the treatment intensity between adjacent cohorts. The reform was highly debated and very salient across German media and sparked a general debate about longer working horizons of younger generations. Further, it affected a large set of individuals and given the large sample size of the administrative data set used in this study is well-suited to conduct an extensive set of heterogeneity analyses. Second, to supplement the results of the 2007 pension reform I analyse the 1999 pension reform using a regression discontinuity approach. The reform amongst others implied an immediate, exogenous cohort-specific and sizeable shift in the early retirement age (ERA) for women born after 1951 from 60 to 63. Crucially, the reform was announced in 1998 and passed in 1999 when the affected women, i.e. women born in 1952 and aged 47, still had a long remaining working life - another setting well-suited to study forward-looking effects.

A range of theoretical considerations point towards plausible forward-looking effects of pension reforms. First, taking a life-cycle perspective, an increase in the retirement age can be interpreted as a negative wealth shock since, ceteris paribus, it reduces the time over which pensions are paid and accordingly individuals' social security wealth. Non-myopic individuals might respond to such a negative wealth shock by increasing their current labour supply, (see e.g. French (2005a); Geyer and Welteke (2021)).[3]

Second, human capital theory, starting with Becker (1962) and Ben-Porath (1967), predicts that the value of human capital investment increases with the payout period of the investment.[4] An increase in the pension age extends the payout period and thus may induce more investment into human capital. Indeed, recent literature provides empirical

[2]In Germany, state pension rules do not only provide economic incentives but, as documented by Seibold (2021), set the legal framework and social norm for the duration of working life. The pension reforms analysed in this paper therefore generate variation in the expected duration of the working life, ultimately allowing to study the forward-looking effects of the pension reform.

[3]Note, however, that an increase in the working horizon from a life-cylce point of view may also reduce the incentive to work, as given the longer contribution period, individuals can collect the same level of pension benefits with fewer contributions in younger years (Geyer and Welteke, 2021).

[4]Generally, this is shown by the strong positive relationship between life expectancy and human capital investment such as school education and/or participation in vocational training documented in most countries over the past century, see e.g. Soares (2005); Cervellati and Sunde (2013).

causal evidence that exogenous increases in the pension age positively impact human capital investment for the employed prior to reaching retirement (Gohl et al., 2021; Bauer and Eichenberger, 2017; Brunello and Comi, 2015), which may increase employability (also see Chapter 2 in this dissertation). In theory, similar results may hold for the unemployed when facing a shift in their working lives. In practice, there are different ways in which the unemployed can accumulate human capital. For example, human capital could be accumulated via the participation in active labour market programs such as vocational training measures or via an increase in job search activity.[5] Hairault et al. (2010), for example, calibrate and simulate a model where job search is dependent on the distance to retirement. In their model search effort is costly and increases with the remaining duration of the ensuing job, implying a shorter unemployment duration for individuals with a longer working horizon. Moreover, not only the unemployed themselves but also job agency officials have an incentive to consider legal working life requirements. For example, one goal of employment agencies could be to reduce the permanent financial pressure on unemployment insurance by arranging more transitions to employment and/or offering more training opportunities in the hope that this will ultimately increase employment.[6]

Last, from an employers' perspective, firms may have a stronger incentive to hire employees that have a longer remaining working life, as this potentially reduces costly job fluctuations and the cost of rehiring and retraining.

Despite these theoretical considerations, forward-looking employment effects of pension reforms are relatively under-studied in economic literature. Existing empirical research predominantly focuses on direct employment effects of pension reforms, generally showing that employment rates between the old and new retirement thresholds increase for individuals with longer statutory retirement ages (Staubli and Zweimueller, 2013; Geyer

[5]Vocational training measures have been shown to entail a short term negative effect on employment, as search intensity and job offers from job center officials decrease during the period of training (van Ours, 2004). In the medium to long term, however, there is wide spread evidence of increased employment due to training programs (Card et al., 2010).

[6]In particular, this goal may be more pronounced for individuals with a longer working horizon, as longer working requirements may imply prolonged unemployment, which in turn implies a longer financial burden on the unemployment insurance system.

et al., 2020; Manoli and Weber, 2016; Atalay and Barrett, 2015; Inderbitzin et al., 2016; Geyer and Welteke, 2021; Lalive and Staubli, 2015). Additionally, some studies find evidence for program substitution such as an increase in the take-up of unemployment or disability benefits after the former retirement age in order to substitute for the abolished retirement option (Staubli and Zweimueller, 2013; Atalay and Barrett, 2016).

In contrast, empirical literature focusing on identifying forward-looking effects of pension reforms on (un)employment and job search responses is scarce and the documented evidence mixed. Hairault et al. (2010), using a difference-in-differences approach to supplement their simulation model, find positive albeit insignificant employment effects of a pension reform in the 1990s in France for men with different distances to retirement. Engels et al. (2017) focus on a cohort-specific reduction in the generosity of the early retirement option for women born in the cohorts 1938-1944 in Germany. They show that a stronger penalty for early retirement reduces duration of unemployment before reaching the early retirement threshold. French et al. (2022) analyse how changes in the way pension contributions are calculated impact employment outcomes before reaching retirement, documenting an increase in labour supply. Geyer and Welteke (2021) focus on individuals just before the retirement threshold and do not find any evidence for forward-looking effects stemming from an increase in the early retirement age. Gohl et al. (2021), using the same reform as Geyer and Welteke (2021) but focusing on younger individuals, find some albeit not very strong evidence for forward-looking employment effects (see Chapter 2). Frimmel (2021), using a pension reform in Austria, documents increases in unemployment training participation for men and employment probabilities for both men and women in response to a shift in the retirement age. Studying a pension reform in Italy, Carta and de Philippis (2021) analyse an increase in the full statutory retirement age from 60 for women and 65 for men to 67 for both using a difference-in-differences inspired approach. They find significant labour force participation effects for women aged 45-59. In particular, they find that younger women exit inactivity to enter unemployment whereas older women exit unemployment into employment in response to the reform. For men they find small and insignificant effects.

This paper contributes to the existing literature in various ways. First, it expands the scarce and mixed empirical literature focusing on forward-looking effects of pension reforms in response to an increase in the statutory retirement age. In particular, it studies the forward-looking employment and unemployment responses to a pension reform that increased the retirement age incrementally and by a small margin from cohort to cohort. Crucially, it confirms theoretical predictions based on human capital theory and existing empirical evidence dealing with large increases in the retirement age by providing new causal empirical evidence that pension reforms that raise the retirement age, even if only by a small margin across cohorts, do indeed increase individuals' subjective expectations about the length of their working lives and imply forward-looking employment effects.

Second, this paper explicitly examines heterogeneous outcomes and, as an important new result, shows that employment effects are particularly pronounced for individuals in service sector occupations that require lower physical intensity and can therefore be performed at older ages.

Third, it provides novel empirical evidence for potential behavioural mechanisms driving the documented employment effects. In particular, it provides suggestive evidence supporting the notion that the expected retirement age plays a key role in job search behaviour.

On a more general level, this paper shows how pension policies, job search and active labour market policies interact. This is crucial as understanding the interaction between these policy tools is key to promoting employment amongst middle-aged and older workers, while simultaneously lifting pressure from pay-as-you go pension systems. In particular, the heterogeneous results documented in this paper highlight that pension reforms affect individuals and occupations differently. To prevent pension reforms and the extension of working lives from exacerbating inequality in old age, policymakers can actively take into account different responses to the rising retirement age when promoting employment in old age through pension reforms or labor market policies. Last, the paper adds to the broader literature examining the relation between pension ages and employment responses in general.

The remainder of the paper is organized as follows. Section 3.2 describes the institutional setting. Section 3.3 describes the data and sample selection. Section 3.4 presents the empirical setting of the study and Section 3.5 presents the main results as well as corresponding robustness checks and alternative specifications. Section 3.6 sheds light on possible mechanisms for the documented effects and Section 3.7 and 3.8 analyse results for further outcome variables as well as heterogeneous effects. Section 3.9 discusses the results and potential policy implications. Finally, Section 3.10 concludes.

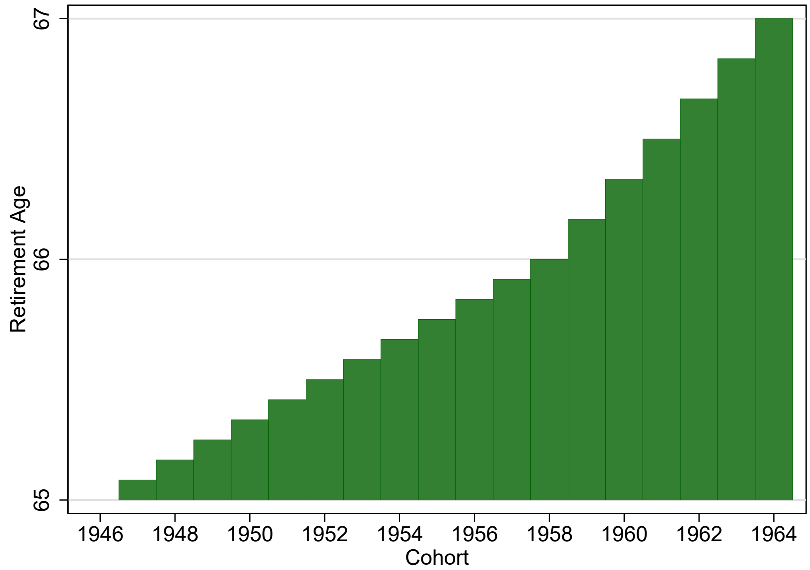
3.2 Institutional Setting

Before turning to the empirical analysis, this section provides a brief overview of the relevant aspects of the German pension system and the 2007 and 1999 pension reforms. Further, it provides an overview of the German unemployment insurance system and its rules, as they might shape age-specific individual employment behaviour and thus need to be taken into account.

The German Pension System The statutory public pension system is the central part of the retirement system in Germany. It covers more than 80% of the workforce with the exceptions of groups that are not subject to compulsory pension insurance, most importantly civil servants and self-employed. It includes old-age pensions, disability pensions, and survivors benefits. The system is financed as a pay-as-you-go (PAYG) scheme and has a strong contributory link: pension benefits depend on the entire working history. The pension system provides several pathways into early retirement, i.e. claiming retirement benefits before reaching the normal retirement age with actuarially fair deductions of 0.3 percent per month.[7] Throughout the 1990s and 2000s the German public pension system was subject to a range of reforms. The paper mainly focuses on the 2007 reform and as an extension the 1999 pension reform.

[7]Actuarial fairness stipulates that the present value of lifetime pension benefits equals the present value of lifetime pension contributions. Actuarially fair deductions ensure that this holds for early retirement entries.

Figure 3.1: Increase in NRA induced by 2007 Pension Reform

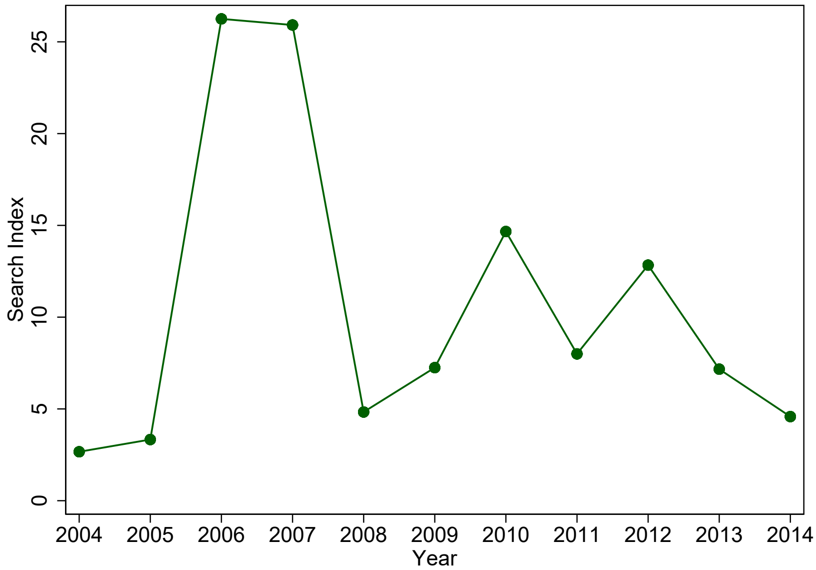


Notes: Figure 3.1 depicts the new normal retirement age (on the y-axis) by cohort (on the x-axis) as stipulated by the 2007 pension reform.

The 2007 Pension Reform In April 2007, Germany officially raised the statutory normal retirement age (NRA) from 65 to 67. The NRA was increased step-by-step for cohorts born between 1947 and 1964. Figure 3.1 illustrates the incremental increase. Starting with cohort 1947 the normal retirement age was raised by one additional month for each subsequent cohort up until cohort 1958. For cohorts 1959-1964 the NRA was increased by two additional months for each consecutive cohort. The reform effectively implied that from 2012 onward individuals could no longer retire at age 65 without deductions. Generally, the reform was very salient and strongly discussed in public. Figure 3.2 plots the Google search trends for "retirement with 67", showing clear peaks in Google searches in 2006-2007 and 2010-2012, the years in which the reform was most heavily debated: discussions and media coverage of the reform started in 2006 and continued throughout 2007. In 2010, the debate concerning retirement with 67 picked up again, with the Social Democrats, the junior party in Germany's grand coalition government at the time, entertaining the idea of postponing the rise in the NRA, which eventually did not happen.[8] In 2012, the first affected cohort, i.e. cohort 1947, retired at an age higher

[8]See for example FAZ.Net (2010)

Figure 3.2: Google Search Trends



Notes: Figure 3.1 depicts the yearly average index value for Google search trends of "retirement with 67". The index can range from 100 to 25 and was aggregated from a monthly to a yearly summary statistic in the figure above.

than 65, which again gained strong media coverage and public attention, as reflected in another increase in Google searches.

The 1999 Pension reform To supplement the results from the main analysis using the 2007 pension reform, I additionally consider the 1999 pension reform, which officially came into force on 1st January 1999 and, amongst others, abolished the *pension for women* for cohorts born after 1951. Prior to the reform, this early retirement option allowed women to enter retirement from age 60 onward.[9] The reform thus effectively raised the early retirement age (ERA) for most women from age 60 to age 63 and therefore increased the working life. Crucially, the reform was cohort-specific. Women born before 1952 could still claim the *pension for women* with the same qualifying conditions as before the reform. The eligibility criteria were: (i) at least 15 years of pension insurance contributions; and

[9]Additionally, the *pension after unemployment or after old-age part-time work* was abolished for individuals born after 1951 at the same time as the *pension for women*. Unemployed individuals born before 1951 could enter early retirement at the age of 63 whereas unemployed individuals born after 1951 could no longer do so. For more details see Geyer et al. (2020). For a more general description on the German pension system, see Boersch-Supan and Wilke (2004).

(ii) at least 10 years of pension insurance contributions after the age of 40. According to Geyer and Welteke (2021), about 60% of all women born in 1951 were eligible for the old-age pension for women. Self-employed women and women working in the civil service by default were not eligible for the old age pension for women. Women born after 1951 were now required to have a total of 35 contribution years in order to qualify for early retirement at the age of 63.

In contrast to the 2007 pension reform, the direct effects of the 1999 pension reform have been extensively studied. For example, Geyer and Welteke (2021) and Geyer et al. (2020) evaluate the labour market effects of the 1999 pension reform, focusing on employment effects between the old and new retirement thresholds. They document that the increase in the ERA has a sizable positive effect on the working life of individuals. More precisely, employment rates for eligible women aged between 60 and 62 increase by about 15 percentage points. The combined effect on inactivity and unemployment has a similar size with about 12 percentage points.

The German Unemployment System Besides the public pension system, the German unemployment system, its benefit payment scheme, and reforms are of central interest to this paper, as they potentially influence employment behaviour and decisions. In Germany, individuals are entitled to unemployment benefits if they have contributed to unemployment insurance for a minimum time of twelve months during the last two years before unemployment. The contribution rate to the unemployment insurance is 2.4 percent of a person's gross earnings, and contributions are usually shared equally between employer and employee. The replacement rate for singles lies at 60 percent of the average net wage in the last 12 employed months before unemployment. The maximum duration of unemployment benefit payments is 12 months for individuals below the age of 50. For individuals above the age of 50 this period further increases to 15 months. For individuals between 55 and 57 the maximum duration of unemployment benefits is 18 months, and individuals above the age of 58 can obtain unemployment benefits for a maximum duration of 24 months.[10] Once the claim to unemployment benefits expires, individuals are entitled to social assistance.[11]

Table 3.1: Benefit Duration (in months) by Age Group

Age	Before 02/2006	After 01/2006	After 12/2007
< 45	12	12	
45-46	18	12	
47-51	22	12	
50-54			15
52-54	26	12	
55-56	26	18	
57	32	18	
≥ 58	32	18	24

Source: Dlugosz et al. (2009), Table 3.1 depicts the different benefit duration rules in each period and for each age group (i.e. before and after the respective reforms). Note that in addition to individuals' age their work experience during the previous years decided eligibility to unemployment benefits.

The paper focuses on a period beginning in 2000, so it is critical to consider changes in benefit duration rules during the period under study. The rules described above were adopted with the Hartz reforms, which were ratified at the end of 2003 and came into force from February 2006. Further, there was a second change in rules in December 2007, marginally adjusting the 2006 rules. Prior to these reforms, the benefit scheme was more generous. For example, already from the age of 52 conditional on having worked for 48 months within the last 7 years individuals could claim unemployment benefits for a maximum duration of 24 months.[12] Table 3.1 illustrates the different benefit duration rules for each period/policy regime. The empirical strategy in this paper fully accounts for differences in claim durations across age groups and changes in benefit duration rules throughout the sample period, as described in more detail below.

[10] In theory, women who enter unemployment at the age of 58 and are eligible for early retirement at 60 thus can bridge the transition into retirement with unemployment benefits.

[11] Arbeitslosengeld II from 2005 onwards, before 2005 Arbeitslosenhilfe and Sozialhilfe.

[12] For a detailed description of unemployment rules prior to the Hartz reforms see Caliendo et al. (2013)

3.3 Data and Sample

I use three different data sets. First, I use the Sample of Integrated Employment Biographies (SIAB). Second, I supplement this administrative data set with data from the SAVE Survey and third I use the IZA Evaluation Dataset Survey.

1) Sample of Integrated Employment Biographies The SIAB data set is a two percent random sample drawn from the Integrated Employment Biographies (IEB) of the German Institute for Employment Research (IAB). The IEB consist of all individuals in Germany who can be described by at least one of the following employment statuses: employment subject to social security contributions[13], marginal employment [14], unemployed and entitled to benefits or basic support, and participation in programs of active labour market policies. Based on these categories, the data documents detailed employment histories from 1975 onward for West Germany and from 1992 onward for the whole of Germany. Following Dauth and Eppelsheimer (2020), I create a yearly panel data set from 2000-2014. Crucially, as key outcome variables I observe whether an individual in a given year is employed and thus contributing to social insurance or registered as unemployed. Further, I observe marginal employment, participation in other active labour market programs or whether individuals received funding to set up their own business.

I explicitly focus on the time period starting from 2000. Before 2000, some of the variables of interest were not recorded coherently[15] and un/employment spells in some regions were incorrectly counted/registered. Moreover, in my main specification, I focus on the years up to 2014 to isolate the effects of the 2007 pension reform. In 2014, there was another pension reform that adjusted the retirement age for an early retirement option, which also had a differential impact on the cohorts studied. In a robustness check, I extend the period to 2017, the last year I observe in the data.

[13]This excludes civil servants and the self-employed.

[14]Employment in so called mini-jobs that are not contributing to social insurance payments.

[15]For example, participation in active labour market programs such as training for the unemployed was not recorded before 2000.

2) Saving and old-age provision in Germany (SAVE) The SAVE Survey is a repeated cross sectional survey, which took place in Germany from 2001-2013 at irregular intervals. Originally launched in 2001, it was repeated in 2003/2004 and then at regular annual intervals until 2013. Crucially for this study, the survey elicits subjective beliefs about retirement ages by asking "*What do you expect - at what age will you presumably retire or receive pension payments*". Additionally, it collects a range of socioeconomic background information such as age, gender and educational background. For a detailed description of the survey see Börsch-Supan et al. (2008).

3) IZA Evaluation Dataset Survey The IZA Evaluation Panel is survey data tracking employment history, behaviour and individual traits for an inflow sample of more than 17,000 unemployed individuals born from 1952 onward. The data was collected for unemployment entries between June 2007 and May 2008, i.e. just after the pension reform. Individuals were tracked for three years after their initial unemployment entry. Crucially, it contains information about the job search behaviour of newly unemployed individuals within the first two months of unemployment, such as the number of applications and job offers received by job agency officials. Further, as described above it tracks unemployed individuals for up to three years, allowing to analyse re-employment probabilities within three years.

Sample In order to analyse employment effects of the 2007 pension reform, I focus on a sample of individuals born in the cohorts 1959-1964 for whom the retirement age was gradually increased in two month steps for each additional cohort. Individuals in these cohorts were aged 44-55 in the post reform period, i.e. 2008-2014.

The reason for exclusively focusing on these cohorts is threefold: first, the unemployment benefit duration reforms described above were implemented shortly before and after the pension reform 2007, i.e. in 2006 and 2008. Due to age-specific changes in the generosity of benefit payment lengths induced by the reforms, they might have affected cohorts differently, thereby potentially undermining identification. For example, Dlugosz et al. (2014) show that younger age groups in cohorts 1959-1964 did not adjust their employment responses to the 2006 unemployment benefit duration reform, whereas older individuals in earlier cohorts did. Further, due to their relatively young age, individuals in cohorts 1959-1964 were not directly affected by the second reform in 2008, which has been shown to have decreased the job search effort of individuals in earlier cohorts who were directly affected by the reform and whose benefit duration was increased by the reform (Lichter and Schiprowski, 2021). Focusing exclusively on younger cohorts, i.e., those born between 1959 and 1964, thus allows excluding potential confounding effects created by the age-specific reforms in the duration of unemployment benefits and their timing.[16]

Second, the shift in the NRA for cohorts 1959-1964 differed by two months for adjacent cohorts, thereby providing stronger variation in the treatment intensity between cohorts, which will be used in the empirical strategy of the paper.

Last, by restricting the analysis to cohorts that are relatively similar in age, I ensure that I only compare individuals whose employment biographies would have evolved similarly in the absence of treatment, a key requirement for identification, as described in more detail in Section 3.4.

As previously described, Carta and de Philippis (2021) find that younger women select out of inactivity into unemployment and older women into employment in response to a pension reform in Italy. Unlike Carta and de Philippis (2021), this paper, in its main specification, only uses variation from individuals that were part of the SIAB dataset already before the reform, i.e. between 2000 and 2007, thereby capturing the response of individuals actively participating in the labour market. In addition, to account for possible selection into employment, I condition on individuals having at least one spell of employment or unemployment prior to 2006, the year in which the reform was first discussed. This ensures that I observe individuals that are likely to actively participate in the labour market and do not select into activity (employment or unemployment) once

[16]The empirical strategy will also control for age group fixed effects ensuring that only individuals from age groups that receive the same benefits are compared with one another. However, these age group fixed effects do not pick up changes over time in (un)employment behaviour in response to reform and thus the sample restrictions described above are necessary.

the reform was first discussed and then passed.

Last, in order to confirm my main results I extend the analysis of forward-looking pension reform effects to the pension reform 1999 using the SIAB data. Here I focus on a sample of women born in 1951 and 1952 observed between 2000-2008 aged 50-58.[17]

3.4 Empirical Approach

In order to analyse the 2007 pension reform, I exploit the variation in treatment intensity between cohorts 1959-1964 in an approach similar to Carta and de Philippis (2021). In particular, I implement a DiD estimator with a multi-valued treatment variable (Callaway et al., 2021) capturing the reform-induced change in the normal retirement age measured in months. Effectively, this approach compares outcome values before and after the reform among groups with different treatment doses, thereby estimating the average causal response of the outcome variables of interest to a one-month increase in the NRA. Equation 3.1 describes the approach in more detail.

$$y_{it} = \beta D_i * post_t + \phi_{tb} + \zeta_{i/c} + \alpha_a + X_i'\delta + \epsilon_{it} \tag{3.1}$$

y_{it} stands for a range of outcome variables such as retirement expectations or employment and unemployment indicators for individual i in year t.

D_i can be thought of as the treatment dosage, which in this setting is a time-invariant variable denoting an individuals' change in the statutory retirement age measured in months induced by the reform. More precisely, D_i is given by $D_i = NRA_{>2007,i} - NRA_{<2007,i}$ where $NRA_{>2007,i}$ is an individuals' normal retirement age after the 2007 pension reform and $NRA_{<2007,i}$ is an individuals' normal retirement age before the reform. Table 3.A.1 in the Appendix details the statutory retirement ages and changes induced by the reform for each cohort respectively. $post_t$ is an indicator variable that is equal to one

[17] I explicitly exclude women above the age of 58 as these women in theory could directly enter retirement at 60 after claiming benefits for two years.

in the post-reform period and zero otherwise. ϕ_{tb} are bi-cohort, b, times year fixed effects accounting for any time trend that similarly affects adjacent cohorts, i.e. individuals that are very similar in age and thus likely to be affected by macroeconomic trends in the same way.[18] $\zeta_{i/c}$, depending on the data set and specification used, are either individual i fixed effects, controlling for any time-invariant unobservables at the individual level or cohort c fixed effects, controlling for any time-invariant cohort-specific unobservables.[19] α_a are age group fixed effects for each age group that receives the same length of benefit entitlement in a given year (see Table 3.1). The inclusion of these fixed effects ensures that only individuals with the same statutory benefit entitlement length are compared to each other. Depending on the specification and the data set, I also control for a range of covariates. In particular, when the inclusion of individual fixed effects is not possible, I include individual-specific socioeconomic covariates such as educational attainment and a female indicator. Finally, ϵ_{it} is an error term, which is clustered at the individual level.

Conditional on the included fixed effects and covariates the point estimate of β identifies the average causal effect of an additional one month increase in individuals' normal retirement age if the parallel trend assumption holds: in the absence of the reform the outcomes of individuals with different treatment doses, i.e. here different changes to their retirement age, would have evolved parallely over time. Additionally, recent advances in the DiD literature have shown that in order for the empirical strategy to truly identify the average causal response on the treatment group (ACTR), it also needs to hold that if groups would have received any level of the same dosage their outcomes would have evolved paralelly (Callaway et al., 2021). This somewhat stronger assumption is only likely to hold for cohorts that are very similar in unobservable and observable characteristics, i.e. adjacent cohorts that are close in age and likely to face similar career and life circumstances. The inclusion of the year times bi-cohort fixed effects, the individual fixed effects (when possible) as well as the general sample restriction to cohorts 1959-1964 ensures that the empirical strategy solely exploits variation between cohorts that are in-

[18]Note that when using the SAVE data, I solely use year fixed effects, as the number of observations is relatively low and the inclusion of bi-cohortly year fixed effects requires a relatively large number of observations in order to ensure sufficient variation.

[19]The SAVE data essentially is a repeated cross section where only relatively few individuals are surveyed repeatedly and thus the inclusion of individual fixed effects is not possible.

deed close in age, controls for any time-invariant individual observables and accounts for age-specific macroeconomic trends.

Further, in order to test for common pre trends, I implement an approach similar to an event study design and run a regression with interaction terms between D_i, the multi-valued treatment variable, and each respective year from 2000-2014. This also allows to analyse the dynamic response to the reform. Equation 3.2 describes the approach in more detail. All variables are the same as in Equation 3.1 and $\sum_{t=2000}^{2014} \beta_t D_i * \phi_t$ stand for the described interaction terms of D_i and year indicators ϕ_t.

$$y_{it} = \sum_{t=2000}^{2014} \beta_t D_i * \phi_t + \phi_{tb} + \zeta_{i/c} + \alpha_a + X'_{it}\delta + \epsilon_{it} \tag{3.2}$$

3.5 Results

The following sections present the main results and corresponding robustness checks for the pension reform 2007. First, I focus on individuals' subjective retirement beliefs and then analyse individuals' (un)employment responses to the pension reform. Last, as a supplementary extension I will present results for the 1999 pension reform.

3.5.1 Main Results

Expectation Results Only if individuals perceive the reform can they react to it. Therefore, for the approach to truly reveal individuals' reactions to pension reform, it is first necessary to analyze whether individuals actually perceive the reform as a shift in their expected work horizon. In order to do so, I use the repeated cross-sectional data provided by the SAVE survey and individuals' expected retirement age as the key outcome of interest. For the relevant cohorts, i.e. cohorts 1959-1964, I observe approximately 3,000 survey responses. Before the reform between 2001 and 2006 individuals in the affected cohorts on average expect to retire at age 63.5. The post-reform average is substantially

Table 3.2: Expectation Results

	(1)	(2)	(3)	(4)
Interaction	1.2292***	1.2333***	0.0005	0.0005
	(0.4363)	(0.4370)	(0.0006)	(0.0006)
Covariates	-	✓	-	✓
Year FE	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓
Reform	True	True	Placebo	Placebo
Cohorts	1959-64	1959-64	1964-69	1964-69
N	2,917	2,917	2,845	2,845

Note: Source: SAVE data, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the expected retirement age measured in months as outcome variable for SAVE Surveys between 2001 and 2013. Before the reform the average expected retirement age was 63.5 among the cohorts under examination. Note that the first post reform year in this specification is 2007, as the survey was conducted from April-August 2007, i.e. when the reform was already in place.

higher with an average expected retirement age just above 65.[20] This simple before-after comparison does not isolate the causal effect of the reform on expectations. I thus implement the DiD approach described above and cluster standard errors at the individual level, as some respondents are interviewed repeatedly in consecutive years. Crucially, the DiD approach allows to identify the causal effect of an additional one month increase in individuals' normal retirement age on their retirement expectations. Note that for the SAVE data I can only include cohort fixed effects, as the majority of individuals is not observed repeatedly.

The results are depicted in Table 3.2. Specification (1) shows the results without covariates. The point estimate for the interaction term implies that a one month increase in the NRA induced by the 2007 pension reform, on average, raised the expected retirement age by approximately 1.2 month. The result holds in specification (2) when controlling for covariates such as educational attainment and a female indicator variable.

Specifications (3) and (4) replicate the estimation for individuals in cohorts 1964-1969,

[20]Individuals state retirement expectations substantially below the statutory normal retirement age. One reason for these lower expectations are potential early retirement options, allowing individuals to retire before the NRA at age 63.

who were not differentially affected by the reform. Crucially, I re-assign D_i as if these cohorts were treated and raise their placebo retirement age by additional two months for each cohort. The results clearly show that cohorts that were not affected by the reform did not adjust their expected working horizon in response to the reform, while the actually affected cohorts adjust their expectations in line with the increase in the statutory retirement age.[21]

Finally, I check for dynamic patterns and common pre-trends by estimating Equation 3.2. Figure 3.A.1 in the Appendix shows the results. Note that due to the relatively low number of observations standard errors are relatively large and estimating such a dynamic specification is not ideal. Nonetheless, the graph allows to provide suggestive evidence on the common pre-trend assumption needed for identification. The results clearly show that before the reform there is no significant difference between cohorts' expectations. In the year of the reform, i.e. 2007, expectations are significantly higher. Similarly, there is a jump in expectations in 2011, which likely is a derivative of increased media coverage of the reform between 2010-2012 (see Section 3.2).

Employment Results If individuals adjust expectations about their working horizon, this potentially translates into their (un)employment outcomes. The following section therefore analyses the forward-looking employment responses to the 2007 pension reform using the SIAB data. Table 3.3 presents regression results for the main specification using individual fixed effects for the two main outcomes: employed at some point in a given year and contributing to social insurance, and registered as unemployed at some point in a given year.[22] Note that the employment and unemployment states are not mutually exclusive. An individual can be both employed and unemployed in a given year. Therefore the point estimates do not need to be equal in absolute value. Further, there are other outside options such as marginal employment, which will be analysed in

[21]The point estimate of 1.2 even suggests a slightly higher rise in expectations than the actual increase in the NRA stipulated by the reform. One potential explanation for this may be that affected individuals anticipate future increases in the NRA for their cohort.

[22]The inclusion of individual fixed effects ensures that only individuals observed before and after the reform are used in the estimation approach, i.e. individuals that are actively participating in the labour market. Further, they allow controlling for all time-invariant confounding factors at the individual level.

Section 3.8.

The results show a clear increase in the probability to be employed in a given year following the reform: a one month increase in the NRA raises the probability by approximately 0.18 percentage points. The point estimate is significant at the one percent level. The probability of being registered as unemployed at some point during a given year decreases by approximately 0.11 percentage points in response to the reform. The point estimate, however, is only significant at the ten percent level. Overall the reform therefore increases the probability of being employed and to a certain extent also the probability of being unemployed.

Table 3.3: Main Results

	(1) Employment	(2) Unemployment
Interaction	0.00178***	-0.00114*
	(0.00064)	(0.00068)
Bi-Cohort Year FE	✓	✓
Individual FE	✓	✓
N	1,763,754	1,763,754

Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Note that the first post reform year in this specification is 2008, as the reform was passed during the year 2007 and employment responses might take time to manifest.

To give an idea of the relative size of the effect, the average employment rate in the pre-reform period across all cohorts was 85.5 % and the unemployment rate was 18.3 %. Taking the pre-reform mean as the baseline, this translates into an increase of 0.21 percent in the probability to be employed and a 0.6 percent decrease in the probability to be unemployed during a given year for each additional monthly increase in the NRA.

On a more general level, it should be noted that the pension reform might have created general equilibrium effects on labour demand and/or wages, thereby potentially creating spillover effects affecting individuals' labour market outcomes.[23] It is therefore important to highlight that the point estimates presented here only capture net of general equilib-

[23]For example, individuals with a longer working horizon may be more likely to find a job, thereby impacting employment chances of individuals with a lower NRA and thus their post reform outcomes.

rium effects, as long as these potential spillover effects do not impact individuals with a two month difference in their statutory retirement age differently.

Figure 3.3: Outcome: Employment

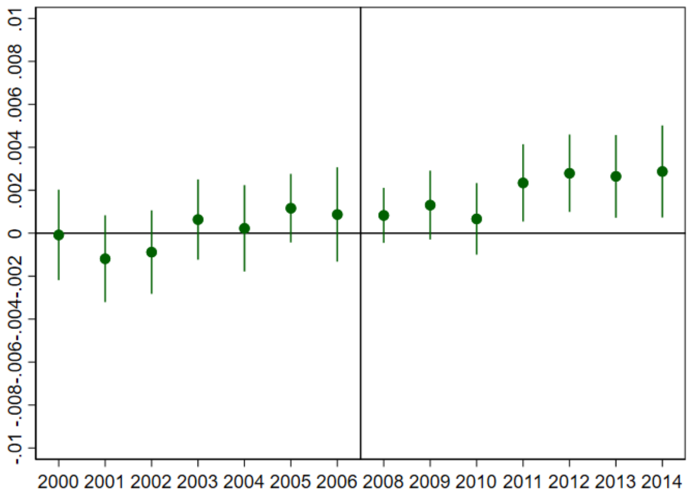
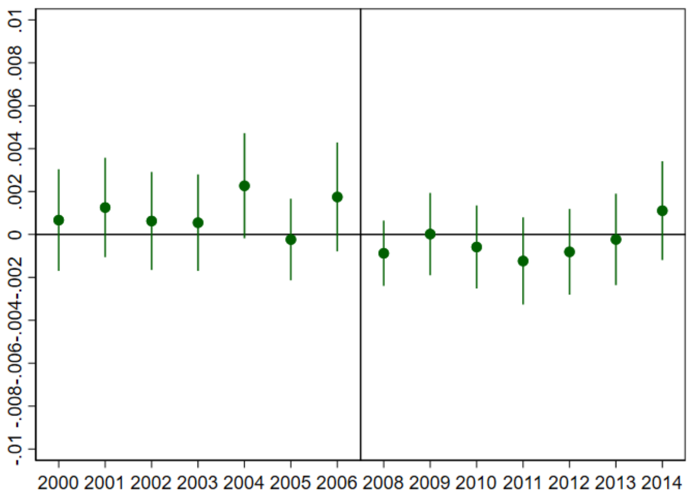


Figure 3.4: Outcome: Unemployment



Note: Figures 3.3 and 3.4 depict point estimates and corresponding 95% confidence intervals obtained from the dynamic specification described in Equation 3.2 for the outcome variables employment and unemployment throughout a given year controlling for individual fixed effects and bi-cohort times year fixed effects. The y-axis depicts the coefficient value and the x-axis the corresponding year. Note that the first post reform year in this specification is 2008, as unlike for the expectation data, observations in 2007 can stem from the pre-treatment period.

In a next step, I focus on dynamic reform responses by estimating Equation 3.2. Figures 3.3-3.4 depict the corresponding point estimates and their 95 % confidence intervals of the respective year-change in retirement age interaction.[24] For the employment out-

[24] As a baseline, I here choose the year 2007, as 2008 is the first full year under the new pension rules

come, there clearly is no difference between cohorts before the introduction of the reform, supporting the necessary common pre-trend assumption. Directly after the reform the point estimates are positive albeit not significantly different from zero. From 2011 on point estimates rise and become significant. One obvious explanation for this dynamic pattern is that between 2010-2012 there was a strong discussion about the reform, as in 2012 the first cohort affected by the reform (cohort 1947) retired at an age older than 65 years, which was strongly covered by the media and discussed by the general public (see Figure 3.2). For the unemployment outcome, there again is no significant difference between cohorts before the reform. Note, however, that the point estimate for the year 2004 is relatively large and significant at the ten percent level. After the reform, point estimates become smaller and in most cases negative, however, individually remain insignificant at any conventional level. This is unsurprising as the point estimate in the pooled specification shown in Table 3.3 is only significant at the ten percent level.

3.5.2 Robustness

A key assumption for the main approach to yield causal estimates is the common pre-trend assumption, which Figures 3.3 and 3.4 support. Crucially, there is no significant difference between cohorts before the reform period. In order to further test that the main approach is not simply picking up differential age patterns between cohorts, I implement a placebo test. As for the expectations outcome, I repeat the estimation for cohorts 1964-1969, which were not differentially affected by the reform and re-assign D_i as if these cohorts were treated, raising their placebo retirement age by an additional two month for each consecutive cohort. Table 3.A.2 in the Appendix clearly shows that there is no significant effect of this placebo treatment on any of the two outcomes, thereby supporting the notion that the main specification identifies causal reform effects.

Another concern might be that the unemployment benefit duration reform in 2006, described above, due to its age specific nature has differentially impacted the cohorts analysed in the main specification. For example, cohorts that were aged below 45 did not

and employment/unemployment responses might take time.

experience a reduction in their benefit entitlement duration throughout the years 2006 and 2007, whereas individuals aged 46 or 47 experienced a substantial reduction by more than 6 and 10 months respectively. As discussed above, Dlugosz et al. (2014) show that exclusively older individuals, i.e. individuals from 50 onward and therefore not the cohorts under examination in this paper, reacted to shorter benefit durations by selecting into unemployment just before the reform was passed and by being less likely to become unemployed directly after the reform was in place. The cohorts under examination in this paper thus in practice have not differentially reacted to the reform, as they were aged 42-47 in 2006, the year of the reform. On paper, however, cohorts 1959, 1960, 1961 and 1962 were also directly and differentially affected by the reform. In the main specification, I address this by including age group fixed effects, which ensure that only individuals who are in the same entitlement length age group in a given year are compared to one another. In an additional robustness check, I now exclusively focus on the cohorts 1963 and 1964. Individuals in these cohorts were aged 42 and 43 in 2006 when the benefit reform was passed and thus not affected by the reform. Since I only observe two cohorts for this additional estimation, I implement a classic DiD design and replace the multi-valued treatment with a dichotomous indicator if individuals were born in cohort 1964. The point estimate thus measures the impact of a two month longer statutory working horizon. Table 3.A.3 shows that the inference remains the same when only focusing on these two cohorts: individuals with a longer working horizon are more likely to be employed in a given year. For the unemployment outcome, the effect is negative albeit insignificant. Figures 3.A.2 and 3.A.3 depict the results for the dynamic specification and clearly show that there is no evidence for differential pre-trends. Further, as in the main specification the positive employment effect is driven by point estimates from 2011 onward.

Next, I adjust the reform date. In the main specification using the SIAB data, I specified 2008 as the first year of the post-reform period, as the pension reform was passed during the year 2007. In a alternative specification, I treat the year 2007 as the first year in the post reform period. Table 3.A.4 shows the results. Inference across outcomes remains the same. The employment effect is smaller in absolute size. This is intuitive as the dynamic specification clearly shows that for the employment outcomes, the point estimates only start to increase from 2011 onward.

Further, in the main specification I restrict my estimation to the years 2000-2014. In June 2014, there was another reform, which essentially reduced the early retirement age for the *particularly long-term insured*. Conditional on having contributed 45 years, individuals born before 1953 were now able to enter early retirement at age 63. For cohorts 1954-1964, this early retirement age, however, was increased step by step for each cohort, in line with the increase induced by the 2007 pension reform. Consequently, individuals born in 1964 who have accumulated 45 years of contributions would not be able to take advantage of this early retirement option until age 65.

In summary, the 2014 pension reform therefore is likely to have reinforced the effects of the 2007 pension reform. Table 3.A.5 in the Appendix repeats the main estimation this time including the years 2000-2017. As expected, the inference remains the same with slightly more pronounced employment effects.

3.5.3 Extension: The 1999 Pension Reform

In order to supplement my results for the 2007 pension reform, I use the 1999 pension reform as a second source of quasi-random variation. Gohl et al. (2021) also use the 1999 pension reform and mainly focus on forward-looking human capital effects but also provide some albeit not very strong evidence on forward-looking employment effects for a sample of older women starting in 2005 using German Microcensus data (see Chapter 2). In contrast to Chapter 2, the additional analysis in this chapter focuses on younger women and earlier observation years.

The 1999 pension reform effectively increased the ERA for women born after 1951 by three years from age 60 to age 63. I therefore exclusively focus on women in the cohorts 1951 and 1952 and employ a regression discontinuity design focusing on observations from 2000 to 2008 and thus on individuals aged 48 to 58.[25]

[25]As described previously, data before 2000 is prone to recording errors. Women after the age of 58 are explicitly excluded, as on paper they can bridge the time into retirement by claiming unemployment benefits, which potentially may lead to mechanical and not behavioural employment responses.

Equation 3.3 details the approach:

$$y_{it} = \alpha_0 + \gamma_1 D_i + \gamma_2(B_i - c) + \gamma_3 D_i(B_i - c) + X_i'\delta + \alpha_a + \phi_t + \eta_f + \varepsilon_{it} \tag{3.3}$$

y_{it}, depending on the specification, stands for the same outcome variables as before. D_i is a dummy specifying treatment, that is equal to one if a woman is born after 1.1.1952 and 0 otherwise. Formally, D_i is given by:

$$D_i = \begin{cases} 1, & if\ B_i \geq c \\ 0, & if\ B_i < c \end{cases} \tag{3.4}$$

where B_i is a woman's month of birth and c is the cut-off date for the increase in early retirement age, ERA (January 1, 1952). In the main specification, one cohort is included on each side of the cut-off. More precisely, I include the cohort 1951 below the cut-off and the cohort 1952 above the cut-off. The difference between a woman's birth month and the cut-off, $B_i - c$, gives the running variable. The running variable is interacted with the treatment variable D_i to allow for different slopes before and after the cut-off. Depending on the specification and to account for potential different functional forms, I include a linear or quadratic specification of the running variable and the interaction term. Further, I account for year fixed effects, ϕ_t, age group fixed effects, α_a, federal state fixed effects, η_f and observable characteristics, X_{it}. In particular, I include educational attainment and the number of days in employment at the beginning of the observation period in 2000 as a measure of total work experience. Last, ε_{it} is the error term. Following Geyer and Welteke (2021), I cluster standard errors at the birth month level to account for correlation between observations for the same individual or individuals born in the same month.

Table 3.4 shows the results for the RDD approach using local linear and quadratic regressions and Figures 3.B.1 and 3.B.2 in Appendix B show the typical RDD graphs plotting the running variable on the x-axis and the corresponding averages of outcome variables on the y-axis, fitting local regressions using a triangular kernel. For the employment

Table 3.4: RDD Results

	(1) Employment	(2) Unemployment
Panel A: Linear Running Variable		
Treatment	0.0188***	-0.0089*
	(0.0067)	(0.0046)
Panel B: Quadratic Running Variable		
Treatment	0.0264***	-0.0167***
	(0.0072)	(0.0044)
Covariates	Yes	Yes
Year FE	Yes	Yes
Federal State FE	Yes	Yes
N	136,199	136,199

Note: Source: SIAB data, own calculations. Standard errors in parentheses are clustered at birth month level. Results obtained from local regressions. Panel A shows the results for a local linear and Panel B for local quadratic specification; significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

outcome, depending on the specification, point estimates range from just under 1.9 percentage points to more than 2.6 percentage points and are significant at the one percent level across specifications. For the unemployment outcome, I find negative effects ranging from approximately 0.9 percentage points to 1.7 percentage points, which are significant at the ten and five percent level respectively. Overall, the results are in line with the findings for the 2007 pension reform and support the notion that pension reforms imply positive forward-looking employment effects. Unsurprisingly, effect sizes are larger than for the 2007 reform, as this reform implied a larger shift of the retirement age by, i.e., by three years.

The above approach isolates the local treatment effect conditional on the assumption that observed and unobserved characteristics around the cut-off are not systematically different between the treated and the control group. In other words, individuals close to the cut-off would have behaved identically in the absence of treatment conditional on the included fixed effects and covariates.

In Appendix B, I provide supporting evidence for this assumption based on balancing tests of covariates as well as by moving the cut-off to a hypothetical placebo date. Table

3.B.2 confirms that pre-determined observable covariates such as educational attainement do not jump around the cut-off. Additionally, Table 3.B.1 shows that cohorts 1952 and 1953, which were not differentially affected by the reform, do not show differences in (un)employment outcomes, supporting the notion that, in the absence of the reform, adjacent cohorts behave similarly. Moreover, there were no other relevant policy reforms between 2000 and early 2007 that affected women in the 1951 and 1952 cohorts differently. As described above, the 2007 pension reform shifted the NRA for the 1952 cohort by an additional month compared with the 1951 cohort, so for the last two years the estimate also captures this additional reform effect.

In summary, both the 2007 and the 1999 pension reform indeed increased the employment probability for individuals with a longer expected working horizon. For the remainder of the paper, I focus on the 2007 pension reform, as additional data needed for further analyses such as the IZA Evaluation Panel is only available for the cohorts affected by this reform. Further, the larger number of observations when using the SIAB data for this reform allows the implementation of an extensive set of heterogeneity analyses. The following sections will first focus on the mechanisms driving the documented results, to then analyse potential reform effects on other outcome variables as well as heterogeneous results.

3.6 Mechanisms

The documented employment effects of the 2007 reform in theory can be driven by a range of combinations of inflows and outflows from one state to another. For example, individuals with a longer expected working horizon might be more likely to remain employed, and thus might be less likely to enter unemployment. Similarly, unemployed individuals with a higher NRA might be more likely to find re-employment.

In order to approximate inflow and outflow patterns from one state to another, I estimate year-to-year transitions using the main DiD specification. More precisely, I repeat the same estimations as before but condition on an individuals' observed labour market state in year t and consider outcome variables that are indicators equal to one if in the following year, i.e. $t + 1$, the observed individual is in one of three possible outcomes states.[26] In addition to the employment and unemployment outcomes used above, I include a third outcome category that captures all individuals that are neither employed nor unemployed throughout a year but are observed in the data set. For example, these are individuals that participate in active labour market programs, such as unemployment training. For each outcome state, I thus condition on being employed, unemployed or in the third category, i.e. *Other*, at time t. Table 3.5 therefore mimicks a transition matrix for the reform effect where the rows show the original state and the columns the state in the next period. Naturally, these transition rates solely serve a descriptive purpose, as the states I condition on are outcomes by themselves. However, they provide suggestive evidence on the drivers of the documented employment and unemployment effects.[27]

Table 3.5: Transition Results

Rows: State t / Columns: State t+1	(1) Employment	(2) Unemployment	(3) Other
Employment	0.00059*	-0.00025	-0.00022
	(0.00016)	(0.00063)	(0.00017)
Unemployment	0.00467*	-0.00201	-0.00018
	(0.00247)	(0.00170)	(0.00083)
Other	0.00498*	0.00215	-0.00660**
	(0.00302)	(0.00168)	(0.00332)

Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Observations in first row: 1,393,145; second row: 272,087; third row: 86,453. The table depicts the point estimates from regressions using bi-cohort year and individual fixed effects and conditioning on employment, unemployment or all remaining possible states observed in time period t (rows) using the observed states in period t+1 as outcomes (columns). Point estimates do not need to cancel out along rows, as the observed states are not mutually exclusive, i.e. individuals can be employed and/or unemployed and/or in the *Other* category in a given year.

[26] Again, employment and unemployment are not mutually exclusive, as throughout a year individuals can be both employed and unemployed. Solely focusing on individuals who are either employed or unemployed in a year would imply that a large part in the variation in employment/unemployment flows would not be captured by the estimation, as individuals who enter either of the states do so throughout a given year.

[27] Also note that the sum of point estimates does not need to be equal to zero, as the observed states are not mutually exclusive, i.e. individuals can be employed as well as unemployed in a given year.

The results in the first row of Table 3.5 show that employed individuals with a higher increase in the NRA are more likely to remain in employment in the following year. The point estimate is significant at the ten percent level. Further, the results indicate that employed individuals with a stronger increase in their NRA are also less likely to be unemployed or in the *Other* category in the following year. Note, however, that both point estimates are insignificant at any conventional level. The results in the middle row point towards a significantly higher probability (10 % level) to find re-employment for unemployed individuals with a longer working horizon. Further, unemployed individuals with a higher NRA are less likely to remain in unemployment and to enter active labour market programs captured by the *Other* category. The point estimates for these latter two outcomes, however, are not significantly different from zero. Last, individuals with a longer NRA in other recorded categories such as active labour market programs are less likely to be in the same category in the following year (significant at the 5% level) and more likely to be employed in the following year with a positive point estimate significant at the ten percent level. The point estimate for the unemployment outcome is positive albeit not significantly different from zero.

In summary, Table 3.5 provides suggestive evidence that employed individuals with a reform-induced longer working horizon are more likely to remain in employment. Further, unemployed individuals and individuals in active labour market or other programs are more likely to find re-employment. One possible explanation for the observed higher probability of staying in the labour force can be found in the empirical literature documenting positive effects of raising the retirement age on investment in on-the-job training, which could have an impact on employability and job security, see e.g. (Gohl et al., 2021; Brunello and Comi, 2015). Similarly, a greater inflow into employment from unemployment or active labor market programs may be explained by firms' possible preference to hire individuals with longer legal employment horizons and/or by an adjustment in individual behaviour. Hairault et al. (2010), for example, suggest an increase in job search intensity by the unemployed in response to a longer remaining working life as a key mechanism.

In order to analyse whether there is an adjustment in individual behaviour and thus a change in the labour supply response, I use the IZA Evaluation data set, which focuses on unemployment inflows in a time frame immediately after the reform, i.e. from June 2007-May 2008 (see Section 3.2). Crucially, the data set includes information about job search behaviour at the beginning of the observed unemployment spell and also allows to observe whether individuals found re-employment within three years of first entering unemployment. In order to measure job search intensity, I focus on the number of applications and job offers received within the first two months of unemployment. More precisely, I construct dichotomous variables that are equal to one if an individual send out or received more than the average number of job applications or offers.

In total, I observe more than 2,500 individuals in the cohorts 1959-1964 who entered unemployment. The data does not include observations from the pre-reform period and the DiD inspired research design used so far is no longer feasible. I therefore implement a design similar to Frimmel (2021). Note that it can not account for possible selection into unemployment, which according to the main results above does exist to a certain extent. The results are therefore of suggestive nature, aiming to show whether individuals who become unemployed and have a longer statutory working horizon, behave differently. The approach is fully detailed by Equation 3.5.

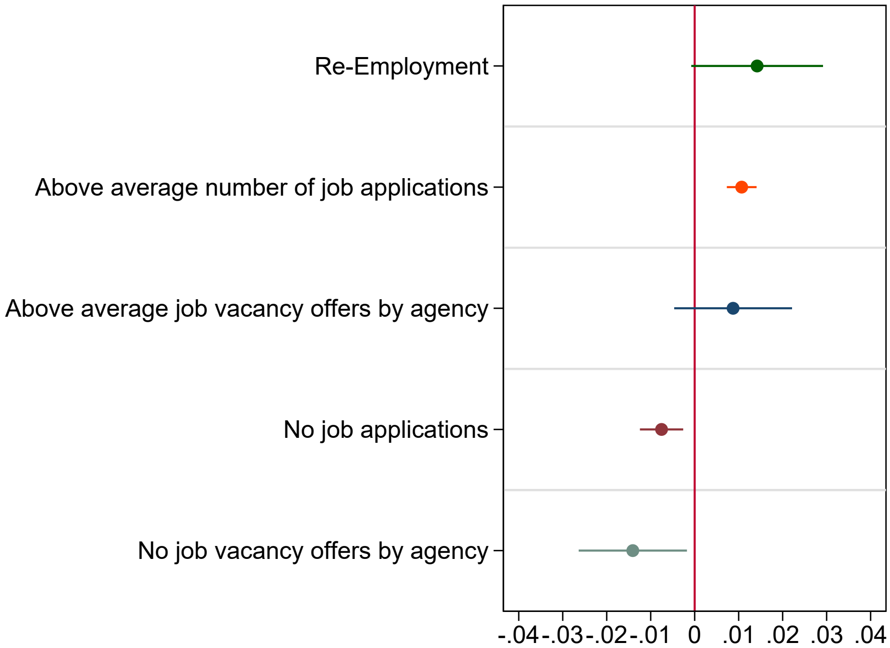
$$y_i = \beta_0 + \beta_1 D_i + \zeta_{bc} + \alpha_a + X'_{it} + \epsilon_{it} \tag{3.5}$$

The approach uses the variation in the statutory retirement age between cohorts. y_{it} is the outcome variable of interest (see above). β_0 is an intercept term and D_i is, as in the main approach, the change in the retirement age induced by the reform. ζ_{bc} are bi-cohort fixed effects, i.e. indicator variables equal to one for cohorts 1959 and 1960, then for cohorts 1961 and 1962 and so on. α_a are age group effects accounting for differences in the entitlement length to unemployment benefits. Both the bi-cohort fixed effects and the age group fixed effects ensure that first I only compare individuals who are one year apart in age and second individuals who share the same rule framework in terms of unemployment benefits. X'_i is a vector controlling for individual characteristics such as gender, educational attainment, the net income received from the last known employment, an indicator whether the individual receives unemployment benefits, a year indicator for the year 2008 and regional control variables such as the local unemployment and vacancy rates. Last, ϵ_{it} is an error term clustered at the cohort level.

Figure 3.5 depicts the point estimates and their corresponding 95% intervals for the coefficient of the change in the retirement variable, thus showing how different outcome variables react to a one month increase in the retirement age. First, the results show a higher probability to be re-employed within three years after individuals first entered the sample. The point estimate is significant at the ten percent level. Second, individuals with a longer working horizon are more likely to have sent out more than 14 applications during the first two months of unemployment, i.e. the average number of sent out applications, and less likely to have sent out no job application at all. Similarly, they are more likely to have received an above average number of job offers by the job agency, i.e. more than one job offer. Note, however, that the corresponding point estimate is not significant at any conventional level. Further, they are less likely to have received no offer at all.

All in all, then, these suggestive results do indeed point to an increase in search intensity as advocated by Hairault et al. (2010). Moreover, the results indicate that individuals with a longer working horizon are actively considered by job agencies and their officials. The next sections will continue exploring possible mechanisms and additional responses to the reform by analysing further outcomes and providing heterogeneous results for the main specification.

Figure 3.5: Job Search Outcomes, IZA Evaluation Survey



Note: Figure 3.5 depicts the point estimates and their corresponding 95 % confidence intervals obtained from regressions of the respective outcome variable displayed on the left hand side of the Figure on a variable measuring the change in the statutory retirement age induced by the 2007 reform for an unemployment inflow sample from June 2007-May 2008 of 2,554 individuals born in the cohorts 1959-1964. All regressions control for the covariates such as gender, educational attainment, the net income received from the last known employment, an indicator whether the individual receives unemployment benefits, a year indicator for the year 2008 and regional control variables such as the local unemployment and vacancy rates as well as bi-cohort and age group fixed effects.

3.7 Further Outcomes and Job Sorting

In this section, I analyse additional outcome variables that may also be affected by the shift in the work horizon. First, I focus on outcomes related to employment or unemployment responses such as employment adjustments on the intensive margin or yearly earnings. Second, I analyse whether individuals sort into particular occupations in response to the reform.

Additional Outcomes Table 3.6 depicts the results for a range of different outcome variables using the preferred DiD specification on the SIAB data. There are a range of other outcome variables that may also be impacted by the reform. For example, the paper so far has considered employment responses at the extensive margin. However, individuals might also adjust their behaviour on the intensive margin by working more hours. First, in column (1) I focus on a part time indicator that is equal to one if a person is employed in part time and zero otherwise. The point estimate is small, positive and insignificant, indicating that there is no overall effect on part-time work. Unfortunately, the SIAB data does not provide information on the actual number of hours worked. Therefore a more precise analysis of responses along the intensive margin focusing on hours worked is not possible. Second, I use marginal employment as an outcome variable. Crucially, marginal employment does not count towards pension insurance. Individuals with a longer work horizon might therefore be less likely to be marginally employed and more likely to switch to employment subject to pension insurance contributions. This indeed seems to be the case, as the corresponding point estimate is small, negative and significant at the ten percent level (column (2)).

In column (3), I look at individuals' yearly labour earnings. In particular, the documented employment effects of the reform may also translate into different yearly earnings. Indeed, I find a positive effect of more than 137 Euros per year, which is significant at the one percent level. This effect corresponds to a 0.55 percent relative increase in yearly earnings using average pre-reform earnings as the baseline and is likely partly driven by the documented employment effects. Further, Gohl et al. (2021) and Gohl et al.

Table 3.6: Other Outcomes

	(1) Part time	(2) Marginal Emp.	(3) Yearly Earnings	(4) Unemp. Training	(5) Days in Unem.	(6) Self-Emp	(7) Inactivity
Interaction	0.00030	-0.00098*	137.44374***	-0.00071**	-0.32385*	-0.00002	-0.00144***
	(0.00081)	(0.00051)	(24.12624)	(0.00033)	(0.18925)	(0.00002)	(0.00033)
Bi-Cohort Year FE	✓	✓	✓	✓	✓	✓	✓
Individ. FE	✓	✓	✓	✓	✓	✓	✓
N	1,763,754	1,763,754	1,763,754	1,763,754	1,763,754	1,763,754	1,763,754

Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table repeats the main specification including individual and bi-cohort year fixed effects using a range of different outcomes depicted in the first row of the table.

(2023) show that individuals with a longer working horizon invest more in human capital accumulation and are more likely to be promoted, which may additionally translate into increases in earnings.[28]

Next, in columns (4)-(6) I examine outcomes related to unemployment. First, I use participation in an *unemployment training program* as an outcome variable. The data used in this study allows to identify participation in *measures of vocational training*. These measures are by law required to be linked to a specific vocational goal. For example, this goal can consist of computer and software courses to maintain and update existing vocational skills, but can also consist of retraining for a new profession and/or training for an additional vocational profession or exam. The results show that individuals with a longer working horizon are less likely to participate in these training measures, which is likely driven by lower inflows into unemployment of those individuals. The same holds for days spend in unemployment per year, which shows a negative point estimate that is significant at the ten percent level.[29] Further, the data allows to observe whether individuals obtain a subsidy for setting up their own business and thus enter self-employment. The point estimate is very close to zero and insignificant. This is not a perfect measure of self-employment, as it just records subsidized self-employment, however, might serve as a suggestive proxy.[30]

[28]When conditioning on employment and repeating the estimation the inference remains the same: yearly earnings are higher for individuals with a longer working horizon. However, the size of the point estimate decreases to approximately 120.

[29]For any of the outcomes related to unemployment, I can repeat estimations conditioning on unemployment. Point estimates for the training outcome and days spend in unemployment remain negative, however, are statistically not different from zero.

[30]When conditioning on unemployment the point estimate for the self-employment outcome remains

Last, exits into inactivity may play a role. Since inactivity is not directly observed in the data, I focus on an inactivity proxy. More precisely, I use an indicator variable that is equal to one if an individual was previously observed in the data but no longer is in the current year. This proxy measure thus does not only include individuals entering inactivity but also other outcome states not captured by the SIAB data, such as civil service or unsubsidized self-employment. The results show that there is a decrease in the probability to enter inactivity, which is significant at the one percent level. This is in line with the main results, showing that a longer legal work horizon increases the probability to be employed while reducing the probability to be in unemployment as well a inactivity.

Sorting into Occupations Another possible response to the reform may be to sort into jobs that can be exercised at older ages and thus enable individuals to remain in employment for a longer period. Jobs with a lower level of psycho-social or physical stress are likely to fulfill this criterion. Differences in physical intensity and stress levels associated with a job indeed strongly affect the likelihood of being employed at older ages (Zwick et al., 2022; Vermeer et al., 2016), as well as actual subjective expectations of being able to work until the official retirement age, and could therefore also play a role in the response to a postponement of the statutory retirement age. To illustrate this, in Figure 3.A.4 in the Appendix I additionally provide descriptive evidence obtained from a survey commissioned by the DGB (Deutsche Gewerkschaftsbund), one of Germany's largest trade union umbrella organisation.[31] Crucially, the survey elicits subjective beliefs about individuals' capacity to work until the statutory retirement age. Individuals were asked whether they expect to work until the statutory retirement age without any restrictions given the requirements of their current job. On average 47 % percent of all individuals answered with yes. Figure 3.A.4 depicts answers to the same question stratified by different physical and psycho-social stress indicators. The results show that for jobs with a high level of physically demanding tasks the percentage of people answering with yes (23 %) is substantially below the overall average rate of 47 %.

negative insignificant.

[31]The survey includes 31,164 workers (members and non-members in the union) across all sectors and regions in Germany between 2012-2016. The summary statistics are taken from Gewerkschaftsbund (2016) as direct access to the data is not possible

In order to explicitly analyse whether job types and the corresponding stress exposure of a given job play a role in the response to the pension reform, I use the *classification of occupations 2010* (*Klassifikation der Berufe 2010*) provided for each observation in the SIAB data.[32] As a first step, using the 2010 occupational classifications allows to differentiate between nine so called occupational areas. Table 3.A.6 in the Appendix lists these areas, ranging from occupation areas in the agricultural and manufacturing sector to occupations within the service sector. In order to asses whether individuals sort into specific occupation areas in response to the reform, I regress indicator variables, which are equal to one if an individual newly entered an occupation area, on the main DiD specification. Table 3.A.7 in the Appendix clearly shows that there are no sorting patterns into any of the observed occupation areas.[33]

As a second step, in order to more explicitly analyse the role of stress exposure both physical and psycho-social, I use exposure indices developed by Kroll (2011) and recently used by Zwick et al. (2022); Mazzonna and Peracchi (2017). Kroll (2011) defines exposure as "conditions with potential physiological and/or psychological effects on the human organism resulting from the characteristics of the activity itself or from its external conditions". Crucially, the indices are available for 138 occupation groups following the *classification of occupations 2010* and therefore allow to study sorting patterns at a more granular level than the nine occupation areas studied above.[34] In total, Kroll (2011) provides four exposure indices with values ranging from zero to ten: (1) a composite overall index of exposure to environmental, temporal, carcinogenic, ergonomic, physical and psycho-social stress indicators and (2) separate indices for physical or psycho-social exposure as

[32]For a detailed description of the classification procedure see Paulus and Matthes (2013). The classification system is based on employers reporting a certain job key when registering the employee. Before the switch to the current system in 2011, there was a different registration key. For employees registered before 2011 the old key is transformed into the new classification system.

[33]Note that for the agricultural occupation area the point estimate of the interaction term is negative and significant at the ten percent level. For all other occupation outcomes point estimates are small and insignificant.

[34]The original indices are calculated at the five and four digit level of the classification system. Since in the SIAB data I only observed classifications based on three digits, I take average values of the index values for 138 three digit occupation groups. The indices are based on responses from a large-scale representative survey with approximately 20,000 participants conducted by the German Federal Institute for Vocational Education and Training (BIBB). The original paper uses the German BIBB/ BAuA-workforce survey 2006 and was later updated for the German BIBB/ BAuA-workforce survey 2011. The index for different job classifications can be retrieved Gesis Archiv.

well as a heavy work index.[35] For each year I merge the respective index value to each observation using an individual's current jobs' three digit occupation classification.[36] I then construct three different outcome variables for each indicator: (1) the immediate index value ranging from 0-10, (2) a dichotomous outcome variable that is equal to one for an above average index value and zero otherwise, and (3) a dichotomous outcome variable that is equal to one if the given index and hence an individual's exposure decreased in comparison to the previous year. Figure 3.A.5 in the Appendix depicts the results. For example, graph [c] shows the results for the last outcome variable, i.e. the indicator variable for a decrease in the index. There is no evidence for sorting into jobs with a lower stress exposure, as point estimates remain insignificant across outcomes and index type. This holds when changing the definition of the outcome variable using the immediate index value as outcome or an indicator variable that is equal to one for an above average index value. For each specification, point estimates remain insignificant.

Overall, there is no evidence for occupational sorting patterns in response to the reform, implying that individuals do not retrain and switch into different occupational areas that may be easier to exercise at an older age. This potentially is explained by the age of the observed individuals who are 44 and older in the post-reform period and thus likely to be relatively settled in their career path and occupational area.

3.8 Heterogenous Results

In this section, I explore whether the effects of the 2007 pension reform vary along a range of dimensions to test for heterogeneous employment and unemployment responses to the reform. In particular, after studying heterogeneous responses along sex, region and education, I revisit the role of occupation types and exposure, testing for different

[35]For example, variables used to construct the physical index ask for the frequency of fulfilling tasks standing or sitting and how often individuals need to lift heavy. For the psycho-social index individuals are, amongst others, asked about time pressure and the general work atmosphere. In total, the indices are constructed using 39 survey items.

[36]In the case of individuals who were not employed in a given year, I use their last observed place of work.

reform effects across occupation areas and jobs' stress exposure. Similar to the argument made above, jobs in certain occupational areas may be more suitable to exercise at an older age, and thus individuals employed in or with experience in certain professions may be more likely to adjust their employment behaviour in response to pension reforms.

Heterogeneous Results: By Sex First, I split the sample into men and women. Table 3.7 shows the results. Specifications (1)-(3) exclusively focus on women and (4)-(6) on men. Overall, the results show that the (un)employment effects are strongly driven by women. This finding is in line with recent results by Carta and de Philippis (2021). One potential explanation for the difference between men and women may be the widely accepted notion that labour supply is less elastic for men. In particular, men tend to be more attached to the labour market and therefore may have less room to increase their labour supply in response to the reform (Carta and de Philippis, 2021).

Table 3.7: Heterogeneity: Women vs. Men

	(1) Employment	(2) Unemployment	(3) Employment	(4) Unemployment
Interaction	0.00333***	-0.00190*	0.00049	-0.00051
	(0.00107)	(0.00101)	(0.00076)	(0.00091)
Bi-Cohort Year FE	✓	✓	✓	✓
Individ. FE	✓	✓	✓	✓
Group	Women	Women	Men	Men
N	834,666	834,666	929,088	929,088

Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns (1)-(2) show the point estimates obtained from estimations including bi-cohort year fixed effects and individual fixed effects for women. Columns (3)-(4) show the results for men.

Heterogeneous Results: East vs. West Another possible dimension for heterogeneity are different regional responses. In particular, there may be differences between former East and West Germany. For example, labour market attachment of women in former East Germany tends to be higher, as historically East Germany had very high female labour force participation rates (Bonin, 2005). Table 3.A.8 in the Appendix thus splits the sample along this regional dimension.[37] The results show that the estimated point estimate for the employment outcome is slightly larger and remains significant for a regression focusing solely on West Germany. For East Germany, the point estimate is slightly smaller and insignificant, which might be due to the substantially lower number of observations and thus a lack of statistical power for East Germany. Alternatively, higher labour market attachment among women in East Germany could explain the slightly lower effects documented for the East. However, note that for the unemployment outcome both point estimates are negative and insignificant with the estimate for East Germany larger in absolute size.[38]

Heterogeneous Results: By Education I then look at heterogeneous results by education and split the sample into individuals with no vocational training, vocational training and a college degree. Table 3.A.9 in the Appendix shows the results. Point estimate for the employment outcome are relatively large and significant for individuals with a college degree. For individuals with some form of vocational training, they are less pronounced and only significant at the ten percent level. Point estimates for unemployment outcome are insignificant for both groups. For individuals without any form of vocational qualification, the effects are insignificant, albeit in case of the employment outcome positive and sizeable. Overall, there is no clear heterogeneous pattern across education groups.

Heterogeneous Results: By Occupation Area In a next step, I analyse heterogeneity across occupation areas. Crucially, while the above analysis of job sorting mechanisms (see Section 3.7) has shown that there is no evidence for individuals selecting into specific occupation areas or jobs that can be exercised at an older age, there may be differential reform effects between individuals whose last recorded job was in a specific occupational

[37]Crucially, there is not sorting along this regional dimension. A regression of an east indicator variable on the main DiD specification renders small point estimates that are not significantly different from zeo.

[38]I also test for potential regional sorting patterns using an East German indicator variable as the outcome. There is no evidence for significant regional sorting patterns, thus the results are not displayed here but can be included upon request.

areas. For example, the effects may be larger for individuals with experience working in occupational areas associated with the service sector, since service-oriented jobs may be easier to exercise at an older age.

As described above, the 2010 occupational classifications allow to differentiate between nine so called occupational areas.[39] In order to study heterogeneous responses to the reform, I split the sample along these groups and thus focus solely on individuals whose last recorded job was within a given occupation area.[40] Table 3.A.10 to 3.A.11 in the Appendix depict the results for each respective outcome variable, i.e. employment and unemployment. Point estimates for the employment outcome are negative and insignificant for the first two areas, i.e. the agricultural and production/manufacturing sector. For the areas *Construction, architecture, surveying and building technology* (3), *Science, geography and computer science* (4) and *Traffic, logistics, protection and security* (5) the point estimates are relatively small and insignificant. For all occupation areas in column (6)-(8) point estimates are positive and either significant or not far off from significance. These occupational areas are traditionally associated with parts of the service sector and include *(6) Commercial services, goods trading, sales, hotel and tourism*, *(7) Business organization, accounting, law and administration* and *(8) Health, social issues, teaching and education.* For category (9) *Linguistics, literature, humanities, social and economic sciences, media, art, culture and design* the point estimate again is small and insignificant. For the unemployment outcome (see Table 3.A.11) the pattern is similar: for occupational areas (6)-(8) point estimates are negative and in case of the latter two significant. Overall, the documented employment effects are thus strongest for subgroups of occupations in the service sector.

The differences in occupational areas could be due to a number of explanations. In particular, individuals' and employers' perceptions of their ability to work in a particular occupation at a particular age are likely to play an important role. Jobs in certain occupational areas or with certain skill requirements might be better suited for employment

[39] Table 3.A.6 in the Appendix lists these areas.

[40] Crucially, as described above Table 3.A.7 in the Appendix shows that there are no sorting patterns into the occupation areas, which allows to study heterogeneous reform responses for each occupation area by splitting the sample along these categories.

at older ages, which in turn may influence the employment response to pension reform. For example, if individuals in certain occupations do not expect to work until the statutory retirement age anyway, they may not have an incentive to respond to pension reform. This may be the case in particular for jobs with a high degree of physical and psycho-social stress (see Figure 3.A.4), which is analyzed in more detail below.

Physical and Psycho-Social Stress Exposure In order to analyse to what extent differences in stress exposure play a role in the documented forward-looking employment effects of pension reforms, I again use the stress exposure indices provided in Kroll (2011) and split the sample into individuals whose last recorded occupation has an index value above or below/equal to the mean occupation index value, thereby measuring whether the documented reform responses are more pronounced in jobs with a low stress intensity.

The results are depicted in Table 3.8. Columns (1)-(2) present the results for individuals in a job with an above mean exposure and columns (3)-(4) with an exposure equal or below the mean. Panel A focuses on the overall exposure of a given job, measured by the *Overall Job Index (OJI)*, and shows that (un)employment responses to the reform are more pronounced for individuals whose last recorded job had a relatively low exposure. Panel B and Panel C specifically focus on the overall physical exposure (*OPI*) and the heavy work exposure *(HWI)* of a job. As for the general index, the results are driven by jobs with a below mean index value, i.e. jobs with a low physical intensity. Last, Panel C splits the sample along a *Psycho-Social Exposure Index (OJI).* As for the other indices the effects are more pronounced and significant in columns (3)-(4), i.e. for individuals in jobs with a below mean value. However, point estimates in both groups are relatively similar in size. In summary, particularly physical stress in a specific occupation influences individuals' responses to the reform.

Last, in order to analyse whether these heterogeneous results again are exclusively driven by women, I also stratify the sample along sex as well as above and below/equal index values for each respective index. Table 3.A.12 depicts the results for women and Table 3.A.13 for men. For women the reform effects remain more pronounced for occupations

Table 3.8: Results by Exposure Indices

	(1) Employment	(2) Unemployment	(3) Employment	(4) Unemployment
Panel A: Overall Exposure Index (OJI)				
Interaction	0.00114	0.00040	0.00215***	-0.00142**
	(0.00084)	(0.00094)	(0.00066)	(0.00069)
N	850,829	850,829	711,316	711,316
Panel B: Physical Exposure Index (OPI)				
Interaction	0.00116	0.00046	0.00232***	-0.00158**
	(0.00081)	(0.00090)	(0.00068)	(0.00071)
N	897,608	897,608	664,749	664,749
Panel C: Heavy Work Index (HWI)				
Interaction	0.00127	0.00020	0.00184***	-0.00124*
	(0.00084)	(0.00095)	(0.00065)	(0.00068)
N	838,080	838,080	723,592	723,592
Panel D: Psycho-Social Exposure Index (OSI)				
Interaction	0.00146	0.00094	0.00163***	-0.00117*
	(0.00093)	(0.00098)	(0.00060)	(0.00070)
N	729,020	729,020	830,521	830,521
Bi-Cohort Year FE	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓
Group	Above Mean	Above Mean	Below/Equal Mean	Below/Equal Mean
Observed Years	2000-2014	2000-2014	2000-2014	2000-2014

Note: Source: SIAB data, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table depicts regression results for the two main outcome variables employment and unemployment throughout a given year for individuals in occupations with an above or below/equal mean value of the respective stress exposure index denoted in each panel.

with a relatively low stress physical as well as psycho-social stress exposure. For men (see Table 3.A.13) there seems to be a small positive pension reform effect on employment for jobs with a below/equal average overall exposure and a below/equal average physical intensity, as the point estimate for this group increases to approximately 0.12 percentage points in Panel A and Panel B and is significant at the ten percent level.

Taken together, the results suggest that irrespective of sex, stress exposure plays a key role in determining to which extent individuals can react to a rise in their statutory retirement age. For women this seems to apply to both physical and psycho-social stress exposure. In contrast, mens' employment response to the reform are small and only positive and significant in jobs with a below/equal average physical stress exposure.

3.9 Discussion and Interpretation of the Results

All in all, the documented results of both pension reforms show that raising the normal and early retirement ages have positive forward-looking employment effects for individuals before they retire. They thus support the notion that increasing the retirement age provides a valid option to relieve pay-as-you go pension systems by reducing the number of unemployment benefit recipients and increasing the number of tax payers even before reaching the retirement age, therefore ultimately reducing the overall fiscal burden. However, the above heterogeneity analyses show that these results are not uniform but different across occupation areas, types and sex. Crucially, they are strongest for individuals in jobs that have a relatively low level of exposure to physical and psycho-social stress and in the service sector. These findings are in line with previous research highlighting differences in old-age unemployment and retirement behaviour. Blekesaune and Solem (2005), for example, show that workers in physically-demanding jobs are more likely to retire early and may also replace early retirement with unemployment or inactivity bridging the time until retirement (Chirikos and Nestel, 1991), which in turn may lead to financial losses. In addition, the findings of this paper suggest that there are differences along the same lines when analysing forward-looking responses to pension reforms, which may exacerbate inequalities amongst older individuals approaching retirement.

On a more general level, this paper has important implications for the policy debate on pension reforms and old-age employment. The findings show how pension policies, job search and active labour market policies are interrelated. A holistic view of these aspects is key to promoting employment amongst middle-aged and older workers, while simultaneously lifting pressure from pay-as-you go pension systems. In particular, the results documented in this paper highlight that pension reforms affect individuals and occupations before they reach official old and new retirement thresholds. Further, individuals' employment outcomes adjust differently to increases in the retirement age - a result policymakers may consider when promoting old age employment through pension reforms or labour market policies.

3.10 Conclusion

This paper provides novel causal evidence for the existence of forward-looking effects of pension reforms on employment outcomes, thereby confirming key predictions of human capital theory and job search models. Specifically, the paper, using two pension reforms from Germany, shows that a longer statutory working horizon increases the expected retirement age, the probability to be employed in a given year and decreases the probability to be unemployed in a given year. It shows that the documented higher employment probability is driven by an increase in employment inflows as well as a decrease in employment outflows. Further, it provides suggestive evidence that the effects are partially driven by a change in individuals' job search behaviour: unemployed individuals with a longer working horizon on average send out more job applications during the first months of unemployment. Last, as a key novel finding the paper shows that the documented employment effects are driven by individuals in non-physical jobs in the service sector. Further, it corroborates existing research that forward-looking employment responses to an exogenous increase in the retirement age are predominantly driven by women.

The findings have important implications for the academic debate on forward-looking and direct effects of pension reforms. First, they confirm recent empirical findings and theoretical considerations that pension reforms indeed imply forward-looking effects. The main pension reform studied in this paper further shows that even relatively small differences in the increase in the statutory retirement age are perceived by individuals and lead to different (un)employment behaviour. Second, the paper provides important insights on the mechanisms behind these forward-looking effects and shows that there are heterogeneities across occupational fields and the stress exposure of jobs. Crucially, the paper advances physical intensity of jobs as an explanation for the observed heterogeneous patterns, as physically demanding jobs may be harder to exercise at older ages.

A limitation of the paper is that the suggestive results documenting changes in job search intensity cannot be interpreted causally. Future research should focus on further investigating these behavioural adjustments among the unemployed in causal empirical settings. Moreover, due to the timing of the 2007 pension reform, this study can only examine forward-looking reform effects for individuals aged 44 and older. Analysing possible forward-looking effects of pension reforms for individuals under the age of forty could be an interesting starting point for future research. In particular, different mechanisms and heterogeneities might be at play for younger individuals than documented in this study. For example, the effects of reforms might differ between future-oriented sectors and dying sectors, and sorting into these specific sectors might matter.

Appendices

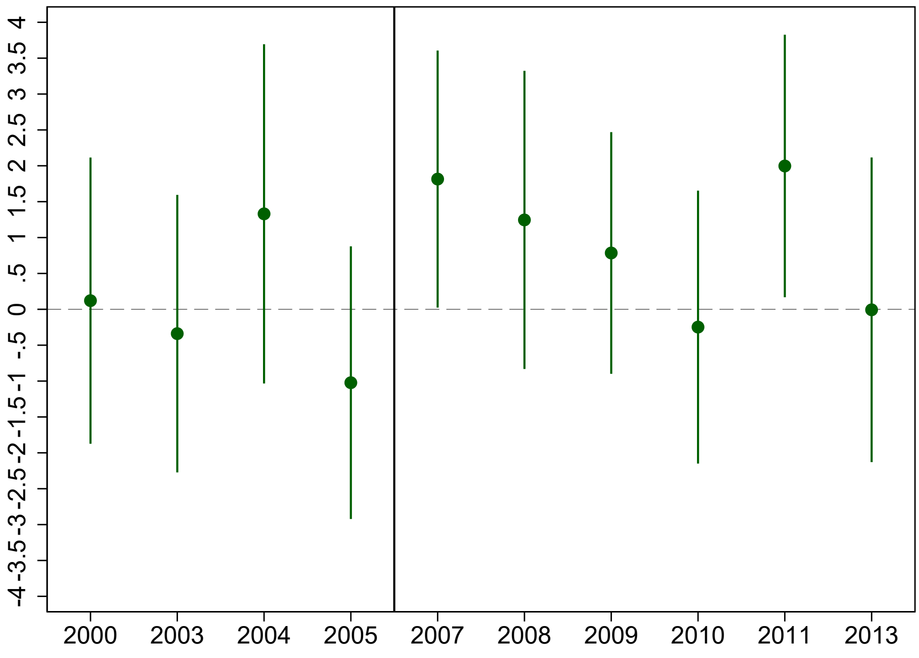
3.A Appendix A: Additional Tables and Figures

Table 3.A.1: Changes in Normal Retirement Age by Cohort denoted in Months

Cohort	$NRA_{<2007,i}$	$NRA_{>2007,i}$	$NRA_{>2007,i} - NRA_{<2007,i}$
1959	780	794	14
1960	780	796	16
1961	780	798	18
1962	780	800	20
1963	780	802	22
1964	780	804	24

Source: The table depicts the the normal retirement age for a given cohort before ($NRA_{<2007,i}$) and after ($NRA_{>2007,i}$) the reform as well as the change denoted in months.

Figure 3.A.1: Dynamic Specification, Outcome Variable: Expected Retirement Age.



Note: Figure 3.A.1 depicts the point estimates and corresponding 95% confidence intervals of year variables interacted with the change in the retirement age variable of regressions including the full set of covariates controlling for education and gender as well as year and cohort fixed effects and an individuals' retirement age expectations as outcome variable. Note that the first post reform year in this specification is 2007, as the survey was conducted from April- August 2007, i.e. when the reform was already in place. The year 2006 thus serves as a baseline.

Table 3.A.2: Placebo Results

	(1) Employment	(2) Unemployment
Interaction	-0.00002	0.00114
	(0.00113)	(0.00114)
Year FE	✓	✓
Individual FE	✓	✓
N	1,507,836	1,507,836

Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table depicts placebo regressions for the cohorts 1964-1969 for the main specification using bi-cohort times year fixed effects and individual year fixed effects.

Table 3.A.3: Cohorts 1963-64 Results

	(1) Employment	(2) Unemployment
Interaction	0.0050***	-0.0008
	(0.0021)	(0.0022)
Year FE	✓	✓
Individual FE	✓	✓
N	622,357	622,357

Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table depicts regressions for the cohorts 1963-1964 using an interaction between a dichotomous treatment variable and a post-reform indicator, and including individual and year fixed effects.

Figure 3.A.2: Outcome: Employment

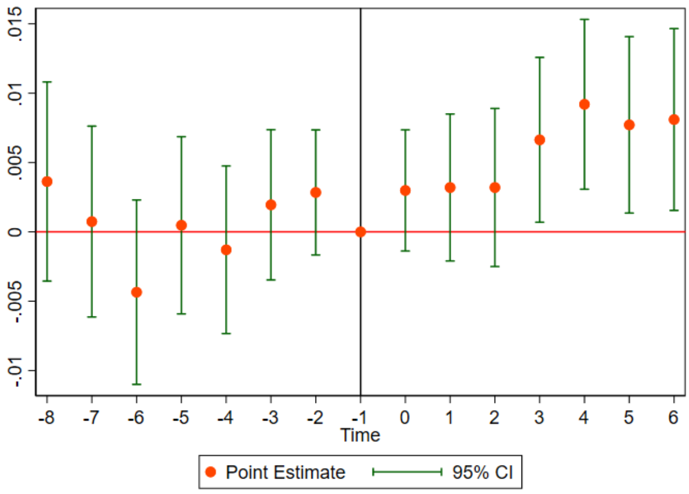
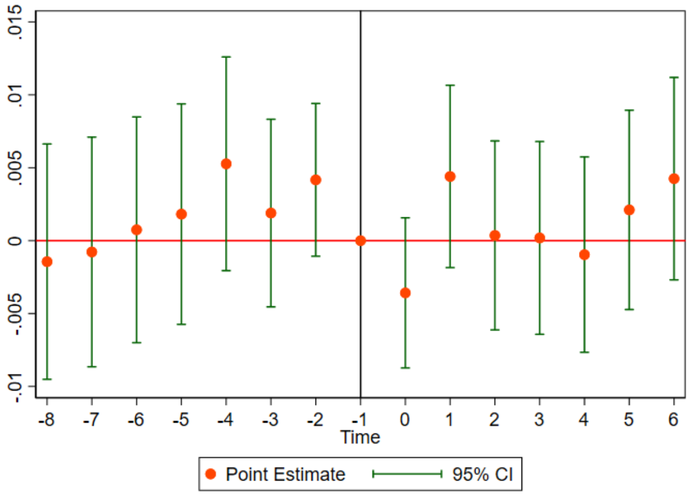


Figure 3.A.3: Outcome: Unemployment



Note: Figures 3.A.2 and 3.A.3 depict point estimates of interactions of a binary treatment and year indicator as well as corresponding 95% confidence intervals obtained from an event study specification for cohorts 1963-64 for the outcome variables employement and unemployment. The x-axis denotes the distance to treatment. The year 2007 serves as a baseline, which is denoted with the value -1 in the graph, i.e. the period before treatment fully happened in 2008 here denoted with the value 0. The y-axis denotes the value of the point estimates.

Table 3.A.4: Results First Post-Reform Year 2007

	(1) Employment	(2) Unemployment
Interaction	0.00150***	-0.00119*
	(0.00068)	(0.00071)
Year FE	✓	✓
Individual FE	✓	✓
N	1,763,754	1,763,754

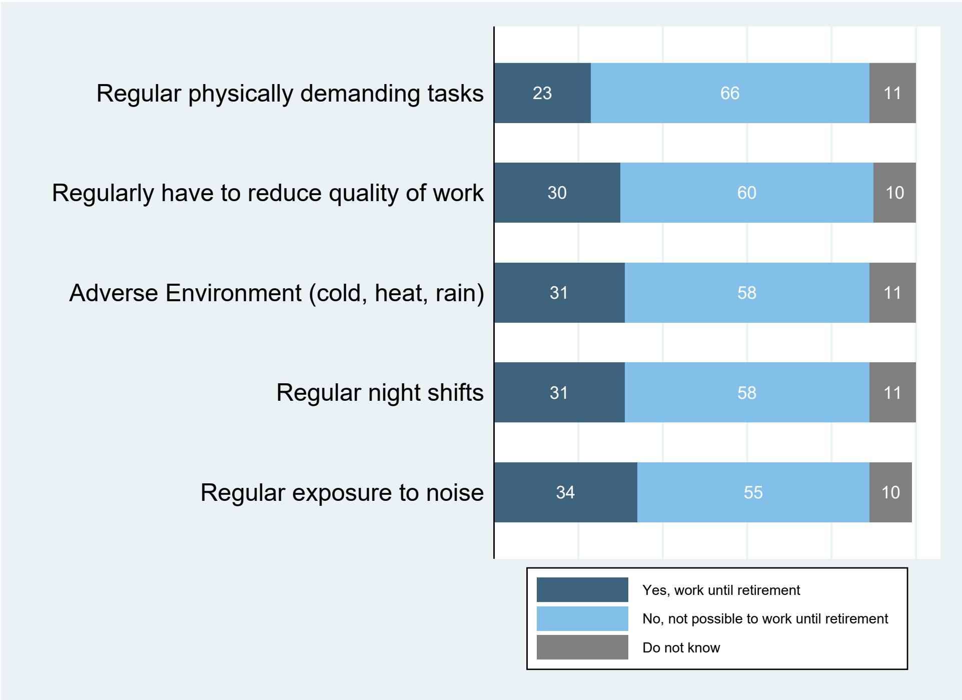
Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Note that the first post reform year in this specification is 2007. The table depicts results for the main specification using cohorts 1959-1964, bi-cohort times year fixed effects and individual year fixed effects.

Table 3.A.5: 2000-2017 Results

	(1) Employment	(2) Unemployment
Interaction	0.00268***	-0.00123*
	(0.00065)	(0.00067)
Year FE	✓	✓
Individual FE	✓	✓
N	2,094,079	2,094,079

Note: Source: SIAB data 2000-2017, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table depicts results for the main specification using cohorts 1959-1964, bi-cohort times year fixed effects and individual year fixed effects.

Figure 3.A.4: Working until Retirement by Physical and Mental Stress Indicators



Note: Figure 3.A.4 depicts average values of answers to the question *Do you think you will be able to work until statutory retirement without any restrictions given the requirements of their current job are maintained?* by individuals' job exposure to physical and psycho-social stress indicators, contained in a survey commissioned by the DGB. The DGB is one of Germany's largest trade union umbrella organisation and survey responses were collected from members and non-members across all sectors and regions in Germany between 2012-2016 of 31,164 participants. Figure is based on Gewerkschaftsbund (2016). Access to the raw data is not possible.

Table 3.A.6: Occupation Areas

Category	Occupation Area
1	Agriculture, forestry, animal husbandry and horticulture
2	Raw material extraction, production and manufacturing
3	Construction, architecture, surveying and building technology
4	Science, geography and computer science
5	Traffic, logistics, protection and security
6	Commercial services, goods trading, sales, hotel and tourism
7	Business organization, accounting, law and administration
8	Health, social issues, teaching and education
9	Linguistics, literature, humanities, social and economic sciences, media, art, culture and design

Note: Table 3.A.6 depicts the different occupational areas in Germany's 2010 occupation classification system.

Table 3.A.7: New Entry into Occupational Area

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Inter	-0.00012*	0.0002	0.000002	0.00008	0.00008	0.0002	-0.0002	-0.00007	0.00006
	(0.00006)	(0.0002)	(0.0001)	(0.00008)	(0.00019)	(0.00015)	(0.00018)	(0.00014)	(0.00007)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ind. FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	162,708	24,349	367,616	93,099	52,529	245,959	160,160	350,445	227,210

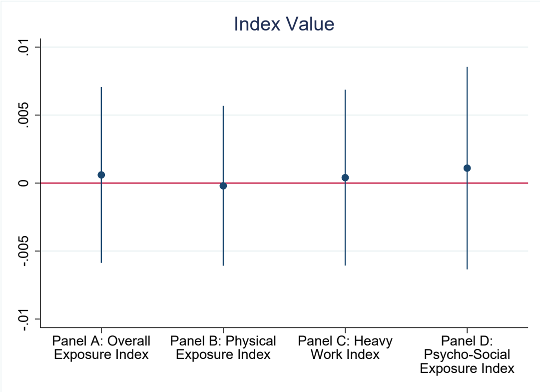
Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table depicts results for regressions using the main DiD specification for dichotomous variables equal to one for a recorded new entry into the occupation area of interest. Column numbers correspond to occupational area: 1 *Agriculture, forestry, animal husbandry and horticulture*, 2 *Raw material extraction, production and manufacturing*, 3 *Construction, architecture, surveying and building technology*, 4 *Science, geography and computer science*, 5 *Traffic, logistics, protection and security*, 6 *Commercial services, goods trading, sales, hotel and tourism*, 7 *Business organization, accounting, law and administration* 8 *Health, social issues, teaching and education* , 9 *Linguistics, literature, humanities, social and economic sciences, media, art, culture and design*

Table 3.A.8: Heterogeneity: East vs. West

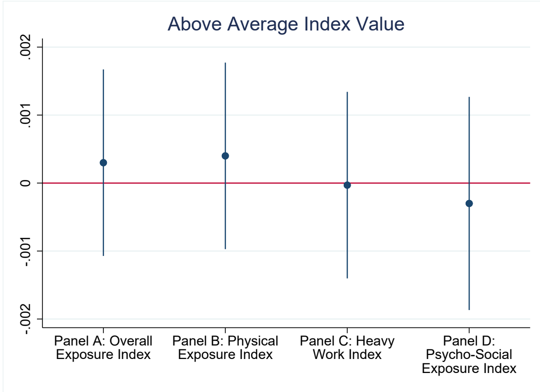
	(1) Employment	(2) Unemployment	(3) Employment	(4) Unemployment
Interaction	0.0018^{**}	-0.0010	0.00015	-0.0017
	(0.0007)	(0.0007)	(0.0014)	(0.0017)
Bi-Cohort Year FE	✓	✓	✓	✓
Individ. FE	✓	✓	✓	✓
Group	West	West	East	East
N	1,357,361	1,357,361	398,479	398,479

Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns (1)-(2) show the point estimates obtained from estimations including bi-cohort year fixed effects and individual fixed effects for West Germany. Columns (3)-(4) show the results for east Germany.

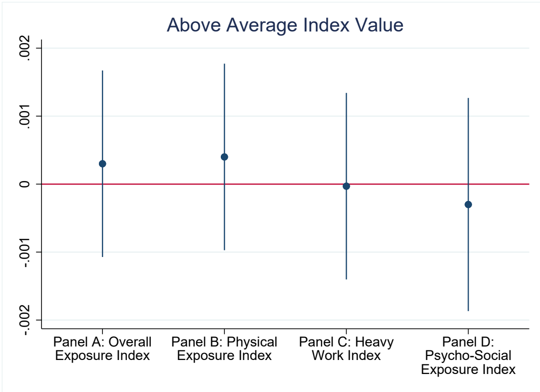
Figure 3.A.5: Sorting and Job Exposure



[a]



[b]



[c]

Note: Source: SIAB data 2000-2014, own calculations. Figure 3.A.5 depicts point estimates and 95 % confidence intervals for regressions using the main DiD specification and different outcomes measuring job exposure. Outcome variables are the respective [a] exposures' index value, [b] an indicator variable that is equal to one if the index lies above the average index value of all observed occupations and [c] an indicator variable that is equal to one if the index decreased in comparison to the previous year. .

Table 3.A.9: Heterogeneity: By Education

	(1) Employment	(2) Unemployment	(3) Employment	(4) Unemployment	(5) Employment	(6) Unemployment
Interaction	0.00346	-0.00092	0.00143*	-0.00062	0.00288**	-0.00216
	(0.00290)	(0.00288)	(0.00074)	(0.00078)	(0.00129)	(0.00147)
Bi-Cohort Year FE	✓	✓	✓	✓	✓	✓
Individ. FE	✓	✓	✓	✓	✓	✓
Group	No Vocational Training	No Vocational Training	Vocational Training	Vocational Training	College	College
N	129,955	129,955	1,321,796	1,321,796	235,893	235,893

Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns (1)-(2) show the point estimates obtained from estimations including bi-cohort year fixed effects and individual fixed effects for individuals with no vocational training. Columns (3)-(4) show the results for individuals with vocational training and columns (5)-(6) for individuals with a university degree.

Table 3.A.10: Employment Outcome by Occupational Area

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Interaction	-0.00655	0.00027	-0.00021	0.00081	0.00109	0.00338	0.00281***	0.00252*	0.0002
	(0.00418)	(0.00066)	(0.00171)	(0.00129)	(0.00168)	(0.00206)	(0.00099)	(0.00130)	(0.041)
Bi-cohort Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ind. FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	24,349	367,616	93,099	52,529	245,959	160,160	350,445	227,210	23,893

Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Column numbers correspond to occupational area: 1 *Agriculture, forestry, animal husbandry and horticulture*, 2 *Raw material extraction, production and manufacturing*, 3 *Construction, architecture, surveying and building technology*, 4 *Science, geography and computer science*, 5 *Traffic, logistics, protection and security*, 6 *Commercial services, goods trading, sales, hotel and tourism*, 7 *Business organization, accounting, law and administration* 8 *Health, social issues, teaching and education* , 9 *Linguistics, literature, humanities, social and economic sciences, media, art, culture and design*

Table 3.A.11: Unemployment Outcome by Occupational Area

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Interaction	0.00237	0.00037	0.00237	0.00030	0.00197	-0.00284	-0.00194**	-0.00280**	-0.0016
	(0.00630)	(0.00103)	(0.00324)	(0.00209)	(0.00187)	(0.00212)	(0.00096)	(0.00125)	(0.0044)
Bi-cohort Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	24,349	367,616	93,099	52,529	245,959	160,160	350,445	227,210	23,893

Note: Source: SIAB data 2000-2014, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Column numbers correspond to occupational area: 1 *Agriculture, forestry, animal husbandry and horticulture*, 2 *Raw material extraction, production and manufacturing*, 3 *Construction, architecture, surveying and building technology*, 4 *Science, geography and computer science*, 5 *Traffic, logistics, protection and security*, 6 *Commercial services, goods trading, sales, hotel and tourism*, 7 *Business organization, accounting, law and administration* 8 *Health, social issues, teaching and education* , 9 *Linguistics, literature, humanities, social and economic sciences, media, art, culture and design*

Table 3.A.12: Results by Exposure Indices for Women

	(1) Employment	(2) Unemployment	(3) Employment	(4) Unemployment
Panel A: Overall Exposure Index (OJI)				
Interaction	0.00293*	0.00185	0.00291***	-0.00239**
	(0.00166)	(0.00149)	(0.00105)	(0.00098)
N	351,628	351,628	393,894	393,894
Panel B: Physical Exposure Index (OPI)				
Interaction	0.00268*	0.00196	0.00321***	-0.00260***
	(0.00161)	(0.00145)	(0.00107)	(0.00099)
N	368,970	368,970	376,586	376,586
Panel C: Heavy Work Index (HWI)				
Interaction	0.00288*	0.00071	0.00274**	-0.00227**
	(0.00153)	(0.00142)	(0.00113)	(0.00101)
N	378,806	378,806	366,834	366,834
Panel D: Psycho-Social Exposure Index (OSI)				
Interaction	0.00246	0.00200	0.00310***	-0.00218**
	(0.00162)	(0.00146)	(0.00107)	(0.00100)
N	355,245	355,245	389,602	389,602
Bi-Cohort Year FE	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓
Group	Above Mean	Above Mean	Below/Equal Mean	Below/Equal Mean
Observed Years	2000-2014	2000-2014	2000-2014	2000-2014

Note: Source: SIAB data, own calculations. Standard errors in parentheses and clustered at individual * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table depicts regression results for the two main outcome variables employment and unemployment throughout a given year for women in occupations with an above or below/equal mean value of the respective stress exposure index denoted in each panel. The mean value is derived from the whole sample and thus the same value as in Table 3.8.

Table 3.A.13: Results by Exposure Indices for Men

	(1) Employment	(2) Unemployment	(3) Employment	(4) Unemployment
Panel A: Overall Exposure Index (OJI)				
Interaction	-0.00000	-0.00053	0.00124*	-0.00031
	(0.00083)	(0.00120)	(0.00066)	(0.00095)
N	499,201	499,201	317,422	317,422
Panel B: Physical Exposure Index (OPI)				
Interaction	0.00020	-0.00052	0.00120*	-0.00034
	(0.00079)	(0.00115)	(0.00070)	(0.00100)
N	528,638	528,638	288,163	288,163
Panel C: Heavy Work Index (HWI)				
Interaction	0.00001	-0.00019	0.00096	-0.00021
	(0.00088)	(0.00128)	(0.00062)	(0.00090)
N	459,274	459,274	356,758	356,758
Panel D: Psycho-Social Exposure Index (OSI)				
Interaction	0.00066	-0.00004	0.00032	-0.00027
	(0.00094)	(0.00131)	(0.00062)	(0.00099)
N	373,775	373,775	440,919	440,919
Bi-Cohort Year FE	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓
Group	Above Mean	Above Mean	Below/Equal Mean	Below/Equal Mean
Observed Years	2000-2014	2000-2014	2000-2014	2000-2014

Note: Source: SIAB data, own calculations. Standard errors in parentheses and clustered at individual level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table depicts regression results for the two main outcome variables employment and unemployment throughout a given year for women in occupations with an above or below/equal mean value of the respective stress exposure index denoted in each panel. The mean value is derived from the whole sample and thus the same value as in Table 3.8.

3.B Appendix B: RDD Pension Reform 1999

This section of the appendix is dedicated to the additional provision of RDD results. Figures 3.B.1 and 3.B.2 show the typical RDD graphs for each outcome variable. Table 3.B.1 shows the results of a placebo test for cohorts 1952-53 between 2000-2006 when they were not differently affected by any reform. The results clearly do not show any significant effect on any of the analyzed outcome variables except for the quadratic specification and the unemployment outcome, which shows a significant point estimate at the ten percent level, albeit positive and thus in the opposite direction of the documented reform effects.

Table 3.B.1: RDD Placebo Results

	(1) Employment	(2) Unemployment
	Panel A: Linear Running Variable	
Treatment	0.0091	0.0031
	(0.0159)	(0.0047)
	Panel B: Quadratic Running Variable	
Treatment	-0.0056	0.0107*
	(0.0210)	(0.0073)
Covariates	✓	✓
Year FE	✓	✓
Individual FE	✓	✓
N	108,830	108,830

Note: Source: SIAB data, own calculations. Standard errors in parentheses and clustered at birth month level. Results obtained from local regressions for cohorts 1952-1953. Panel A shows the results for a local linear and Panel B for local quadratic specification; significance levels
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A key assumption underlying the RD design is that covariates vary smoothly across the cut-off. I provide support for this assumption by using individual control variables as outcomes. More precisely, I regress an indicator variable whether individuals have a German high school degree (Abitur) and a variable measuring work experience at the beginning of the sample period (measured in days) on the linear and quadratic RDD specification. Table 3.B.2 shows that there is no significant effect on these covariates, supporting the notion that in the absence of the policy individuals close to the cut-off are similar and would have behaved similarly.

Table 3.B.2: RDD Balancing Test

	(1) Abitur	(2) Work Experience
	Panel A: Linear Running Variable	
Treatment	0.0013	-91.71
	(0.0063)	(80.12)
	Panel B: Quadratic Running Variable	
Treatment	0.0013	-90.9
	(0.0075)	(102.73)
Year FE	✓	✓
Individual FE	✓	✓
N	136,199	136,199

Note: Source: SIAB data, own calculations. Standard errors in parentheses and clustered at birth month level. Results obtained from local regressions. Panel A shows the results for a local linear and Panel B for local quadratic specification; significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3.B.1: Outcome: Employment

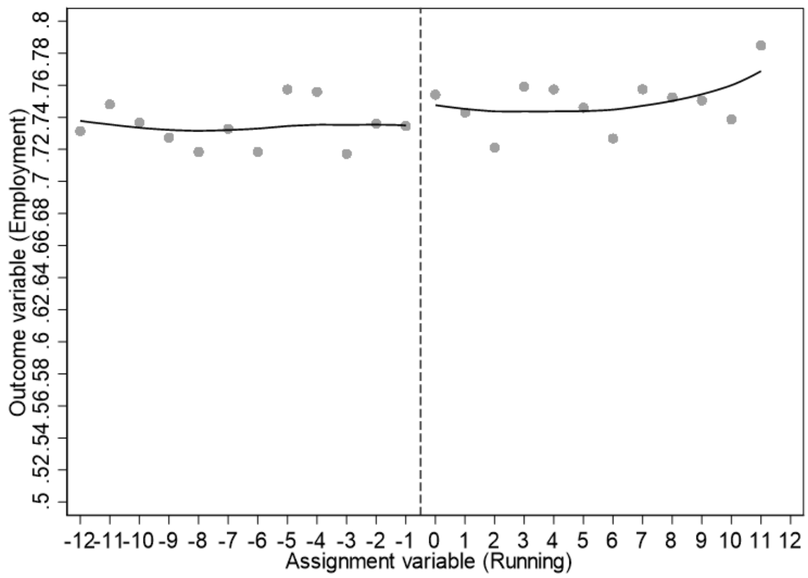
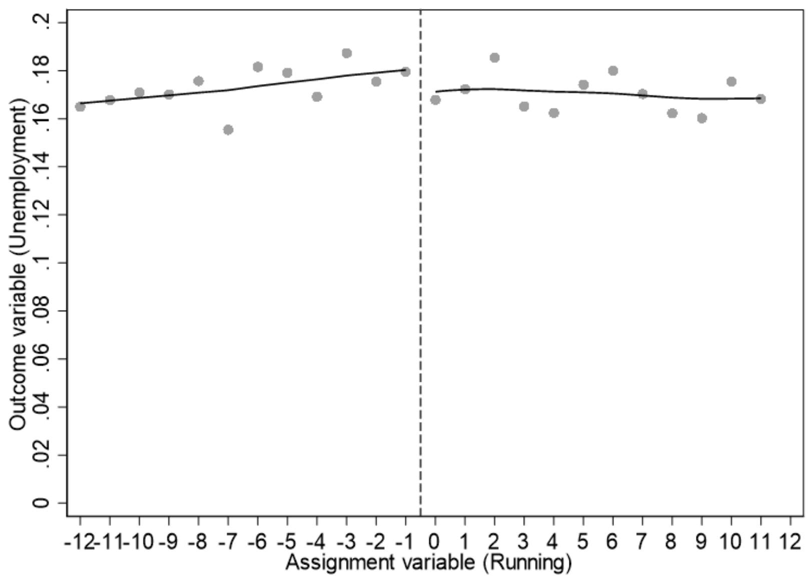


Figure 3.B.2: Outcome: Unemployment



Note: paper2/graphs3.B.1 to 3.B.2 show RDD graphs for the outcomes (a) Employment and (b) Unemployment. The x-axis depicts the running variable and the y-axis average values of the respective outcome variables with local quadratic polynomials with a bandwidth of 12 months on each side and a triangular kernel.

Chapter 4

House Price Expectations.[1]

Abstract

This study examines short-, medium-, and long-run price expectations in housing markets. We derive and test six hypothesis about the incidence, formation, and relevance of price expectations. To do so, we use data from a tailored household survey, past sale offerings, satellites, and an information RCT. As novel finding, we show that price expectations show no evidence for momentum-effects in the long-run. We also do not find much evidence for behavioural biases in expectations related to individual housing tenure decisions. Confirming existing findings, we find momentum-effects in the short-run and that individuals use aggregate price information to update local expectations. Lastly, we corroborate existing evidence that expectations are relevant for portfolio choice.

[1]Co-authored with Peter Haan, Claus Michelsen and Felix Weinhardt.

4.1 Introduction

Housing is a key component of wealth, consumption and the best collateral. This makes housing critical to household decisions on consumption and saving, bank lending decisions and asset pricing. Since these factors are relevant for economic growth and financial stability, housing markets are important. As a result, understanding housing markets or determinants of tenure decisions has long been in the focus of both micro- and macro-economists. For example, it is well understood that housing markets are characterized by short-run momentum (Case and Shiller, 1989) and longer-run mean reversion (Capozza et al., 2002; Glaeser and Gyourko, 2006). In contrast to the evidence about observed market prices, evidence about the incidence and formation of price expectations is scarce and only speaks to the short or medium run., see e.g. Armona et al. (2019) or Kindermann et al. (2021). Armona et al. (2019) document momentum-effects in price expectations for periods up to five years. While five years is a long period to form expectations over, in terms of housing cycles, five years is rather short. Therefore, it is also important to understand long-run price expectations.

This paper systematically examines the incidence, formation, and relevance of price expectations over a period that captures a full housing cycle. Specifically, we study price expectations over comprehensive periods relevant for housing markets: in the short run (2 years), medium run (10 years), and the long run (30 years). In the next section, we derive six hypothesis regarding: (1) local housing markets and momentum-effects, (2) individual characteristics such as financial literacy, (3) the role of past tenure decisions, (4) the role of land supply, (5) aggregate information, and (6) the relevance of expectations for portfolio choice, which we then test empirically in the data. The key finding of the paper is that while different factors matter for price expectations over different time dimensions (discussed in detail below), there is clear evidence for momentum-effects in the short-run but *not* in the long-run. Thus, we show that there is no tension between price expectations and market cycles.

At the heart of our analysis is the design and inclusion of specific and novel housing-related questions in a representative household panel survey, the Innovation Sample of the German Socio Economic Panel (SOEP IS). We elicit beliefs and preferences, alongside a rich set of individual and household characteristics, including detailed information about current tenure and past tenure decisions. Specifically, in the survey, we ask representative households about their expectations about the development of house prices in their current neighborhood over a horizon of 2, 10, and 30 years, respectively. Besides these novel questions, respondents also complete the SOEP-core questions, meaning that we have access to an unusually large set of individual- and household-level background characteristics that might affect the formation of price expectations, including measures of educational background and financial literacy. For our analysis, we augment the survey-information further and merge with a number of additional data sets: local data on housing prices, as well as satellite data and remotely sensed terrain data to construct an index of housing supply in the spirit of Saiz (2010). These additional data allow us to study how price expectations correlate with local housing market characteristics. Moreover, we conduct a randomized information treatment to study causal effects of information on past aggregate price trends on expectations and on a hypothetical portfolio choice outcome.[2]

Our first important, purely descriptive, finding is that price expectations for the long-run appear very low, in particular when compared to short-run expectations. This result is in contrast to the only existing evidence on longer-run expectations that we are aware of: Case and Shiller (2003) and Case et al. (2012) present evidence that individuals in both hot and cold markets across the USA generally have very high annualised growth expectations over a ten-year period. These findings contributed to the notion of strong behavioural biases in housing markets. We discuss a potential reason for this difference in findings with the literature below.

We then proceed to test for our hypotheses. First and foremost, examining the role of local markets, we find that local price trends have very strong predictive power for local expectations in the short term. Each percentage point higher experienced local house prices growth leads to a 0.18-0.19 increase in expectations of annual growth rates.

[2]We discuss the “aggregate” treatment in detail in section 4.3.2.2.

This estimate is extremely close to the recent experimental estimate of 0.2 percentage points provided by Armona et al. (2019). Importantly, and as a new finding, for long-run expectations we do not find evidence of economically significant momentum-effects, i.e. these seem to be largely independent of past local market conditions. The longer the time-span considered, the smaller the estimates of momentum effects: our best elasticity-estimate for annual expectations are at 0.0353 for the ten-year and merely at 0.0126 for the thirty-year period. These thirty-year estimates are 7.5 to 14.5 times smaller compared to the corresponding short-run estimates. Importantly, these expectations are estimated with precision. In fact, some of the long-run estimates remain significant at conventional levels of statistical significance. Thus, these results show that, economically, current market conditions do not affect long run expectations by much. These findings speak against "irrational exuberance" in the long-run and are consistent with individuals internalising market cycles. Second, we find that individual characteristics have only modest effects on expectations. This holds for the short run, the medium run, and the long run. Females are more conservative and the well-educated have slightly lower expectations for the short run but higher expectations in the longer term. Third, we refute our hypotheses related to historical decisions and regret and loss aversion: we find that the own tenure status does not correlate with expectations, both in booming and not-performing markets. On the other hand, consistent with Kindermann et al. (2021), our results show that renters who have experienced a recent rental rise have higher price expectations. Fourth, in line with more rational forward-looking behaviour, land supply measured by available building land using satellite imagery, as stipulated by housing market theories, correlates negatively and strongly with long-run expectations. Fifth, in line with recent empirical literature, our findings also document that individuals factor in more aggregate, distant information, such as past aggregate OECD-level trends, when forming expectations. This hypothesis is tested using an information RCT that presents data on past (averaged) price developments of OECD countries. Last but not least, we show that this information RCT also shows effects on a portfolio choice question.

This paper is related to a new and quickly growing literature that examines how price expectations are formed in the housing market and if these affect individual behaviour.

In particular, Kuchler and Zafar (2019) show that individuals use regional price trends to form expectations about future prices at the national level. In a different context, Bailey et al. (2018) show that housing decisions of distant friends on Facebook have effects on local behaviour. Niu and van Soest (2014) show that medium run expectations are positively related to past house price developments and perceived economic conditions. Kindermann et al. (2021), focusing on short term price expectations in Germany, find that renters have higher and more accurate price expectations. In a quantitative model, they explain the difference and show that renters are relatively well informed about house prices. Most closely related to this study, Armona et al. (2019) study how local market conditions and information affect the formation of price expectations. They find that local markets have strong effects on individuals' expectations in the short- and medium-run, without any evidence for mean reversion. Moreover, they show that information about markets affects expectations and stated investment decisions. So far, most studies consider price expectations in the short- or medium-run. For example, Armona et al. (2019) find evidence for momentum-effects over one-year periods, with no evidence for mean-reversion for periods of up to five years. While five years is a long period to form expectations over, in terms of housing cycles, five years might not be long enough. Whether evidence for psychological driven volatility or momentum-effects (Shiller, 2015, e.g.) persists in price expectations also over longer time-horizons is an open question. Moreover, there exists no evidence how individual characteristics such as education, numeracy, tenure type or local housing market characteristics affect price expectations over longer horizons. This study starts filling this gap and we regard the lack of evidence for momentum-effects over the long-run as the most interesting contribution of this paper.

Finally, this paper also relates to a literature that studies how price expectations can be extracted using surveys (e.g. Manski (2004), Binswanger and Salm (2017) or Breunig et al. (2021b)). We chose to ask for point estimates[3] for total expected price changes

[3]Due to the in-person nature of the SOEP survey the number of additional questions that we could include into the SOEP-IS was limited. This is why we did not also ask for probabilistic expectations in housing markets or for histograms. Breunig et al. (2021b) show that point expectations for the mean are highly correlated with the means derived with probabilistic questions or histograms in the context of the stock market. Therefore, we regard this as minor limitation of our study. However, the measurement of long-run expectations in surveys is a interesting field of research in itself and testing to what extent housing market expectations differ depending on the elicitation method a route for future work.

over periods up to thirty years, which allows us to back out average annual growth rates. The reason for asking directly for long run expectations rather than for annualised growth rates is that there is evidence that individuals have problems to understand the compounding of growth rates (or interests). This might become particularly problematic when considering longer-run expectations. Case and Shiller (2003) and Case et al. (2012) ask for average annual growth rates over ten-year periods and document very high total expectations. The different question might be part of the explanation why our descriptive results differ, although theoretically it should not matter if one asks for total price changes or expected average annual growth rates over the same periods. However, we acknowledge that important questions on how to best elicit long-run price expectations remain and see this as important avenue for future work.

The remainder of the paper is structured as follows: the next section describes the hypotheses tested in this study in more detail. Section 4.3 gives a brief overview of the German housing market and introduces the data sets used in the paper. Section 4.4 describes the empirical setting of our study. In Section 4.5 we present our results and robustness checks, respectively. Last, Section 4.6 concludes.

4.2 Formation of Expectations: Derivation of Hypotheses

H1: No Momentum-Effects in the Long-Run We explore the hypothesis that local housing markets matter for the formation of expectations in the short-, medium-, and long-run by testing the effect of past local prices trends on expectations. The hypothesis is motivated by previous literature, as studies show that past local prices trends do positively affect price expectations for periods of up to five years; see e.g. Case et al. (2012); Niu and van Soest (2014); Armona et al. (2019). Also looking beyond the short-run, Armona et al. (2019) find evidence for momentum-effects for price expectations in the short-run (i.e. after one-year) as well as for periods of up to five years (medium-run). However, there is ample evidence that housing cycles exist and that markets show mean-reversion over longer periods, i.e. Wheaton (1999); Glaeser and Gyourko (2006). Given the duration of the most recent cycles (Bracke, 2013), we do not necessarily expect to see evidence for mean reversion over ten years - but certainly over the thirty-year period momentum-effects should not dominate. Therefore we hypothesize that we will not find evidence for momentum-effects in long-run expectations.

H2: The Role of Individual Characteristics Individual characteristics can influence expectations in the short-, medium-, and long-run. A large literature documents gender-differences in willingness to take financial risks (Charness and Gneezy, 2012; Almenberg and Dreber, 2015, e.g.). Applied to the formation of expectations, we hypothesize that males are more likely to hold large and positive price expectations, see e.g. Breunig et al. (2021a). A second individual factor that may influence the formation of expectations is the level of education. More highly educated individuals should be better in processing information for the formation of price expectations. Behavioural biases, such as not compounding interests and financial illiteracy, are shown to diminish with better education, see e.g. Lusardi (2008). We test this hypothesis using educational information (the highest educational degree obtained) and information about the financial literacy and self-assessed numeracy of the individual.

H3: The Role of Housing and Tenure The own housing and the tenure situation may affect the expectations of prices. In particular, owner-occupiers could be more optimistic compared to renters - simply because they partly select into the respective housing tenure based on their expectations. Equally, reference point theory would suggest that past decisions could have additional influences on the formation of expectation: Individuals who have purchased a home (owner-occupiers) might not downward-adjust expectations to rationalize past decisions ex-post, see e.g. Lamorgese and Pellegrino (2019); Genovese and Mayer (1997). More generally homeowners might attach a greater value to their home, see e.g. Goodman et al. (1992); Chan et al. (2016); Bao and Gong (2016). Similarly, homeowners who decide against selling at times of relatively high housing prices might experience regret aversion and, consequently, would rather not consider selling their house, see e.g. Seiler et al. (2008). In sum, selection into owner-occupation as well as behavioural biases would suggest that owner-occupiers should hold more positive price expectations compared to renters. Expectations of renters, on the other hand, could be affected by recent rental rises. This, in turn, could lead to higher price expectations of renters (Kindermann et al., 2021).

H4: The Role of Local Housing Markets In addition, the elasticity of land is shown to play a key role for long-run housing supply and house price developments (Saiz, 2010). In theory, more available land for residential development should dampen future price increases. This could also be the case when examining price expectations, in particular for the long-run.

H5: The Role of Aggregate Information The literature highlights that information, such as housing decisions of distant friends on Facebook, affects local behaviour; see e.g. Bailey et al. (2018). Further, Kuchler and Zafar (2019) show that local price trends significantly shape national, aggregate house price expectations. We hypothesize that information on past aggregate house price developments affects expectations. Specifically, we test if information on residential real estate prices in 14 different countries, including Germany and the United States, that on average approximately quadrupled since 1945, has a positive effect on individual house price expectations. We test this using a RCT, which we describe further in Section 4.3.

H6: The Relevance of Expectations for Investment Decisions In the final hypothesis, we move beyond the formation of expectations and directly test if expectations about house prices affect planned investment. To provide a causal answer to this question, we estimate if information on past aggregate house price developments (see H5), which we provide using an RCT, affects stated investment decisions of individuals.

4.3 Institutional Background and Data

4.3.1 The German Housing Market

In comparison to most OECD countries, the German housing market has a relatively low share of homeownership, with owner-occupied dwellings accounting for less than half of all occupation forms (Voigtlaender, 2010). In rural and suburban areas, especially in eastern Germany, the housing market is relatively stable. In contrast, in urban agglomerations and student cities, severe housing shortages dominate the local housing markets, ultimately putting upward pressure on housing prices in these areas (Kholodilin et al., 2016). However, even within big cities, substantial variations in local housing markets exist (BMF, 2017). The regional differences in the German housing market are, in turn, reflected in rent and house price developments. Particularly, since 2007/2008, housing prices have strongly increased in popular urban areas, including the Big Seven cities[4] but stagnated or only increased marginally in many rural areas. The German context hence offers rich variation, making it a relevant setting for the study of the formation of price expectations.

4.3.2 Data

4.3.2.1 The German Socioeconomic Panel Innovation Sample

For the analysis, we use a range of different data sets. The main data set is the German Socioeconomic Panel's Innovation Sample (SOEP IS). The SOEP IS an annual representative survey providing information on a large set of socioeconomic and demographic variables and specific survey modules, see Richter and Schupp (2015) and Goebel et al. (2019). As part of the SOEP IS waves in 2016-2018, we were able to design a specific module of questions to elicit price expectations of households. For this analysis, the questions on the short-, medium-, and long-term formation of house price expectations of

[4] Berlin, Hamburg, Munich, Cologne, Duesseldorf, Frankfurt am Main, and Stuttgart

individuals are of central interest. In all waves, we ask for expectations of local housing prices over a two- and thirty-year horizon. More precisely, individuals were asked the following question:

The following section focuses on your expectations of house price developments in your area. In your opinion: How will house prices develop in comparison to today?

Participants then had the option to state whether prices will fall, rise, remain the same or not answering at all. Depending on their answer individuals were asked *By how much - in percent- do you think prices will fall/rise.* For the 2017 wave, we additionally asked for expectations over a ten-year horizon. Crucially, the survey is regionally representative, sampling individuals from a total of 183 postal codes covering 75 out of 90 residential postal code regions in Germany (see Figure 4.B.1).

In addition, the SOEP IS data provides a wide range of socioeconomic variables. For example, individuals were asked to assess their own math skills as rather proficient or rather bad. Further, individuals received several questions testing their financial literacy. Based on whether individuals correctly answer these questions, we construct three measures of financial literacy; see Grohmann et al. (2019).[5]

4.3.2.2 RCT on past Aggregate Price Trends

To test the importance of aggregate house price information, we designed a randomized controlled information trial that we could implement within the SOEP IS household survey. As part of the RCT, we provide randomly selected survey participants with

[5] More precisely, individuals were asked the following three questions:

1. Assume the interest of your savings account is 1% per year and the inflation rate is 2% per year. What do you think: One year from now, could you buy the same, more, or less than today with the credit at your savings account (Answer options: More, less, the same, I do not know).
2. Assume that you have a 100 euros deposit in your savings account. This credit bears interest at 2% per year and you leave it on this account for 5 years. How much credit will your savings account have after 5 years? (Answer options: More than a 102 euros, less than a 102 euros, Exactly a 102 euros, I do not know)
3. Is the following statement correct or wrong? The investment into shares of one company is less risky than into an equity fund.

background information on past aggregate housing price developments, see Figure 4.A.1. Additionally, the treatment group received the information that the graph depicts the development of average prices in residential real estate in 14 different countries, including Germany and the United States, and that prices on average approximately quadrupled since 1945. In the 2018 wave of the SOEP IS, we repeated the RCT for the same individuals.

In the 2018 wave, individuals were additionally asked to allocate assets between different investment options. More precisely, participants were asked the following:

Suppose you have some spare money and have decided to invest this money. How would you in per cent allocate your money between the following investment options?

The given options were stocks, real estate, state bonds, savings account, and gold.

Note that due to the inclusion of this RCT into the in-person SOEP household survey it was not possible to use local price information as treatment, like in Armona et al. (2019). The rational for proceeding with aggregate information is given by Kuchler and Zafar (2019), who show that individuals do not sharply distinguish between aggregate and local price information for housing markets.

4.3.2.3 Postal Code, House Price and Land Supply Data

We use three additional data sets to obtain information on regional demographics and housing markets. First, we use the *empirica ag* housing data bank for the years 2012 to 2016 (pre-period) to derive postal code specific hedonic regional house price trends. The *empirica ag* data bank contains all listings and deals from 2012 to 2016 conducted through Germany's largest online real estate platforms such as *Immoscout* and *immowelt/immonet*, newspapers and local online platforms. In total, the data bank provides 484,604 observations for rental and non-rental properties over the given time period. Table 4.A.7 provides summary statistics of the observed transaction prices. In addition, the data base offers a wide range of background information on the listed properties such as the dwelling's size, age, state, room number and equipment. We use this information to construct hedonic, quality-adjusted average annual house price trends for each postal code over the 2012-2016 period. Appendix B describes the derivation of the hedonic price trends in more detail and Figure 4.B.1 in Appendix B depicts the hedonic price trends for each postal code region in Germany. The quality-adjusted trends are one indicator that we employ to assess how local housing markets impact housing expectations.

As a second indicator, we construct a postal code specific developable land index that serves as a proxy for local housing supply elasticity. For this, we use Corine Land Cover 2018 data to identify land that is potentially developable. The Corine Land Cover data is based on 100m resolution satellite data that, via remote sensing, provides the basis for over 37 land cover and land use categories, such as *sports and leisure facilities* or *wetlands*. Using this information, we construct an index of developable land supply for Germany similar to Saiz (2010).[6] More precisely, the index measures the share of developable land in each postal code. As an example, Figure 4.A.2 in the Appendix depicts the index as constructed for Berlin and surroundings. The pink areas show land that is potentially developable. As expected, more land is developable in the rural areas surrounding Berlin, whereas within the city itself developable land is relatively scarce. Lastly, in order to control for postal code specific characteristics, such as population density, we use Census data.[7]

[6]We abstract from zoning decisions due to a lack of data, but see these as potentially malleable in the long run anyway. Accordingly, we classified as developable all categories that could be used for supply, in principle: Non-irrigated arable land; Vineyards; Fruit trees and berry plantations; Pastures; Complex cultivation patterns; Land principally occupied by agriculture, with significant areas of natural vegetation; Agro-forestry areas; Broad-leaved forest; Coniferous forest; Mixed forest; Natural grasslands; Moors and heathland; Transitional woodland-shrub; Beaches, dunes, sands; Bare rocks; Sparsely vegetated areas.

[7]Note that the last Census was conducted in 2011. Data on population size is available.

4.4 Empirical Approach

The empirical approach follows directly from the hypotheses presented in Section 4.2. The first set of results that we use to test the first hypothesis is descriptive in nature and we take great care in establishing representativeness of our respondents.

In order to test the relation between house price expectations and the impact of individual characteristics, housing/tenure, information, as well as local housing markets, respectively, we adopt the empirical framework described below in Equation 4.1. Y_i are the two, ten, and thirty-year expectation outcomes, respectively. Depending on the hypothesis we seek to test, we separately include sets of different explanatory variables.

$$Y_i = \beta_0 + C_i'\beta_1 + H_i'\beta_2 + \beta_3 I_i + L_i'\beta_4 + \beta_5 + \sum_y^Y \eta_y + \sum_r^R \theta_r + u_i \tag{4.1}$$

First, we run the regression on past price changes, i.e. L_i. More precisely, we measure past price changes using a hedonic regression to back out the average annual growth rate over the past six-years at the postal code level. Our findings are robust to other measures of past price changes. Next, to examine the role of individuals characteristics, we regress the expectation outcomes on C_i, which is a matrix of individual characteristics, like the level of education, gender, as well as variables measuring financial literacy and self-assessed numeracy skills. Third, we regress the expectation outcomes on housing and tenure characteristics contained in H_i, such as the tenure choice and past rental increases. Fourth, we test the impact of local supply using the land supply index from remote sensing described above. Finally, in order to analyze the role of information, we regress the expectation outcomes on a variable indicating whether the individual received the information treatment, I_i. Here, we naturally restrict our sample to the years in which the RCT was conducted, i.e. 2017 and 2018. In addition, we use a portfolio choice question asked in 2018 as another outcome variable to complement the analysis and to test if expectations have an effect on behaviour.

Depending on the specification, we control for the year of the survey by including year fixed effects, η_y. Further, in order to control for potential unobserved regional variation in expectations and housing markets, we include commuting zone fixed effects, i.e. θ_r. Commuting zones are based on the so called *Raumordnungsregionen* provided by the German government. In total there are 96 of these regions across Germany. Crucially, the commuting zones are larger than postal codes in Germany, which, in turn, allows us to exploit the within commuting zone variation whilst controlling for time-invariant unobservables at the commuting zone level. This is crucial since there might be unobservable characteristics that simultaneously are correlated to price expectations and local housing market performance. Further, we control for a range of individual and postal code covariates contained in X. More precisely, when appropriate, we control for the information treatment half of the participants in the 2017 and 2018 samples received. In addition we control for age, personal income, tenure duration and the postal code log population density.

4.5 Results

4.5.1 Descriptive Statistics

Table 4.1 provides summary statistics for house price expectations in two, ten and thirty years. As previously described expectations over ten years were solely part of the survey in 2017. Thus, observation numbers are smaller by construction for this outcome. For the two and thirty year outcomes it is important to note that more individuals (2687 vs. 2079) answered the question over the shorter time frame than over the long horizon. In order to ensure that there is no systematic difference in the expectation outcomes stated by those who answered only one question, we in addition to the summary statistics for all observations (Panel A) provide results for survey participants only that answered both of the questions on two- and thirty-year expectations (Panel B).

On average individuals expect house prices to increase by just under 10 percentage points

Table 4.1: Summary Statistics: House Price Expectations

Expectations	2 years	10 years	30 years
Panel A: All observations			
Mean	9.7940	16.3846	29.1819
Median	10	5	20
25th Percentile	5	10	10
75th Percentile	15	20	40
Annual Growth Rate	4.80	1.53	0.87
N	2679	754	2071
Panel B: Only obs. that answered 2/30 year questions			
Mean	9.8478	16.1357	29.1411
Median	10	10	20
25th Percentile	5	7	10
75th Percentile	15	20	40
Annual Growth Rate	4.83	1.51	0.87
N	1984	561	1984

Notes: Table 4.1 depicts mean values of house price expectation for 2/10/30 years. Each mean is derived controlling for RCT and year indicators. Panel A depicts summary statistics for all individuals and Panel B only for those who answered both the 2 and 30 year expectation questions. Source: SOEP Innovation Sample 2016-2018; 10 year expectations only exist for 2017 sample. Panel A depicts summary statistics for all individuals and Panel B only for those who answered both the 2 and 30 year expectation questions.

in the next two years which corresponds to a compounded average annual growth rate of just under 5 percentage points. Over a ten-year horizon individuals on average expect an increase of more than 16 percentage points. This corresponds to a compounded annual growth rate of approximately 1.53 percentage points. Over a thirty-year frame the survey participants expect prices to increase by only about 30 percentage points with a compounded annual growth rate of under one percent. Figure 4.1 represents the finding of lower annual growth rates graphically. Crucially, there is no substantial difference in the results of Panel A and B. This supports the notion that there is no systematic difference in the reported outcomes of those who answer both questions and those who only answer one. We thus continue our analysis using the answers of all respondents.

Examining the annual growth rates in Table 4.1, it is immediate to see that the longer the time horizon over which the expectations are given, the lower the corresponding annual growth rate. Figure 4.1 represents this finding graphically and depicts hypothetical growth trajectories based on the respective compounded annual growth rates derived from the different time frames over which expectations where collected. Crucially, we assume a constant annual growth to calculate the growth trajectories. For example, based on an annual average compounded growth rate of just under five percent derived from the two year time frame the corresponding 30 year ahead forecast considering compounding growth rates and under the assumption of constant annual growth is approximately 308 percent. Asked over a ten-year time frame the 30 year ahead forecast is reduced to almost half, i.e. approximately 158 percent, and for a time frame of 30 years we obtain an average forecast of just under 30 percent. The black triangle indicates the actual expectations stated when directly asked about the relevant time frame and hence corresponds to the mean values stated in Panel A of Table 4.1. Most importantly, we see that the growth trajectories become flatter the longer the time horizon over which expectations are elicited.

Figure 4.1: Growth trajectories based on the annual growth rates for 2/10/30 year expectations assuming constant annual growth.

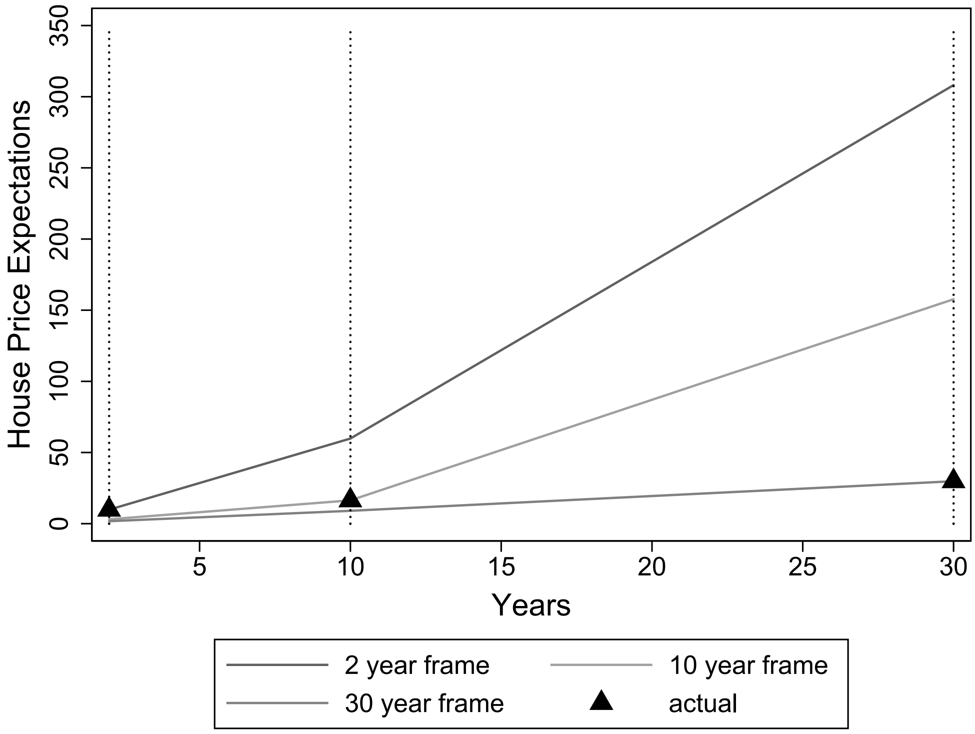


Figure 4.1 depicts growth trajectories based on annual growth rates derived respectively from the 2/10/30 year expectation means stated in the SOEP IS 2016-2018, see Table 4.1. Crucially, each forecast assumes that annual growth rates remained constant. The black triangle indicates the actual expectations stated when directly asked about the relevant time frame and hence corresponds to the mean values stated in Panel A of Table 4.1.

4.5.2 Main Results

Momentum Effects over different Time Periods (H1) To test for momentum effects, we regress (total) future price expectations on experienced annual local house price changes using specification described in Equation 4.1. Table 4.2 shows the estimates for periods of two, ten and thirty years in columns (1)-(6), each time first without and then with controls for year fixed effects, commuting zone fixed effects, age, personal income, tenure duration and the postal code log population density. For example, the estimate in column 2 shows that individuals who experienced a one percentage point higher annual price growth in their postal code expect that prices will increase by 0.4175 percentage points more over the next two years. This estimate is highly statistically significant. For the longer periods, the estimates for effects on total price changes become slightly larger,

Table 4.2: The Role of Momentum Effects

Expectations:	2 years		10 years		30 years	
	(1)	(2)	(3)	(4)	(5)	(6)
Average Yearly Price Trend	0.4294***	0.4175***	0.7380***	0.4185	1.4598***	0.9831
	(0.0761)	(0.1516)	(0.1932)	(0.3867)	(0.4954)	(0.8145)
Compounded Yearly Effect	0.1873	0.1817	0.0632	0.0353	0.0249	0.0126
Year FX	No	Yes	No	Yes	No	Yes
Commuting Zone FX	No	Yes	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes	No	Yes
Observations	2670	2534	751	710	2063	1972

Notes: Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Standard errors clustered at individual level. Source: SOEP Innovation Sample 2016-2018, own calculations. All of the above regressions control for individuals receiving RCT in 2017 and 2018 wave. Price trends are derived at the postal code level, which are smaller regional units than the commuting zones used as fixed effects. The compounded yearly effects are calculated as the difference between the compounded annual growth rate (derived from the mean values of each expectation question) and the compounded annual growth rates derived from the mean value of each expectation question in addition to the corresponding point estimate of the regression.

0.4185 in columns 4 (ten years) and 0.9831 in column 6 (thirty years), but start loosing significance. However, given the higher overall expectations for the longer periods (see descriptive results in Table 4.1), this increase in total expectations appears rather small.

To ease comparison and interpretation we convert each estimate into the corresponding compounded yearly elasticity (see Table 4.2).[8] The two-year estimate from column 2 corresponds to a yearly elasticity of 0.1817. This is remarkably close the the experimental estimate of 0.2 provided by (Armona et al., 2019), which they estimate by randomly providing information on last year's local price changes in combination with measuring the updating of knowledge about past growth and one-year ahead expectations. In contrast, the longer-term compounded yearly elasticities show less evidence for such momentum effects. The elasticity is at 0.0353 in column 4 (ten years) and only at 0.0126 in column 6

[8]The elasticity in this context then measures the percentage point change in house price expectations in response to a one percentage point change in the postal code specific hedonic average house price growth rate.

(thirty years). This means that current local price changes, measured as average annual changes over the six previous years, hardly affect expectations for long-run price growth. This rejects momentum effects. In other words, individuals in hot and cold markets have similar long-run expectations.

The Role of Individual Characteristics (H2) Table 4.3 depicts regressions of the different house price expectation outcomes on a range of individual socioeconomic variables frequently suggested in the literature to be central for the formation of price expectations (see Section 4.2). Table 4.3 depicts the regression results for the different expectation outcomes on a female dichotomous variable, a college indicator variable, and a numeracy dummy. The numeracy dummy indicates whether individuals would classify their math skills as rather proficient. Further, we include three variables testing financial literacy. More precisely, we include indicator variables that are equal to one if an individual answered the respective financial literacy question (see Section 4.3.2.1) correctly and zero otherwise. Depending on the specification, we include the potential explanatory variables without controls or we include the full set of control variables and the commuting zone and year fixed effects.

Individuals with a college degree, on average, have significantly lower short-run expectations of about 1.69 percentage points when including the full set of controls. Conversely, women appear to be less optimistic in the medium- to long-run, with significantly lower expectations over the ten- and thirty-year horizon. The self-assessed numeracy skill is significant over the medium- and long-run. The financial literacy indicator for the compound interest question, shows a slightly significant point estimate in the short and medium term, indicating that individuals who grasp the concept of compound interest have slightly lower expectations in the short term. The estimate, however, is only significant at the ten percent level.

Our findings overall are in line with previous findings that document more pessimistic expectations amongst women. Moreover, there is limited evidence that individuals with a higher degree of education, as measured by the college indicator, the self-assessed numeracy skills and the financial literacy measures, expect lower house price growth in the short-term and higher increases in the long-term. Surprisingly, only one of the financial literacy indicators is significant at the ten percent level, indicating that individuals who answered the compounding interest question correctly expect prices to increase less strongly over the 2 and 10 year horizon. The results support the hypothesis that, to a limited extent, education, financial literacy, and numeracy play a role in shaping expectations, which is consistent with Kindermann et al. (2021).

Table 4.3: The Role of Individual Characteristics

Expectations:	2 years		10 years		30 years	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.1506	-0.1902	-3.8077***	-2.1976	-11.7928***	-11.5719***
	(0.5284)	(0.5422)	(1.4500)	(1.4394)	(2.7661)	(3.2071)
Numeracy	0.0068	0.1705	3.8800***	4.0593***	3.4311	4.3474*
	(0.5372)	(0.5259)	(1.4257)	(1.4618)	(2.5964)	(2.5356)
College	-1.2867**	-1.6920**	0.9081	-2.3375	3.8324	1.5321
	(0.5863)	(0.6641)	(2.5233)	(2.0912)	(5.2797)	(5.9815)
Fin. Lit.: Inflation	-0.3157	-0.3548	-0.3452	2.8756	1.8017	0.9288
	(0.7641)	(0.9320)	(2.7511)	(2.8488)	(4.4510)	(5.3990)
Fin. Lit.: Investment	0.2668	0.3702	-0.6958	0.8661	4.1244	1.9133
	(0.5951)	(0.6056)	(1.8898)	(1.7163)	(2.6989)	(2.7330)
Fin. Lit.: Interest	-1.5737*	-1.5369*	-2.5334	-4.1532*	4.7487	2.9486
	(0.9040)	(0.8939)	(2.2850)	(2.3324)	(2.9935)	(3.0889)
Year FX	No	Yes	No	Yes	No	Yes
Commuting Zone FX	No	Yes	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes	No	Yes
Observations	2677	2541	752	711	2069	1978

Notes: Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Standard errors clustered at individual level. Source: SOEP Innovation Sample 2016-2018, own calculations. All of the above regressions control for individuals receiving the information treatment in the 2017 and 2018 wave. Each variable related to financial literacy (here abbreviated as Fin. Lit.) is an indicator variable that is equal to one if an individual answered the respective question about inflation, investment or interest rates correctly.

The Role of Housing and Tenure (H3) In Table 4.4, we focus on the effect of housing and tenure features and regress the respective expectation outcomes on an indicator variable for ownership-occupation and an indicator variable whether the monthly rental payment was increased during the last four years.

Table 4.4: The Role of Tenure and Housing Characteristics

Expectations:	2 years		10 years		30 years	
	(1)	(2)	(3)	(4)	(5)	(6)
Owner-Occupier	-0.8938	-0.3870	-1.5984	-0.7035	0.1486	1.1623
	(0.6527)	(0.6301)	(1.6055)	(1.8613)	(2.6368)	(2.5875)
Rental Raise	1.3851*	0.1219	7.7947***	6.1226**	11.0273	6.6989
	(0.7282)	(0.8091)	(2.6497)	(2.9223)	(6.7435)	(6.2837)
Year FX	No	Yes	No	Yes	No	Yes
Commuting Zone FX	No	Yes	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes	No	Yes
Observations	2679	2543	754	713	2071	1980

Notes: Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; Standard errors clustered at individual level. Source: SOEP Innovation Sample 2016-2017, own calculations. All of the above regressions control for individuals receiving RCT in 2017 wave.

The tenure type, i.e. owning or renting, does not appear to play a role in explaining differences in expectations. The results are at odds with frequently documented endowment effects for different markets, including the housing market. For example, homeowners are shown to over-estimate the value of their homes (see Kahneman et al. (1990), Goodman et al. (1992), Bao and Gong (2016), Chan et al. (2016)). In light of this literature, we would expect a higher expected house price change by homeowners. However, the results in Table 4.4 show that there does not appear to be a significant difference in homeowners and tenants expectations regardless over which time frame. One possible explanation for this difference is that home-owners and renters might be more alike in other observable and unobservable characteristics in Germany where roughly half of the population owner-occupies, when compared to countries with very high owner-occupancy rates.

In turn, tenants who experienced a rental increase in the past four years, have higher expectations. However, when including the full set of covariates, only the point estimate for the ten-year horizon is statistically significant.

The Role of Local Housing Markets (H4) In a next step, we test if local housing market conditions affect future housing price expectations. In Table 4.5, we regress house price expectation on the postal code specific housing supply index constructed from satellite data.

Table 4.5: The Role of Local Housing Markets

Expectations:	2 years		10 years		30 years	
	(1)	(2)	(3)	(4)	(5)	(6)
Supply Index	-0.0475***	-0.0429*	-0.1149***	-0.0652	-0.2075***	-0.2300**
	(0.0111)	(0.0258)	(0.0267)	(0.0435)	(0.0536)	(0.0928)
Year FX	No	Yes	No	Yes	No	Yes
Commuting Zone FX	No	Yes	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes	No	Yes
Observations	2679	2543	754	713	2071	1980

Notes: Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Standard errors clustered at individual level. Source: SOEP Innovation Sample 2016-2018, own calculations. All of the above regressions control for individuals receiving RCT in 2017 and 2018 wave. Note that the supply index is derived at the postal code level. Postal codes are smaller regional units than the commuting zones used as fixed effects.

The supply index is significant in predicting short term expectations. When including controls and fixed effects the point estimate, however, only remains significant at the ten percent level indicating that a one percentage point increase in the share of available land reduces two-year ahead house price expectations by 0.04 percentage points. When looking at the 10 and 30 year house price expectations, the supply index gains in absolute size and significance, particularly over the long-term horizon of 30 years. The sizable and significant effect of the supply index is consistent with theories and empirical evidence assuming housing supply to be elastic in the long run, see e.g. Saiz (2010). In our context, an increase of the land supply index by one percentage points indicates more developable land in the respective ZIP code and hence a more elastic supply response in the housing market. Standard theory would therefore suggest lower house prices as a consequence of a more elastic supply response. This is confirmed by our results. Overall, the results suggest that individuals observe local market conditions and take them into consideration when forming expectations.

The Role of Information on Expectations (H5) In Panel A of Table 4.6, we report regression results of the expectation outcomes on the randomly assigned information treatment. The treatment showed aggregate and long-run price developments to a randomly selected group of survey participants. The results show that individuals in the treatment group, on average, expect stronger price increases in the short- and long-run. However, the coefficients of the treatment variable in the short-run specification are statistically insignificant with p-values above the 10 percent level of significance. The coefficient in the long-run specification, however, is significant at the ten percent level. When including the additional control variables and fixed effects, the point estimate decreases and is no longer significant. Overall, the results suggest that, on average, treated individuals approximately expect higher long-run price changes. Thus, in the long-run aggregate long-run information appears to matter when forming expectations about future house price developments, although not very much. The result confirms findings, see e.g. Bailey et al. (2018), that, in addition to local information, individuals also factor in more distant, aggregate information, such as pooled average OECD, long-run price developments used as an information treatment in our setting.

Again it is important to note that more individuals (1426 vs.1054) answered the expectation question over the shorter time frame than over the long horizon. In order to ensure that there is no systematic difference in the expectation and portfolio outcomes stated in response to the RCT between those who answered only one question and those who answered all questions, Table 4.A.1 in the Appendix repeats the estimations for individuals that answered both, the two- and thirty-year expectation question. The results point towards a slight increase in the point estimates for short and medium term expectations in comparison to the estimation with all individuals. The point estimates for the long term outcome remain similar.

Table 4.6: The Role of Information

	Panel A: Information and House Price Expectations					
	2 years		10 years		30 years	
	(1)	(2)	(3)	(4)	(5)	(6)
RCT	1.0048	1.1734	0.3634	0.7965	9.0660**	6.4394
	(0.8260)	(0.9096)	(1.4855)	(1.4928)	(3.9293)	(4.6905)
Year FX	No	Yes	No	Yes	No	Yes
Commuting Zone FX	No	Yes	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes	No	Yes
Observations	1426	1361	754	713	1054	1013

	Panel B: Information and Investment Decisions		
	(1)	(2)	(3)
RCT	3.7674	3.7954	1.9903
	(2.5137)	(2.4911)	(2.5158)
Year FX	No	Yes	Yes
Commuting Zone FX	No	No	Yes
Control Variables	No	Yes	Yes
Observations	751	723	723

Notes: Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01.
Panel A: Standard errors clustered at individual level. Source: SOEP Innovation Sample 2017-2018 for the 2 year and 30 year horizon, SOEP 2017 for the ten year horizon, own calculations.
Panel B: Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; ; Heteroskedasticity robust standard errors. Control group mean is 33.71 per cent for the portfolio choice question, i.e. more than 33 percent would be invested in housing. Source: SOEP Innovation Sample 2018, own calculations.

The Relevance of Expectations for stated Decisions (H6) We also use the RCT to test the relevance of expectations on a hypothetical portfolio investment question which was asked in 2018. The key finding is that individuals who received the information treatment invest relatively more in housing (Panel B in Table 4.6). However, the estimated coefficients in the first two specifications are only borderline significant with p-values just above the threshold of 0.10. The estimated coefficient loses in significance when including commuting zone fixed effects which might be due to the relatively small observation number and hence a low within commuting zone variation. Overall, our findings therefore provide slightly imprecise albeit causal evidence that long term expectations can matter for investment decision, thereby confirming previous findings by Kuchler and Zafar (2019).

4.5.3 Robustness Checks

In this section we provide robustness checks for all hypotheses tests: inference holds when clustering standard errors at the commuting zone level. Further, we shed light on whether the expectation outcomes can be interpreted in nominal or real terms and find that our results are robust to possible inflation considerations.

Clustering of Standard Errors In the main specification, we cluster standard errors at the individual level, as standard errors are likely to be correlated within multiple observations for the same individual. However, standard errors might also suffer from heteroskedasticity at larger regional clusters. Thus, we repeat our estimations for all expectation outcomes using standard errors clustered at the commuting zone level. Tables 4.A.3 to 4.A.6 in the Appendix show that the results hardly change across the different specifications and that the interpretation remains the same.

Nominal vs. Real Expectations As described above, individuals received the following text when asked about their expectations: *The following section focuses on your expectations of house price developments in your area. In your opinion: How will house* *prices develop in comparison to today?* Participants then had the option to state whether prices will fall, rise, remain the same, or not answering at all. Depending on their answer individuals were asked *By how much - in percent- do you think prices will fall/rise.* From this question alone it does not become clear whether individuals state their expectations in real or in nominal terms. Therefore as a follow-up in 2019 we asked whether individuals interpret the question in nominal or real terms. More precisely in an additional question participants were asked:

In the previous section we asked you several questions about your expectations of house price developments in your area. When answering the questions did you factor in general price developments, i.e. inflation? Please be aware that it is also correct if you did not.

The majority of participants (close to 60 percent) states that they did indeed factor in inflation whereas over 40 percent state that they did not.

When regressing house price expectations on an indicator whether individuals state that they factored in inflation, we do not find a significant difference in long term and short term expectations between both groups (see Table 4.7 below).

Table 4.7: Expectations and Inflation

Expectations:	2 years		30 years	
	(1)	(2)	(3)	(4)
Inflation Indicator	-0.0291	-0.9555	6.3316	6.8777
	(0.9196)	(1.0781)	(5.8150)	(7.7347)
Year FX	No	Yes	No	Yes
Commuting Zone FX	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes
Observations	619	596	445	430

Notes: Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; Standard errors clustered at individual level. Source: SOEP Innovation Sample 2019, own calculations.

All in all, there is suggestive evidence that individuals appear to state their expectations in real terms, however, a relatively large proportion, i.e. approximately 40 percent, state that they did not. Interestingly there is no significant difference in the expectations in the short- or long-run between those who self-report that they factored in inflation and those who did not. One potential explanation could be that individuals do not fully grasp the concept of inflation as well as how nominal and real price developments differ. Alternatively, individuals could expect relatively low general price appreciation, which potentially explains why expectations between both groups do not significantly differ.

4.6 Discussion and Conclusions

In this paper, we test six sets of hypotheses regarding price expectations over different time horizons. To do this, we develop survey questions to elicit information on how individuals form house price expectations, including novel questions regarding the long-run.

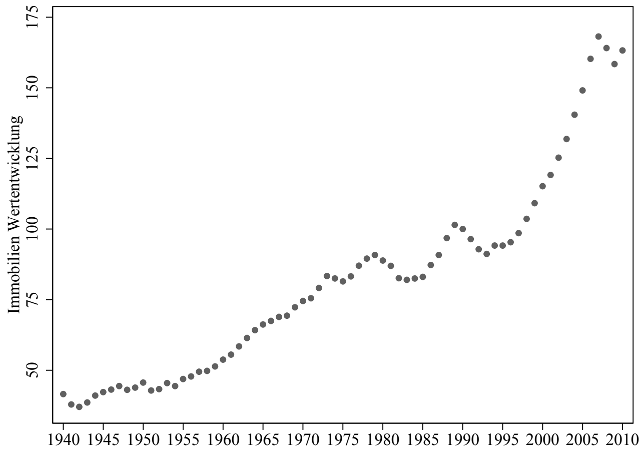
As our key finding, we show that individual house price expectations show no evidence of momentum effects in the long-run. Further, we show that most tenure decisions are not significantly correlated with future price expectations, whereas individual characteristics, such as gender and education, as well as local housing market experiences, play a small role. A land supply proxy constructed from satellite data is strongly correlated with (long-term) price expectations.

We also replicate a range of empirical findings from the literature, mostly from the US, in the context of Germany. This shows that Germans in the short–run behave similarly when forming house price expectations. All in all, our findings are generally consistent with empirically observed and theoretically formulated models of housing cycles that document short-run momentum effects and long-run mean reversion.

Appendices

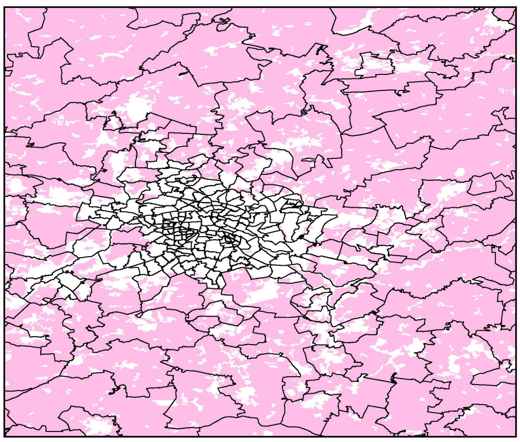
4.A Appendix A: Additional Figures and Tables

Figure 4.A.1: Real House Prices in the OECD



Note: Graph 4.A.1 is based on Knoll et al. (2017) and depicts the depicts the development of average prices in residential real estate in 14 different countries, including Germany and the United States. Prices on average approximately quadrupled since 1945.

Figure 4.A.2: Developable Land Index Berlin



Note: Source: Corine Land Cover database 2018, photointerpretation of satellite imagery. The pink area depicts the share of land in each postal code that is developable. Conversely, the white are is the developed share of land.

Table 4.A.1: The Role of Information: Only Obs. that answered 2/30 Year Questions

Panel A: Information and House Price Expectations

	2 years		10 years		30 years	
	(1)	(2)	(3)	(4)	(5)	(6)
RCT	1.8072*	1.9191	1.8678	2.1924	9.6453**	6.3102
	(1.0802)	(1.2553)	(1.6829)	(1.8470)	(4.1418)	(4.9725)
Year FX	No	Yes	No	Yes	No	Yes
Commuting Zone FX	No	Yes	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes	No	Yes
Observations	992	953	561	534	992	953

Panel B: Information and Investment Decisions

	(1)	(2)	(3)
RCT	2.2267	1.5777	3.1104
	(2.8995)	(2.8782)	(2.9559)
Year FX	No	Yes	Yes
Commuting Zone FX	No	No	Yes
Control Variables	No	Yes	Yes
Observations	423	411	411

Notes: Significance levels: ∗ p 0.10, ∗∗ p 0.05,∗ ∗ ∗ p 0.01. This table replicates Table 4.6, however only focuses on observations from individuals who answered both the 2 and 30 year house price expectation question. Note that the 10 year question was only asked once in 2017. The rates of response for the 2, 10 and 30 year expectations in the full sample are 79.67, 75.90 and 58.88 percent respectively. 94.23 percent of individuals answered the portfolio choice question. Panel A: Standard errors clustered at individual level. Source: SOEP Innovation Sample 2017-2018 for the 2 year and 30 year horizon, SOEP 2017 for the ten year horizon, own calculations. Panel B: Heteroskedasticity robust standard errors. Control group mean is 36.76 per cent for the portfolio choice question, i.e. more than 36 percent would be invested in real estate.Source: SOEP Innovation Sample 2018, own calculations.

Table 4.A.2: The Role of Local Housing Markets

Expectations:	2 years		10 years		30 years	
	(1)	(2)	(3)	(4)	(5)	(6)
Average Annual Price Trend	0.4294***	0.4175**	0.7380***	0.4185	1.4598**	0.9831
	(0.1175)	(0.1584)	(0.2641)	(0.4572)	(0.5910)	(0.9518)
Year FX	No	Yes	No	Yes	No	Yes
Commuting Zone FX	No	Yes	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes	No	Yes
Observations	2670	2534	751	710	2063	1972

Notes: Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; Standard errors clustered at commuting zone level.
Source: SOEP Innovation Sample 2016-2018, own calculations. All of the above regressions control for individuals receiving RCT in 2017 and 2018 wave.

Table 4.A.3: The Role of Individual Characteristics

Expectations:	2 years		10 years		30 years	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.1506	-0.1902	-3.8077**	-2.1976	-11.7928***	-11.5719***
	(0.5274)	(0.5688)	(1.6758)	(1.8815)	(3.1274)	(3.8205)
Numeracy	0.0068	0.1705	3.8800***	4.0593***	3.4311	4.3474*
	(0.5677)	(0.5624)	(1.0665)	(1.3463)	(2.4448)	(2.3649)
College	-1.2867**	-1.6920***	0.9081	-2.3375	3.8324	1.5321
	(0.5632)	(0.5928)	(2.4092)	(2.2036)	(5.5308)	(6.2514)
Fin. Lit.: Inflation	-0.3157	-0.3548	-0.3452	2.8756	1.8017	0.9288
	(0.9992)	(1.2017)	(3.2470)	(3.8871)	(5.0648)	(5.7925)
Fin. Lit.: Investment	0.2668	0.3702	-0.6958	0.8661	4.1244	1.9133
	(0.6919)	(0.7275)	(2.1221)	(2.1582)	(2.9642)	(2.6626)
Fin. Lit.: Interest	-1.5737*	-1.5369*	-2.5334	-4.1532*	4.7487	2.9486
	(0.8029)	(0.8492)	(2.5021)	(2.1985)	(2.9759)	(3.1078)
Year FX	No	Yes	No	Yes	No	No
Commuting Zone FX	No	Yes	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes	No	Yes
Observations	2677	2541	752	711	2069	1978

Notes: Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Standard errors clustered at commuting zone level.
Source: SOEP Innovation Sample 2016-2018, own calculations. All of the above regressions control for individuals receiving RCT in 2017 and 2018 wave.

Table 4.A.4: The Role of Tenure and Housing Characteristics

Expectations:	2 years		10 years		30 years	
	(1)	(2)	(3)	(4)	(5)	(6)
Owner-Occupier	-0.8938	-0.3870	-1.5984	-0.7035	0.1486	1.1623
	(0.7153)	(0.6365)	(1.7252)	(1.9179)	(2.7293)	(2.7461)
Rental Raise	1.3851**	0.1219	7.7947**	6.1226	11.0273*	6.6989
	(0.6732)	(0.8701)	(3.0052)	(3.7939)	(6.1878)	(5.4531)
Year FX	No	Yes	No	Yes	No	Yes
Commuting Zone FX	No	Yes	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes	No	Yes
Observations	2679	2543	754	713	2071	1980

Notes: Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; Standard errors clustered at commuting zone level.
Source: SOEP Innovation Sample 2016-2018, own calculations. All of the above regressions control for individuals receiving RCT in 2017 and 2018 wave.

Table 4.A.5: The Role of Local Housing Markets

Expectations:	2 years		10 years		30 years	
	(1)	(2)	(3)	(4)	(5)	(6)
Supply Index	-0.0475***	-0.0429**	-0.1149***	-0.0652	-0.2075***	-0.2300***
	(0.0102)	(0.0203)	(0.0246)	(0.0398)	(0.0491)	(0.0809)
Year FX	No	Yes	No	Yes	No	Yes
Commuting Zone FX	No	Yes	No	Yes	No	Yes
Control Variables	No	Yes	No	Yes	No	Yes
Observations	2679	2543	754	713	2071	1980

Notes: Significance levels: ∗ p 0.10, ∗∗ p 0.05, ∗ ∗ ∗ p 0.01; Standard errors clustered at commuting zone level.
Source: SOEP Innovation Sample 2016-2018, own calculations. All of the above regressions control for individuals receiving RCT in 2017 and 2018 wave. Note that the supply index is derived at the postal code level. Postal codes are smaller regional units than the commuting zones used as fixed effects.

Table 4.A.6: The Role of Information

Expectations:	2 years		10 years		30 years		
	(1)	(2)	(3)	(4)	(5)	(6)	
RCT	1.0048	1.1734	0.3634	0.7965	9.0660**	6.4394	
	(0.8260)	(0.9824)	(1.4855)	(1.3395)	(3.9293)	(4.7058)	
Year FX	No	Yes	No	Yes	No	Yes	
Commuting Zone FX	No	Yes	No	Yes	No	Yes	
Control Variables	No	Yes	No	Yes	No	Yes	
Observations		1426	1361	754	713	1054	1013

Notes: Significance levels: $*$ p 0.10, $**$ p 0.05, $***$ p 0.01; Standard errors clustered at commuting zone level.
Source: SOEP Innovation Sample 2016-2018, own calculations. All of the above regressions control for individuals receiving RCT in 2017 and 2018 wave.

Table 4.A.7: Summary Statistics: Housing Data

Variable	**Mean**	**Min.**	**Max.**	**N**
Price per sqm	2,283.943	35	182,03.881	172,328
Rental Price per sqm	7.772	1.02	66	309,079

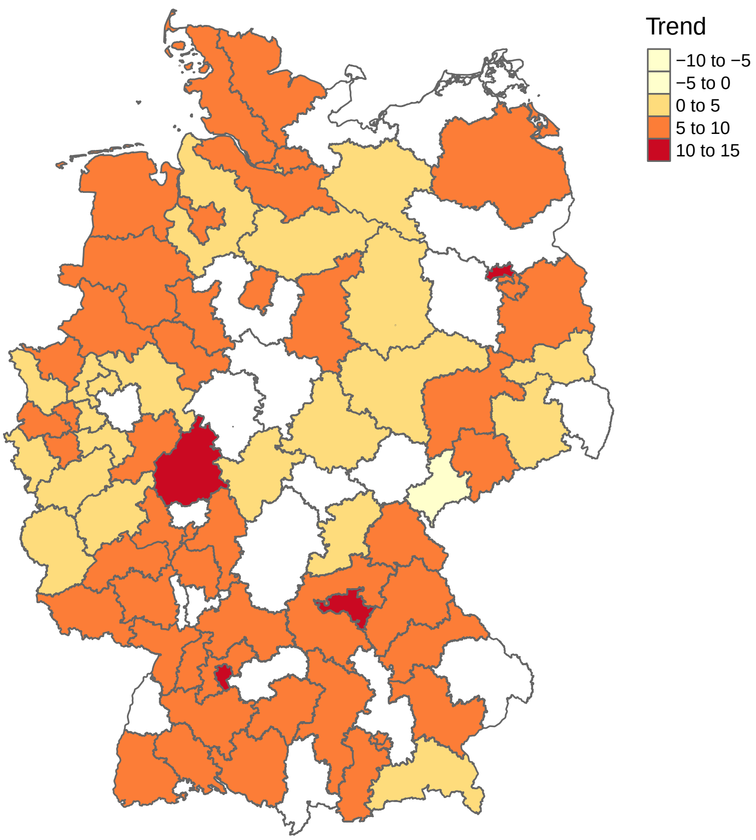
Note: Table 4.A.7 depicts summary statistics for advertised rental and purchase prices during the time period 2012-2016 in Germany. Source: *empirica ag*

4.B Appendix B: Calculation of Hedonic Rent Price Index

For the derivation of the hedonic rent price index we use real estate data provided by *empirica ag* for the years 2012-2016. The data is based on advertised house prices and includes both apartments and houses for sale and for rent. In our derivation of hedonic house price changes we focus on dwellings for sale to measure isolate house price changes. For the years 2012, we regress the logarithm of the advertised sales price per square metre on a range of housing and quality characteristics such as the number of rooms, the apartment size, the age of the dwelling, the existence of a balcony, garden or cellar and the type of building the dwelling is situated in. Further, we include year and postal code (see Figure 4.B.1) dummies and interactions between both. Based on these interaction terms we can then derive the quality adjusted house price changes from year to year.

For illustrative purposes Figure 4.B.1 shows the average yearly price within the postal code regions contained in our data.

Figure 4.B.1: Average Yearly Price Change: Postal Code Region



Note: Figure 4.B.1 depicts the average yearly price change between 2012-2016 across postal code regions in Germany. Postal code regions are characterized by the first two digits of each postal code. Source: empirica ag.

Chapter 5

Gentrification and Affordable Housing Policies.[1]

Abstract

We use a quantitative spatial equilibrium model to evaluate the distributional and welfare impacts of gentrification as well as policy responses such as the recent temporary rent control policy in Berlin, Germany. We calibrate the model to key features of Berlin's housing market. We find that gentrification benefits rich homeowners, while poor renter households lose. In a counterfactual simulation, we find that rent control reduces welfare for all households, but the percentage change in welfare is largest for the poorest households. We also study alternative affordable housing policies such as targeted housing subsidies and rezoning policies, which are better suited to address the adverse distributional consequences of gentrification.

[1] Co-authored with Rainald Borck.

5.1 Introduction

Over the last decades, cities in the developed world have seen enduring trends of immigration and gentrification, that is, the movement of higher income households into central cities. The ensuing increases in housing prices and rents have spurred the fear that poor households would be priced out of their neighborhoods and central city housing would become unaffordable for lower income households. Indeed, rising housing costs have led politicians to call housing “the new social issue of our time” (Sagner et al., 2020). Policy responses have included a range of affordable housing policies, including rent control, subsidized low-income housing and residential construction. The primary goal of these policies has been to relieve the burden of growing housing expenditures, especially on poorer households, and thus mitigate the distributional consequences of gentrification and rising housing costs.

In this paper, we use a quantitative spatial economic model to assess the impact of gentrification and several policy measures to counteract the distributional effects of gentrification on the welfare of heterogeneous households. In the model, households differ by income and sort into locations based on housing prices and amenities. Our paper is calibrated to key economic variables in Berlin, Germany, and does a good job at replicating the observed gentrification patterns over the last years. Our model has non-homothetic preferences over housing and (endogenous) amenities, which leads to movement of richer households into central city, high amenity neighbourhoods. We run a counterfactual which shows that gentrification benefits more affluent, homeowner households and hurts poorer renter households in the lower half of the income distribution.

We then run counterfactual simulations to assess the distributional impacts of several policy measures. One of these is based on the rental cap policy that was introduced (and later repealed following a ruling of the Constitutional Court) by the state government of Berlin. In essence, the policy set maximum rents for all existing and new rental contracts with the exception of newly built housing. Our simulation shows that this policy reduces welfare for all households. What is more, not only does the policy reduce welfare for all household types, but the welfare reduction is largest in relative terms for the poorest households. While rents fall for all households, they fall most for richer households who choose to live in expensive, high amenity locations mainly in the city centre, and it is there that rent control is binding. Rent control also increases segregation by making these locations even more attractive for high income households. The policy is thus regressive in the sense that it hurts poor households most, even though richer households bear some of its cost in the form of reduced rental income.

We also consider alternative measures, namely, a targeted housing subsidy for low income households and two rezoning policies which essentially increase housing construction (in one part only or the entire city). We find that both housing subsidies and rezoning increase welfare most for the poorest households. The subsidy by design redistributes from rich to poor households, but these effects are also partly amplified by a reduction in segregation. Rezoning increases supply and thus reduces rents for all households. The rezoning policies also lead to a slight reduction of segregation in the city and, depending on the spatial distribution of new housing, lead to a larger relative increase in welfare for poor households. This suggests that rent control may not only be inefficient, as has long been suggested by economic research but also that rent control may fail to help the poor households who suffer most from gentrification and rising housing costs. In part, the effects of rent control stem from sorting of households across locations, which amplifies segregation and differences in endogenous amenities to the detriment of poor households.

We contribute to the economic literature on gentrification and housing affordability, and in particular on rent control and other policies aimed to combat the adverse distributional effects of this gentrification. The combined trends of gentrification and rising house prices in developed cities have been studied extensively over the last years, mostly for urban America. Hwang and Lin (2016), Couture and Handbury (2020) and others have documented the increase in the number of college educated households in American cities. Baum-Snow and Hartley (2020) and Couture and Handbury (2020) show that these trends can be explained by demand for central city amenities. A central concern of analysts and advocates has been the potentially harmful effect of gentrification on poor residents' welfare. A number of papers have examined this issue empirically. For instance, Vigdor (2002) documents that low-status households experienced increased housing costs without discernible changes in self-assessed housing unit quality, public service quality, or neighborhood quality. In line with this, Couture et al. (2019) quantify a spatial sorting model, which shows how increasing incomes at the top of the distribution lead to gentrification due to the demand of rich households for central city amenities. This hurts poor residents who are either displaced or end up paying higher rents for amenities they do not value highly. Brummet and Reed (2019), in turn, document positive effects of gentrification on a range of subjective well-being measures and school outcomes of original residents and their children.[2]

In recent years, policy makers across the globe have responded to gentrification with a range of affordable housing policies, such as rent control, tax credits, housing subsidies or rezoning policies. In the United States rent control measures are a strongly debated topic in local ballots, particularly in larger agglomerations such as the San Francisco Bay area, New York, New Jersey or Maryland and have been in place in some areas for several decades (Diamond et al., 2019). In Europe, rent control has become an equally topical issue. Many countries such as Spain, the Netherlands or Germany have highly protected rental sectors and have recently increased rental protection by introducing additional rent control measures in areas with tight housing markets. A large economic literature has analyzed the economic impacts of rent control. Findings include a negative impact on tenants' mobility (Diamond et al., 2019), misallocation in housing markets (Glaeser and Luttmer, 2003), price appreciation in decontrolled housing market segments (Autor et al., 2014) as well as a negative impact on housing investment and residential construction in the controlled segments of the market. Hahn et al. (2022) and Sagner and Voigtländer (2022) present reduced-form evidence on the short term impact of the rental cap introduced in Berlin, which we consider. They find a decrease in the number of advertised dwellings as well as a decrease in rental prices directly after the rental cap

[2]Chetty and Henderson (2018a) more broadly document that exposure to relatively rich or poor neighbourhoods during childhood years plays an important role for inter-generational mobility and later labour market outcomes such as earnings and human capital accumulation.

was enacted.[3] Favilukis et al. (2022) build a quantitative dynamic model with incomplete risk sharing and show that affordable housing policies (especially targeted at the poor) can increase welfare by providing insurance against housing price increases. Other recent papers such as Mense et al. (2019) provide evidence of an increase in construction activity and investment in the unregulated segments of the market.

Our paper contributes to the literature in several ways. First, we add to the quickly developing strand of research that examines and quantifies patterns and causes of gentrification and spatial sorting, see e.g. Brummet and Reed (2019); Couture and Handbury (2020); Baum-Snow and Hartley (2020); Couture et al. (2019). While much of the literature studies US cities, we explicitly model gentrification patterns in Berlin, Germany. Studying a European city may be interesting since the location patterns of residents as well as housing markets, zoning policies, transport etc. substantially differ between European and American cities.[4] Like other papers before us, we find that gentrification benefits rich homeowners while hurting poor renters. Second, our most important contribution is to the literature on affordable housing polices. We do so by providing counterfactual long-run evidence on a range of affordable housing policies such as rent control, subsidies and rezoning policies. In particular, we provide novel counterfactual long-term evidence on the impact of a unique rental cap temporarily enacted in Berlin. This policy differs for instance, from the rent stabilization policies studied by Favilukis et al. (2022) and the taxation of high quality housing in Couture et al. (2019). We stress that in order to understand the redistributive effects of gentrification and housing policies, modeling spatial sorting of households is necessary. For example, the "traditional" view of rent control argues that rent control benefits renter households in controlled dwellings at the expense of landowners (Arnott, 1995). However, if poor households sort into locations where rent control does not bind, they may not benefit from rent control, even if they do not own housing. Our sorting model is similar to but spatially more explicit than those in Couture et al. (2019) and Favilukis et al. (2022), who analyze stylised cities consisting

[3] See also Jofre Monseny et al. (2022) and Kholodilin et al. (2022) on a similar rent control policy in Barcelona, Spain.

[4] See for example Nivola (1999) for a discussion of sprawl in the context of differences between European and American cities. For a related application to the Dutch city of Amsterdam see Almagro and Dominguez-Iino (2021).

of a downtown and a suburb.[5]

We find that the Berlin rental control is not suitable to address the adverse consequences of gentrification and other policies such as targeted housing subsidies or rezoning perform better. Our findings contribute to the quantitative evaluation of gentrification and affordable housing policies. Unlike the empirical "reduced form" literature, our quantitative model is among the first to evaluate the long run equilibrium effects of gentrification, rent control and other housing policies. Finally, whereas early literature has discussed the efficiency effects of rent control, we contribute to the literature about the redistributive effects of this as well as other policies.

The remainder of the paper is structured as follows. Section 5.2 presents the model and its theoretical foundations. Section 5.3 gives a brief overview of the German housing market and its institutional setting. Section 5.4 presents the data and calibration of the model. Section 5.5 and 5.6 present a range of counterfactual simulations as well as corresponding model extensions and sensitivity checks, while Section 5.7 concludes.

5.2 The Model

We develop a discrete choice model of a city made up of several neighborhoods, where households are heterogeneous in terms of income and preferences.

There are K discrete regions or parts of the city (ZIP codes in our example) indexed $k = 1, ..., K$. Households differ in their income, indexed by $j = 1, ..., J$, and their preference, where i indexes the household's preference type. A household of type i, j who lives in part k of the city has a utility function of the Stone-Geary type,

$$u_{ik}^j = (A_k^j)^\beta (q_{ik}^j - q_0)^\alpha (c_{ik}^j)^{1-\alpha-\beta} \eta_{ik}^j, \quad q_0 > 0, \tag{5.1}$$

[5] In their main analysis, Couture et al. (2019) use the average of 27 CBSAs as a representative city which has a downtown and suburban area; Favilukis et al. (2022) study the downtown and suburban area of the New York MSA.

where q is housing consumption in square metres and q_0 is minimum floorspace, c is consumption of a composite good, A is a composite amenity index, and η^j_{ik} is household i, j's idiosyncratic taste parameter, which governs its preference for living in part k of the city. The composite amenity index is given by

$$A^j_k = Z_k (n^H_k)^{\zeta^j}, \tag{5.2}$$

where Z is a measure of amenities such as bars, restaurants and shops in region k. Further, households may benefit from living together with high income households, see e.g. Couture et al. (2019), Diamond (2016) and Guerrieri et al. (2013). Our model captures this endogenous amenity via the term $(n^H_k)^{\zeta^H}$, where n^H_k is the total number of households of the higher income groups living in region k. In the baseline model, we assume that only high-income households benefit from endogenous amenities, so $\zeta^j = 0$ for all lower-income types, but in an extension, we assume that there is also a cross-type externality, which implies that low-income households also benefit from mixing with high-income types, for example because of positive neighborhood externalities in schooling or social capital (see Section 5.6.4).

The household's budget constraint is

$$w^j + a^j R = c^j_{ik} + p_k q^j_{ik}, \tag{5.3}$$

where w^j is the household's labour income, p_k is the housing rent per square meter, R is the average aggregate housing rent per household in the city and a^j is the share received by a type-j household, i.e. the home-ownership rate for that household type.

Maximizing utility subject to (5.3) gives optimal housing consumption and indirect utility, v:

$$\begin{aligned} q^j_k &= \frac{(1-\alpha-\beta)p_k q_0 + \alpha(w^j + a^j R)}{(1-\beta)p_k} && (5.4) \\ v^j_{ik} &= (A^j_k)^\beta (w^j + a^j R - p_k q_0)^{1-\beta} p_k^{-\alpha} \eta^j_{ik} && (5.5) \end{aligned}$$

The Stone-Geary utility function has several attractive properties. First, the budget shares for housing,

$$\frac{p_k q_k^j}{w^j + a^j R} = \frac{(1-\alpha-\beta)p_k q_0 + \alpha(w^j + a^j R)}{(1-\beta)(w^j + a^j R)} \tag{5.6}$$

are decreasing in income. This is in line with observed data (see Table 5.4 below).

Second, the marginal willingness to pay for amenities

$$\begin{aligned} \frac{dp_k^j}{dA_k^j} &= \frac{\partial u_{ik}^j / \partial A_k^j}{q_{ik}^j \partial u_{ik}^j / \partial c_{ik}^j} && (5.7) \\ &= \frac{\beta p_k (w^j + a^j R - p_k q_0)}{A_k^j((1-\alpha-\beta)p_k q_0 + \alpha(w^j + a^j R))} && (5.8) \end{aligned}$$

is increasing in income. The Stone-Geary function thus captures a key aspect of gentrification as the demand for inner city amenities has been identified as an important driver of gentrification (Brueckner et al., 1999; Couture et al., 2019).

In the spirit of the discrete choice literature, individuals have heterogeneous tastes for which part of the city to live in. In line with recent quantitative models, an individual's idiosyncratic taste parameter η_{ik}^j for living in part k of the city is drawn from an independent Fréchet distribution

$$G(\eta_{ik}) = e^{B_k^j (\eta_{ik}^j)^{-\epsilon^j}} \tag{5.9}$$

where the scale parameter B_k^j gives the average utility of living in part k of the city for individuals of type j, and the shape parameter $\epsilon^j > 1$ controls the dispersion of idiosyncratic utility for each household type. After observing their idiosyncratic taste parameter, individuals choose their place of residence to maximize utility, for given attributes of the locations. Note that the variance of idiosyncratic tastes decreases with ϵ^j, which implies that individuals become more responsive to policy changes when ϵ^j increases.

The choice probabilities conditional on household type for living in part k of the city are given by

$$\pi_k^j = \frac{B_k^j ((A_k^j)^\beta (w^j + a^j R - p_k q_0)^{1-\beta} p_k^{-\alpha})^{\epsilon^j}}{\sum_{\ell=1}^K B_\ell^j ((A_\ell^j)^\beta (w^j + a^j R - p_\ell q_0)^{1-\beta} p_\ell^{-\alpha})^{\epsilon^\ell}}, \quad k, \ell = 1, ..., K. \tag{5.10}$$

Housing supply in part k of the city is produced by profit maximizing housing firms under perfect competition. We assume a reduced form profit function[6]

$$\Pi_k = \Theta_k^{\frac{1}{\theta}} p_k Q_k - \frac{\theta}{1+\theta} Q_k^{\frac{1+\theta}{\theta}} - R_k \tag{5.11}$$

where Q_k is housing quantity supplied, θ is the housing supply elasticity and Θ an underlying profitability parameter of region k. Producers have costs $\frac{\theta}{1+\theta} Q_k^{\frac{1+\theta}{\theta}}$ and pay land rent R_k. Solving the profit maximizing problem and using the zero profit condition $\Pi_k = 0$ gives housing supply and land rent

$$Q_k = \Theta_k p_k^{\theta} \tag{5.12}$$

$$R_k = \frac{1}{1+\theta} \Theta_k^{\frac{1+\theta}{\theta}} p_k^{1+\theta}, \tag{5.13}$$

with the average land rent

$$R = \frac{1}{N} \sum_{k=1}^{K} R_k. \tag{5.14}$$

The housing market clearing conditions are

$$Q_k = \sum_{j=1}^{J} n_k^j q_k^j, \quad k = 1, ..., K, \tag{5.15}$$

where n_k^j is the number of type j residents in part k of the city. Using (5.4), the set of equations (5.15) define the equilibrium housing prices in regions $k = 1, ..., K$.

Total city population of type j is exogenous and given by N^j. To close the model, the location equilibrium is defined by the following equations:

$$n_k^j = \pi_k^j N^j, \quad j = 1, ..., J, k = 1, ..., K-1. \tag{5.16}$$

Given (5.5), (5.10), and (5.15), the equilibrium is defined by the $J \times (K-1)$ equations

[6] This profit function can be derived by assuming, as in Ahlfeldt et al. (2015), that firms combine capital and land to produce housing floor space.

in (5.16). This pins down the number of individuals in each part of the city.

In order to compute the welfare effects of housing policies, in the counterfactual simulations, we will compute the expected welfare of a type j resident

$$\mathbb{E}(u^j) = \Gamma\left(\frac{\epsilon^j - 1}{\epsilon^j}\right)\left[\sum_{k=1}^{K} B_k (v_k^j)^{\epsilon^j}\right]^{1/\epsilon^j} \tag{5.17}$$

where $\Gamma(\cdot)$ is the gamma function, and the expectation is taken over η_{ik}^j.

In order to calibrate the model we use a range of small scale housing and socioeconomic data for the German capital, Berlin. Before describing the data and the calibration in more detail, however, the following section will give a brief overview of the German housing market, with a particular focus on Berlin.

5.3 Institutional Background and the German Housing Market

In comparison with most OECD countries, the German housing market has a relatively large rental sector, which accounts for approximately half of all dwellings (Voigtlaender, 2010). In large cities the rental sector is even larger, with a home-ownership rate of only 17 per cent in Berlin (Bundesamt für Statistik, 2019).

The corner stone of rent setting according to federal government regulations is the so called "*Mietpreisspiegel*" or "rental price barometer". The "*Mietpreisspiegel*" is essentially a survey of characteristic regional rents conducted or recognised by the municipality or by tenants' and landlords' associations. It serves to derive a comparable local reference rent and is legally required to be updated every two years (Kholodilin et al., 2016, p.8). The local reference rent is computed as an average of new and existing contract rents concluded in the previous four years. This reference rate sets the maximum limit for rent increases within existing contracts.[7]

[7]If the calculation of the "*Mietpreisspiegel*" is not possible, the local reference rate is derived by an

There is strong regional variation within the German housing market. In rural and suburban areas, especially in eastern Germany, the housing market has been relatively stable. In contrast, in urban agglomerations and college towns, strong demand and supply shortages have dominated local housing markets, putting upward pressure on housing prices (Kholodilin et al., 2016). However, even within big cities, substantial variation in local housing markets exists. While some districts are characterised by high levels of housing demand and supply shortages, other more peripheral areas show relatively stable housing markets (BMF, 2017, p.5). The regional differences in the German housing market are also reflected in rent and housing price developments. Particularly since 2007/2008, housing prices have strongly increased in popular urban areas such as the Big Seven cities[8] and stagnated or only increased marginally in many rural areas.[9]

Rent changes have been particularly pronounced in the capital city Berlin, which is the focus of our quantitative exercise below. In Berlin, quality-adjusted nominal rental prices of new rental leases have almost doubled between 2007 and 2020, as shown in Figure 5.A.1 in the Appendix.[10] As a consequence of quickly rising housing prices, affordable housing, especially for low- and middle income households, has become one of the most salient political topics in Berlin. In 2020 the state government of Berlin passed a rental cap for new rental contracts as well as existing rental contracts, effectively setting maximum permissible rents for most segments of the city's rental market, in the hope to address concerns about rising housing costs.[11] However, the rental cap was declared unconstitutional by Germany's Constitutional Court in April 2021 and hence was only in place for less than a year.

In our simulation, we will take Berlin as a case study to analyse the effects of gentrification on the distribution of welfare across households and the effect of various policies intended

expert's reported estimate or by rents in at least three comparable dwellings owned by other landlords (Kholodilin et al., 2016).

[8]The "Big Seven" are: Berlin, Hamburg, Munich, Cologne, Duesseldorf, Frankfurt am Main and Stuttgart.

[9]In the last couple of years, the coronavirus pandemic and the onset of an economic downturn have led to a slowing of price increases in the major cities.

[10]Note that over the same time period, the average annual inflation rate was relatively low with 1.29%. Thus real rental prices also increased substantially over the same period (see Federal Office for Statistics).

[11]Dwellings built from 2014 on are exempt from the rental cap. See Section 5.5.3 for more detail.

to mitigate the rise in housing prices and rents that comes with this gentrification. In particular, we will study whether a rent control policy modelled along the lines of the enacted one can effectively address the problem of housing affordability for lower income households.

5.4 Data and Calibration

5.4.1 Data

In order to calibrate the model, we need small scale regional information on current numbers of households, rental prices, income distribution and amenities for Berlin. We employ different data sets to collect the required data.

We use the GFK Demographic data set which provides information on the number of households and their approximate net household wage income across the ZIP codes of Berlin. In total this allows to differentiate between seven income types. The income groups and the relative size of each group for the city of Berlin are depicted in Table 5.3. For the calibration of the model we use the midpoint of each income bracket.[12] Table 5.3 shows that more than half of all households living in Berlin belong to the lowest three income brackets and more than thirty per cent of all households are part of the highest three income groups.

Table 5.1: Household Distribution across Income Groups 2019

Type	1	2	3	4	5	6	7
Net Income	850	1300	1750	2300	3300	5750	9000
Share of HH	19.5	14.7	17.8	14.9	18.3	12.9	1.9

Note: Table 5.3 shows the household distribution by income group and depicts the share of households by income group in percentage terms. Source: GFK Demographic Data 2019

Figure 5.B.1 in the Appendix depicts the spatial distribution of all households types

[12]For the highest and lowest income bracket we use values of 9000 Euro and 850 Euro which are the median income values of households above/below the respective threshold derived from the GSOEP data.

of the lower three income groups across Berlin's ZIP code areas. In particular, in the richer districts of Berlin in the southwest, households of the lower three income brackets constitute the minority of the total ZIP code population. In contrast, in the city centre the shares of low and high income type households are fairly similar. Districts such as Neukoelln, parts of Tempelhof, Wedding, Spandau and Marzahn and Hellersdorf in turn have a relatively high proportion of households in the lower three income brackets.

For each ZIP code, we then calculate an amenity index value using Open Street Map data. We do so by collecting the number of available amenities, such as restaurants, bars, cinemas, nightclubs, ATMs, shops and parks in each ZIP code. Using these values we then construct an index employing principal component analysis.[13] Figure 5.B.2 in the Appendix shows that as expected high amenity levels are predominantly found in the central districts of Berlin.

Lastly, we need information on average rents within each cell. Most studies analysing housing prices and rent control measures focus on transaction or advertised rents and hence capture only new rental contracts. However, stricter protection of existing rental contracts in Germany implies that there is a fundamental wedge between new and existing contract prices. Therefore, exclusively focusing on transaction prices would bias price estimates upwards and would hence not reflect the actual average rental prices within each ZIP code. In order to obtain rent estimates that reflect prices of existing rental contracts as well as new lease prices, we therefore make use of the aforementioned rental barometer and geo-coded data. The rental barometer separates Berlin into different residential area classifications, differentiating between "simple", "medium" and "good" residential areas. For each category, the rental barometer then lists different square metre rents. The rental barometer differentiates further between the dwelling's overall floor space and the dwelling's age and thus provides regional estimates for different housing categories. We make use of openly available geo-data to obtain information on buildings' age, average square metres per dwelling and the residential classification across Berlin.[14] Using this information, we can interpolate average values for each ZIP code across Berlin.

[13]For more detail see Appendix A.
[14]See FIS-Broker, an online catalogue for Berlin-specific geo-data.

We thus obtain estimates of average floor space size, dwelling age and residential area classification for each ZIP code. Using the interpolated averages and the metrics of the 2019 rental price barometer, we construct estimates of average rental price levels for each ZIP code. Figure 5.B.3 in the Appendix shows that rents per sq. meter are higher in central and western Berlin, whereas rents particularly in many less central, eastern parts of Berlin are relatively low.[15] Still, our estimates should be more accurate than those relying on new contract rents only.

5.4.2 Calibrating the Model

Using the price and amenity index estimates for each ZIP code as well as the share of each of the seven household types living in each ZIP code, we can calibrate the model as follows.

For the composite amenity index A_k, we use two data components as described above. First, we use the index calculated from the Open Street Map data, Z_k. Second, we account for endogenous amenities accruing to high-income types, as shown in (5.2). For households of the highest four income groups, we assume that the value of amenities is re-enforced by the number of other high income type households living in ZIP code k. We set the parameter governing the strength of endogenous amenities to $\zeta^j = 0.01$ for the four highest income types, and for the three lowest income types, we set $\zeta^j = 0$. However, since not much is known about the "true" value of this parameter, we will vary it in our sensitivity analysis in Section 5.6.2. In Section 5.6.4, we further assume that endogenous amenities also accrue to low-income households, which potentially impacts the efficiency effects of gentrification and rent control policies.

[15]Note that these estimates are only approximations of the "true" rental prices in each post code. In order to obtain the actual average rental prices of each ZIP code we would require access to very detailed data on existing rental contracts or large scale survey data for a large enough number of households in each ZIP code, which is not available. Note also note that our approximation of rental prices might be biased when a zip code contains different quality residential areas and the share of households living in these areas differs widely. For instance, some zip codes have a mix of medium and simple residential quality, but many more households live in the simple areas. Since we don't know how many households live in which area, our estimates would be biased in this case.

Next, we obtain values for the parameters q_0, a^j, α, β and Θ as follows. First, we infer the home-ownership rate for each income group in Berlin, i.e. a^j, from the German Socioeconomic Panel (GSOEP) for the year 2018. The GSOEP is an annual representative household survey with more than 18,000 households. We use the data for households living in Berlin, but in order to have sufficient observations we include Hamburg and Bremen (the two other federal "city states" in Germany) and compute the average home-ownership rate for each income group. Results are shown in Table 5.2. Berlin has a comparatively low ownership rate with only 17.4 per cent owning property (Investitionsbank, 2019). Based on the GSOEP data we find a slightly lower average home-ownership rate of approximately 16 percent. The table shows that homeownership increases from around 10 percent or less for the lowest four income groups to 75% for the highest income earners.

Table 5.2: Home-Ownership across Income Brackets

Type	1	2	3	4	5	6	7
Income Bracket	850	1300	1750	2300	3300	5750	9000
a	0.0138	0.0919	0.1176	0.1090	0.2182	0.4051	0.7568

Table 5.2 depicts the average home-ownership rate by income bracket. Estimates are based on the German Socioeconomic Panel 2018.

For q_0 we assume a value of 30 square metres, which we then use to estimate β and α in a method of moments approach to target the average housing expenditure share in net household incomes across the city.

We target an average expenditure share for housing of 28.5 per cent, see Investitionsbank (2019). Choosing α and β to minimize the difference between this expenditure share and the modelled expenditure share, we obtain values of $\beta = 0.215$ and $\alpha = 0.1229$. While the expenditure share is 28.5% in the data and our model on average, it varies from 17.73 % for the richest households to 38.73% for the poorest ones in our model calibration (see Tab. 5.4 below). Table 5.3 provides an overview of our model parameters.

We then numerically solve for the Θ_k's using equation (5.15). For θ, the housing supply elasticity, we base our calibration on structural estimates for Germany calculated in Mense (2020). Mense (2020) finds a nation-wide housing supply estimate of 5.8 in the medium run between 2011 and 2017. This estimate is in line with Green et al. (2005)'s findings

Table 5.3: Parameter Overview

Parameter	Value	Source
Demand Parameters		
q_0	30	by assumption
α	0.1229	calibrated using housing expenditure shares
β	0.2150	calibrated using housing expenditure shares
Scale Parameter		
B_k^j	by ZIP code	calibrated
Shape Parameter		
ϵ^j	3.4 – 6.5 depending on type; see Table 5.A.3 in Appendix A	estimated, see Appendix A
Endogenous Amenities		
ζ^j	0.01 for high income households; 0 for low income households	alternative values in robustness check
Supply Parameters		
Elasticity		
θ	3.5	literature; see Mense et al. (2019); Green et al. (2005)
Productivity Parameter		
Θ	by ZIP code	calibrated

Note: Table 5.3 gives and overview of the model parameters and briefly describes how they are obtained.

for the metropolitan areas in the US but lies above Saiz (2010) who finds an average elasticity of 1.75 of urban centres in the US. Since the estimates in Mense (2020) contain results for all of Germany, and large cities typically have lower supply elasticities, we use a slightly lower value of $\theta = 3.5$, which lies between the averages from Mense (2020) and Saiz (2010).

Following the literature, we can use Equation (5.10) to estimate the Fréchet parameters, ϵ^J. Since these vary by income type, we estimate seven separate regressions for each of them. In order to address potential endogeneity concerns, we estimate instrumental variable regressions, as described in more detail in Appendix A. As shown in Table 5.A.3, the estimates by type range from 1.75 to 4.79, which seems in line with other recent estimates such as Monte et al. (2018), Ahlfeldt et al. (2015) and Heblich et al. (2020).

Finally, we solve the model numerically. The model is exactly identified, so given the distribution of types by ZIP codes and the values of the parameters described, we can solve the location equilibrium in (5.16) numerically for the level parameters, B^j_k. That is, by construction, our model perfectly replicates the distribution of types for each ZIP code.[16]

5.5 Results

In this section, we first present our baseline results and then run a range of counterfactuals to compare the results of our baseline calibration to key outcomes of the counterfactual simulations. More precisely, we compare the developments of housing prices, floor space consumption and rental income in our counterfactual to their baseline values. Additionally, we use Equation (5.17) to calculate the average expected welfare for each household type. Using the expected welfare values we can then calculate the corresponding equivalent variation (EV) denoted in Euros for each household group and each counterfactual.

[16]For a proof of existence and uniqueness in a similar model, see e.g. Monte et al. (2018).

5.5.1 Baseline Results

Table 5.4: Results Basic Calibration

	Type 1	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
Budget Share (%)	38.73	30.04	26.37	23.93	21.35	18.95	17.73
Price (EUR per sq. m)	7.86	7.85	7.85	7.84	7.86	7.96	7.93
Floor Space (sq. m)	42.56	52.98	62.47	73.46	95.36	146.37	217.88
Rental Income (EUR)	12.14	80.88	103.49	95.93	192.03	356.51	666.02
Amenity	21.43	20.21	19.96	29.92	29.78	30.92	29.34

Source: Own model calculations. Budget share expressed in percentage terms. Prices denoted in Euros. Floor space denoted in sqm. Rental income is denoted in Euros and amenity denotes the average amenity index value for each household type.

Table 5.4 depicts the results of our baseline calibration. Floor space consumption increases with income and on average approximately equals 78 square metres per household. Precise data on the accuracy of these estimates is difficult to find. Based on a projection of 2011 census data, the government-sponsored *Annual Report on the Berlin Housing Market 2019* finds an average dwelling size of 73.2 square metres for residential dwellings across Berlin (Investitionsbank, 2019, p. 72), which is slightly below the value predicted by our model. Housing prices per square metre vary slightly with income. Households of the five lowest income groups pay similar per square metre prices ranging from 7.86 to 7.84. The highest two income types pay a slightly higher average price per square metre with the highest income type paying 7.93 Euros on average. Figure 5.B.3 in the Appendix depicts the spatial distribution of prices across ZIP codes. Further, the model replicates the fact that average expenditure on housing decreases with income and approximates estimates for Berlin rather well (Investitionsbank, 2019, p. 83). Average rental income by income group increases in income which reflects the higher home-ownership rates among higher income groups. The average rental income for the highest income bracket is 666.02 Euros and households with a representative income of 850 Euros on average earn a rental income of just above 12.14 Euros, as the large majority of this income group consists of tenants. Lastly, the final row depicts the population weighted type specific average mean value of the amenity index, A_k^j. The average amenity index is larger for the four highest income groups. This is driven in part by amenities such as bars and restaurants, as the rich sort into high amenity locations, and in part by the endogenous compositional amenity.

5.5.2 Gentrification

In the first counterfactual, we try to assess the effects of gentrification. Gentrification is usually defined as an increase in the number of highly-educated, high income households in central districts of cities with an originally large share of poorer households, see e.g. Couture et al. (2019). In order to analyze the effects of gentrification in Berlin, we solve the model using the city-wide household distribution of income groups for the year 2013 instead of 2019.[17] This period has often been associated with strong gentrification (see Döring and Ulbricht (2018)). More precisely, we use the 2013 values for the citywide number of household by income group and solve the location equilibrium for the number of households by type in each ZIP code.[18] That is, the only difference to our baseline simulation is the distribution of the total number of households for each income group. The main driver of gentrification is then the change in the city wide income distribution. This is similar to the analysis of gentrification in Couture et al. (2019).

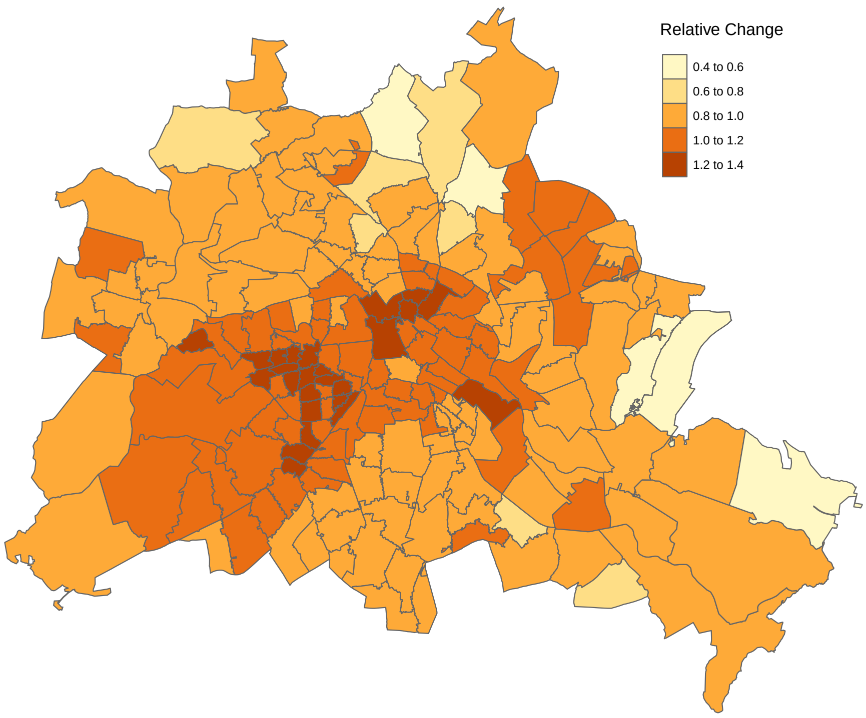
Figure 5.1 shows the change in the share of households of the highest four income brackets between 2013 to 2019 in each ZIP code relative to the average change of these household types across ZIP codes. Our model predicts the actual shares of household types across ZIP codes very well, with a correlation coefficient between actual and predicted shares of 0.96 (see Figure 5.B.4 in the Appendix). This serves as a validation of the model, since the baseline calibration allows for a quite accurate out of sample prediction.

Overall the share of households in the upper four income brackets increased across all ZIP codes. Increases were particularly pronounced in central areas such as and Mitte, Schoeneberg and Prenzlauer Berg where they ranged from six to nine percentage points;

[17]The city-wide distribution for 2013 is contained in the GFK data.
[18]We solve for the housing market equilibrium using an interpolation of (5.15).

substantially higher than the average of 5.34 percentage points. In addition, some more peripheral, already relatively wealthy ZIP codes particularly in southern Berlin also experienced above average increases in their share of high income households.

Figure 5.1: Change in share of high income households per ZIP code relative to average change across ZIP codes



Note: Figure 5.1 depicts the percentage point change in the share of households of the highest four income brackets in each ZIP code relative to the average percentage point change across ZIP codes.

We now look at the effect of gentrification on income segregation. To this end, we use the well-known dissimilarity index of segregation. Let n_k^P and n_k^R be the number of poor (lowest three income brackets) and rich (highest four income brackets) households in ZIP code k. Let $d_k = \left|\frac{n_k^P}{n^P} - \frac{n_k^R}{n^R}\right|$, where $n^J = \sum_k n_k^J$ for $J = P, R$. The dissimilarity index is defined as

$$D = \frac{1}{2}\sum_{k=1}^{M} d_k.$$

For each ZIP code, d_k describes how the ZIP code's income mix deviates from the city's aggregate income mix. Below, we assume that this will affect poor households' welfare. For now, the overall dissimilarity index measures the fraction of rich (or poor) households that would have to change ZIP code in order for rich and poor to be evenly distributed throughout the city.

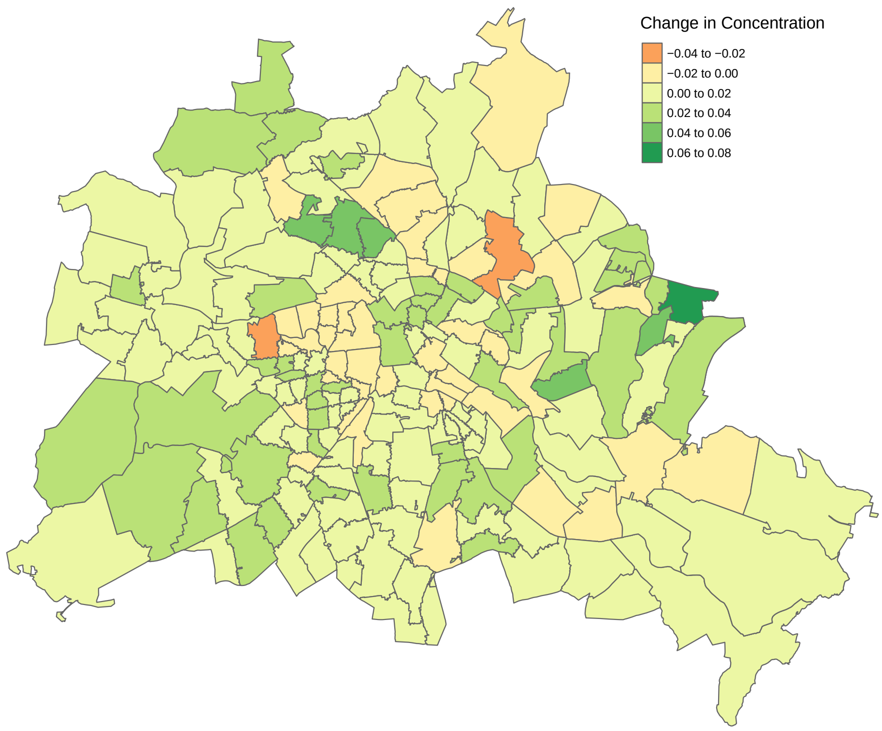
For 2013, we find a value of $D = 0.2230$. In 2019, the index increases to 0.2323, which suggests a slight increase in city-wide segregation. To get a glimpse at which parts of the city experienced changes in concentration, Figure 5.2 shows the values of d_k per ZIP code. An increase from 2013 to 2019 shows that the ZIP code experienced a greater concentration of either rich or poor households, relative to the city average. The figure shows that this is indeed the case in wealthier outskirts but also in central areas such as Mitte, Friedrichshain and Prenzlauer Berg. These districts had relatively large shares of low income households in 2013[19] and experienced an above average increase in the share of high income households (see Figure 5.1). Areas with an even larger share of low income households such as Neukölln, in contrast, experienced an influx in the share of high income households but also a fall in the concentration measure, suggesting an increase in income dispersion.

These different patterns might indicate different stages of gentrification. In the early stage of gentrification, the income distribution becomes more dispersed, as high income households move into areas with initially high shares of low income households. In a more advanced stage of gentrification, however, the concentration of high income households increases, as poorer households are successively displaced. Overall, the findings document patterns that have frequently been attributed to gentrification such as an increase in the concentration of rich households in central areas with an initially high share of low income households.

Lastly, we quantify the impact of gentrification on the different types of households. Table 5.5 summarizes the results of the counterfactual relative to the baseline. From 2013 to 2019 prices rose between 1.53 percent for households of the low income group and 3.25 for

[19] See Figure 5.B.5 in Appendix B.

Figure 5.2: Change in Concentration per ZIP Code



Note: Figure 5.2 depicts the change in the within ZIP code concentration from 2013-2019.

households in the highest income bracket. Average floor space decreased across all income brackets, with slightly larger decreases for the higher income groups. The type specific amenity index decreased for the lower income households. In contrast, it increased across household types four to seven.

Table 5.5: Counterfactual: Gentrification 2013-2019

Change in %	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
Sqm Price	1.5252	1.6080	1.7270	1.8641	2.4405	2.5586	3.2510
Floor Space	-0.5584	-0.5465	-0.7134	-0.9880	-1.4735	-1.7288	-2.4177
Amenity Index	-0.0567	-0.0265	-0.0135	0.1579	0.1922	0.1913	0.2014
Rental Income	9.3077	9.3077	9.3077	9.3077	9.3077	9.3077	9.3077
Welfare	-0.5451	-0.0155	-0.0053	-0.0734	-0.0289	0.0257	0.0197
EV	-4.35	-0.22	-0.11	-2.02	-1.2	1.92	2.37

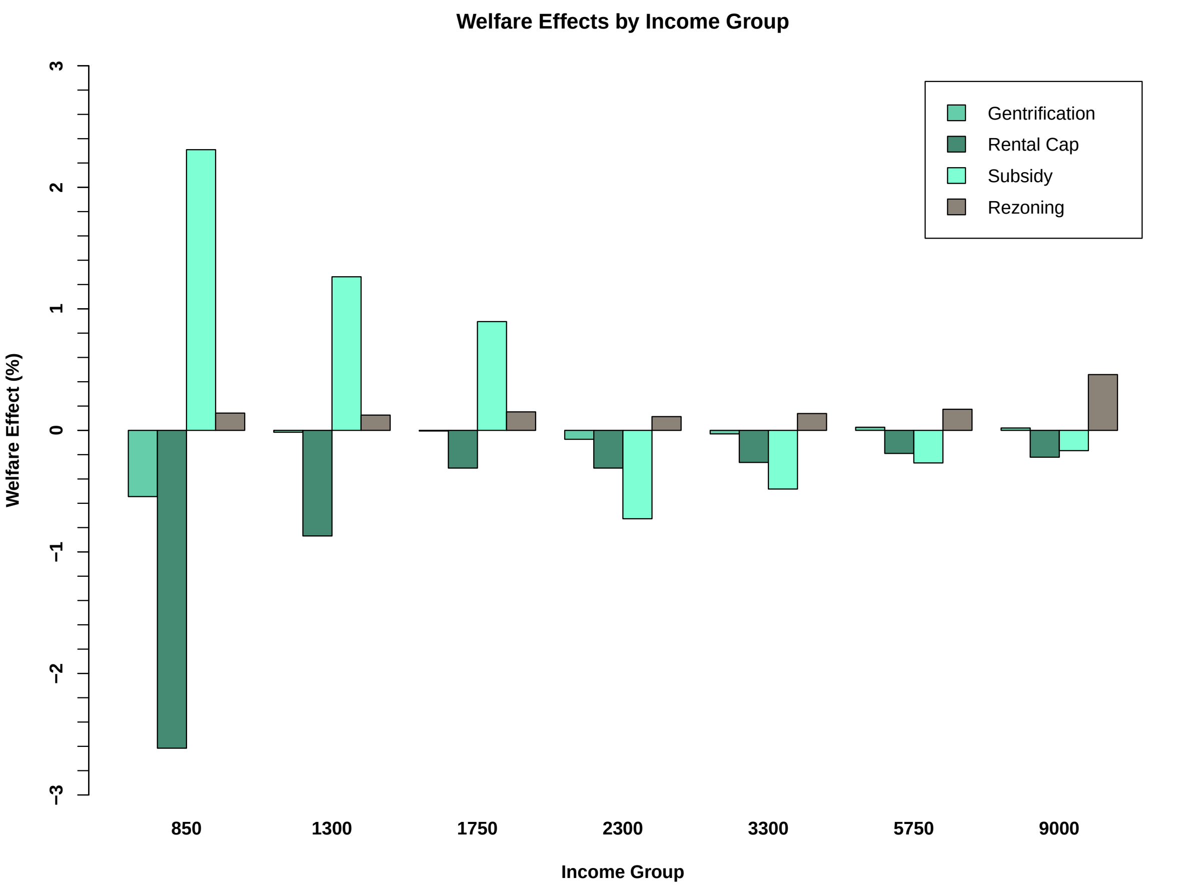
Source: Own model calculations; values expressed in percentage changes. Equivalent variation (EV) expressed in Euros. The above table lists the change in relevant outcomes of all counterfactuals in relation to the baseline calibration.

This reflects two separate effects of gentrification: first, high income households tend to move to gentrifying areas with lots of bars, restaurants etc., whereas low income households tend to move out of these areas. Second, for high income households, this is reinforced by the growing shares of other high income households in these areas. Average rental income earned by homeowners rose by 9.3 percent. Ultimately, the change in the income distribution results in a reduction of expected welfare for households of the lowest five income brackets (see Table 5.5 and Figure 5.3). Households of the lowest income bracket suffer the greatest welfare loss of 0.55 percent. This corresponds to an equivalent variation of minus 4 Euro and 35 Cent per month. For the other income groups, welfare changes are relatively small. The highest two income groups benefit slightly, since the increase in rental income and endogenous amenities outweigh the increase in housing prices.[20]

[20] Note that the welfare effects are not, as one might think, monotonic in households' income. For example, the relative welfare loss is larger for type 4 than for type 3, but smaller for type 5 than type

In order to address the consequences of gentrification and housing affordability, the city state of Berlin in 2020 passed a rental cap, as described above. In the following sections we first implement a counterfactual, which mimicks this rental cap and quantify its long-term welfare effects. We also look at alternative policies and can thus analyze whether and to which extent different income groups benefit from different policies designed to mitigate the documented distributional consequences of gentrification. Figure 5.3 gives an overview of the welfare consequences of gentrification and each analysed policy by income group.

Figure 5.3: Welfare Effects by Income Group



Note: Figure 5.3 depicts the welfare change in percent for all four main counterfactual analyses conducted in the paper.

4 households. This is driven in part by the low homeownership rate of type 4 households, and in part by the fact that in 2013 many of these households already lived in central, gentrifying areas and thus were relatively strongly affected by the price increases in these gentrifying areas. The particular location of this group relative to others could itself be driven by the difference in homeownership rate or by the exogenous location preference (i.e. the parameter B_k^j).

5.5.3 Rental Cap

As described above, in 2020 Berlin passed a city-wide rental cap which applied to new as well as existing rental contracts.[21] In April 2021 the policy was declared unconstitutional by Germany's federal constitutional court and ceased to apply effective immediately. However, popular support for this policy as well as other interventions (such as the proposed expropriation of large housing companies) remains strong. Our counterfactual is intended to simulate the long-run equilibrium consequences of a strict rental cap modelled along the lines of the policy that was introduced in Berlin.

Under the rental cap, reference rents for maximum permissible rents were based on the dwelling's age and type, as well as the category of the residential area where the dwelling is located. Reference rents were calculated based on a table published by the state government of Berlin.[22] Existing contract rents were allowed to lie above the reference rent by a maximum of 20 per cent, whereas new rental contracts were strictly bound by the reference rent.

In our counterfactual analysis we take the reference rate as the maximum permissible rent without the buffer of 20 per cent. In order to emulate the maximum permissible rent according to the rental cap for each ZIP code, we again use freely available geo-data from the FIS-Broker (see above). We can then obtain values for the average age of dwellings and the category of the residential area as well as the predominant dwelling type in each cell. This enables us to approximate the maximum permissible average rent within each ZIP code. Newly built dwellings, i.e. those that were built after 2013, were exempt from the regulation. Figure 5.4 shows areas where the rental cap applies according to our model.[23] In particular, popular central areas in the districts Mitte, Prenzlauer Berg, Kreuzberg and Schöneberg, but also the outskirts of Western Berlin are affected. We then analyse how rent control affects welfare and the housing market within our model,

[21] Rents could increase with the inflation rate by up to 1.3 per cent p.a. Moreover, rents could increase following the modernization of a dwelling. The law did not apply to dwellings that were completed after Jan. 1, 2014.

[22] See Berliner Mieterverein.

[23] This is obviously a simplification, since we assume that within ZIP codes, all dwellings are the same.

by simulating a new equilibrium subject to rent control.

Figure 5.4: Application of Rental Freeze across Berlin

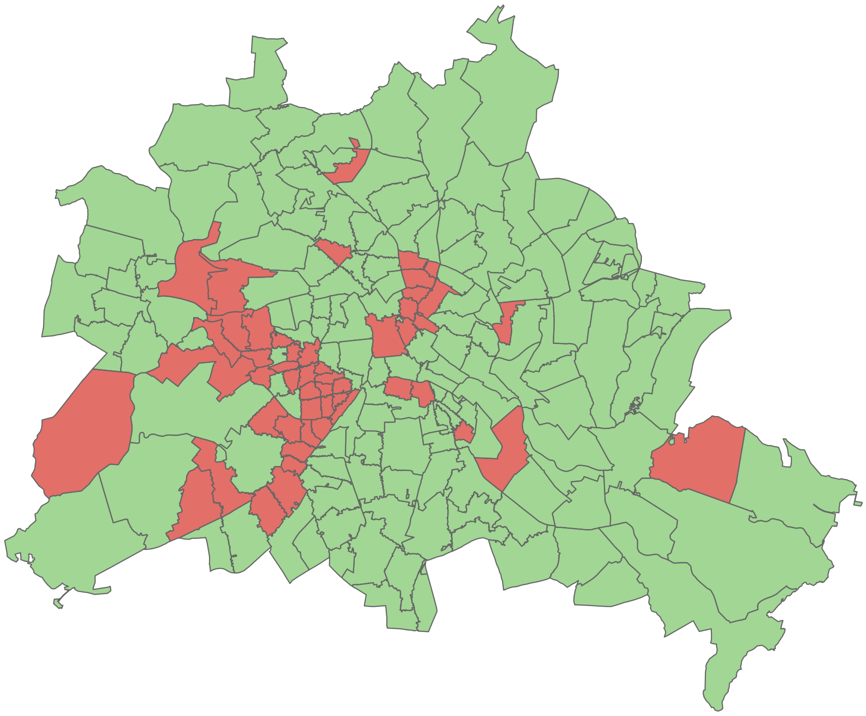


Figure 5.4 shows postal codes in which the rent control is binding in red. In these postal codes the free market rent lies above maximal permissible rents stipulated by the rent control.

With rent control, the housing market equilibrium conditions are replaced by

$$\bar{Q}_k = \Theta \hat{p}_k^{\theta}, \tag{5.18}$$

where $\bar{Q}_k$ is the aggregate quantity of housing, and the rent per sq. meter is defined as

$$\hat{p}_k = \min[p_k^*, \bar{p}_k],$$

where $\bar{p}_k$ is the controlled rent in k and p_k^* the equilibrium rent in case the cap is not binding.

We assume that when rent control binds, housing is allocated to households in proportion to their share in total housing demand at the regulated rent. Denoting type j's housing consumption in k by $\bar{q}_k^j$:[24]

$$\bar{q}_k^j = \frac{q_k^j(\hat{p}_k)}{Q_k(\hat{p}_k)} \bar{Q}_k.$$

Further, note that p_k^* is a function of n_k^j, which we can interpolate using equation (5.15). Combining the above, we use the definitions of $\hat{p}_k$ and $\bar{q}_k^j$ to solve equation (5.16) for n_k^j to obtain the counterfactual allocation of households across ZIP codes.

Table 5.6 shows the results of the counterfactual simulation relative to the baseline. The table shows that average prices fall across household types. The fall in rental prices ranges from 1.5% to 2.5%, and interestingly, is most pronounced for richer households. This is due to the fact that the rental cap reduces rents most strongly in areas that are inhabited by mostly rich households (see Figure 5.4). In response to the fall in rents, supply decreases, which leads to a reduction of housing floor space between 1.1% and 5.3%. Further, the type specific average amenity index falls across all income types, however, less so for more affluent households. The rental cap thus seems to redistribute amenity consumption from poor to rich, since the rich move into high-amenity areas where the rental cap applies.

Ultimately, rent control results in a reduction of welfare for all household types (see Figure 5.3). Importantly, low income households suffer the greatest relative welfare loss in percentage terms. For example, expected welfare on average decreases by 2.62 percent for households in the lowest income bracket and by only 0.22% for the richest household type. The equivalent variation corresponds to 20.71 Euros per month for the poorest and to 26.52 Euros for the richest households. For homeowners, welfare losses are compounded by the decrease in rental income, which amounts to 4.66 percent. Nonetheless, this is not sufficient to outweigh the smaller welfare loss of rich households due to the stronger fall

[24]This assumption implies that there is misallocation of housing under rent control, since dwellings are not necessarily allocated to the households with the highest willingness to pay (Glaeser and Luttmer, 2003; Mense et al., 2019). Other possible rationing mechanisms include efficient and random allocation. Obviously, the rationing mechanism affects the distributional effects of rent control (see below).

Table 5.6: Counterfactual: Rental Cap

Change in %	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
Sqm Price	-1.9879	-1.5556	-1.5144	-1.5049	-1.7643	-2.4070	-2.5469
Floor Space	-1.0973	-3.5945	-3.8701	-3.9099	-4.3408	-5.4621	-5.3158
Amenity Index	-0.9403	-0.3070	-0.2434	-0.1126	-0.0430	-0.0018	-0.0029
Rental Income	-4.6582	-4.6582	-4.6582	-4.6582	-4.6582	-4.6582	-4.6582
Welfare	-2.6153	-0.8689	-0.5720	-0.3103	-0.2639	-0.1898	-0.2209
EV	-20.71	-12.65	-11.77	-8.53	-10.94	-14.18	-26.52

Source: Own model calculations; values expressed in percentage changes. Equivalent variation (EV) expressed in Euros. The above table lists the change in relevant outcomes of the rental cap counterfactual in relation to the baseline calibration.

in housing rents.[25] In summary, it appears that in the long run, capping rents in Berlin would be regressive, as richer households suffer proportionately less than poorer ones.

We find that rent control leads to an increase in segregation: compared to the baseline, the city-wide dissimilarity index rises from 0.2323 to 0.2606. When looking at the ZIP code specific change in concentration, we find that the increase in concentration is particularly driven by an increase in the concentration of high income households in the regulated central city ZIP codes (see Figures 5.B.6 and 5.B.7 in Appendix B). This effect is driven by the larger willingness to pay for central amenities of high income households, as the rental cap mainly reduces prices for housing in central, high-amenity ZIP codes.[26] Conversely, the concentration measure decreases in less central areas with fewer amenities. This finding stands in contrast to popular arguments, which view rent control as a measure

[25]Note that the welfare loss gets weaker as income increases from type one to six, whereas it is stronger for type 7 than for type 6. This is due to the fact that from type 6 to 7, the homeownership increases more than proportionately with income. Also, a relatively larger share of households of type 7 lives in suburban ZIP codes where the rental cap did not bind, so that the fall in rents for these households is correspondingly lower.

[26]Note that this effect is driven by the assumption that the restricted housing supply is allocated proportionally to households types' specific housing demand at restricted prices in a given region. A random allocation mechanism or a guarantee that a certain amount of the restricted housing supply is given to lower income households, for example, could potentially counteract the negative segregation effects of the simulated rent control policy. Such allocation mechanisms, however, were not included in Berlin's rental cap policy.

to preserve a relatively even income mix in central cities. Note that the amenity index decreases most for poor households, which amplifies the adverse distributional effects.

5.5.4 Housing Subsidy

An alternative measure to implement affordable housing, particularly for lower income households, is a targeted housing subsidy. We hence run a counterfactual policy that redistributes income from high income to low income households and requires the subsidy to be spent on housing. More precisely, the counterfactual policy subsidizes 30 square metres of housing for each tenant household of the first three income brackets, i.e. our q_0 in the model. All households with an income below 2300 Euros are eligible to receive housing benefits. The subsidy in turn is financed by a flat tax levied on households of the upper four income brackets. In our example we implement a flat tax of 20 Euros per month which implies a subsidy of 62 cents per sq. meter for the first thirty square metres of eligible households.

Table 5.7: Counterfactual: Housing Subsidy

Change in %	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
Net Sqm Price	-5.4534	-4.4086	-3.7484	-0.0031	-0.0003	0.0017	0.0020
Floor Space	0.8603	0.6965	0.5924	-0.5445	-0.4199	-0.2715	-0.1840
Amenity Index	0.0135	0.0035	0.0013	-0.0018	-0.0009	-0.0004	-0.0002
Rental Income	-0.0018	-0.0018	-0.0018	-0.0018	-0.0018	-0.0018	-0.0018
Welfare	2.3096	1.2638	0.8949	-0.7273	-0.4827	-0.2680	-0.1669
EV	18.52	18.49	18.48	-19.98	-20.00	-20.02	-20.05

Source: Own model calculations; values expressed in percentage changes. Equivalent variation (EV) expressed in Euros. The above table lists the change in relevant outcomes of the subsidy counterfactual in relation to the baseline calibration. Note that the change in average prices for the subsidized households is derived as the weighted average of the subsidzed price for the first 30 sqm and the unsubsidized price for any sqm in excess of 30 sqm.

Table 5.7 shows the counterfactual results. As a consequence of the subsidy, net prices per sqm for subsidized households fall. The fall is most pronounced for households of the lowest income bracket as the subsidized 30 sqm constitute a larger share of Type 1's housing consumption than for the other households. The lower housing rents in turn induce an increase in average floor space for the subsidized households. In addition, the subsidized households on average benefit from better amenities. Ultimately, lower net rental prices and the increase in housing consumption as well as amenities spark an increase in welfare for the subsidized households, which ranges from 0.89 percent for Type 3 households to 2.31 percent for the poorest households (see Figure 5.3).

For high income households, lower net wages reduce housing consumption. The changes in average rental income induced by general equilibrium effects are small but negative. Thus the negative effect on rental income stemming from a decrease in average floor space for the higher income households slightly dominates the positive effect on rental income stemming from an increase in housing demand for the subsidized households. Welfare for richer households falls by between 0.17 percent for the richest households and 0.73 percent for Type 4 households. City-wide segregation is essentially unchanged. Note also that unlike for the rental cap, the locational sorting here leads to a slight fall in the amenity index for rich households, whereas amenities actually increase for poorer households. Overall, then, welfare is redistributed from rich to poor households.[27]

5.5.5 Rezoning: Residential Development of the Tempelhofer Feld

Another policy that is often proposed to address rising housing prices is to increase housing supply through more construction. Since space for new construction projects in cities is scarce, rezoning policies play a key role for housing supply in urban agglomerations.

We run a counterfactual in which the former city airport in Tempelhof (currently a recreational green space) is used for the construction of new affordable housing. A partial

[27]Note, however, that the policy is still partially regressive, as among the rich households the welfare loss is proportionately smallest for the richest households. Obviously, combining the subsidy with financing via the existing progressive income tax or a tax on rental income could undo this effect.

residential development of the Tempelhofer Feld was subject to a referendum in 2014 but was rejected by the majority of voters. Parties such as the Christian Democrats (CDU), the Liberals (FDP) and Social Democrats (SPD), however, have expressed support for a new referendum. In our counterfactual we simulate a scenario where 9000 flats are built at the periphery of the Tempelhofer Feld.[28] Existing proposals for the partial development of the Tempelhofer Feld impose the construction of affordable dwellings.[29] In an attempt to mimic this, we calculate housing prices for the new dwellings based on the adjacent ZIP code's parameters, i.e. Θ and B. The housing prices for the new flats on the Tempelhofer Feld thus reflect rental barometer rents of similar adjacent regions which tend to be lower than market rents for exclusively newly-built dwellings.

Table 5.8: Counterfactual: Rezoning Tempelhofer Feld

Change in %	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
Sqm Price	-0.1064	-0.1068	-0.1073	-0.1068	-0.1091	-0.1153	-0.1176
Floor Space	0.0419	0.0527	0.0604	0.0669	0.0755	0.0897	0.0942
Amenity Index	-0.0389	-0.0336	-0.0324	-0.0296	-0.0319	-0.0345	-0.0470
Rental Income	-0.0691	-0.0691	-0.0691	-0.0691	-0.0691	-0.0691	-0.0691
Welfare	0.1422	0.1256	0.1522	0.1132	0.1385	0.1737	0.4590
EV	1.12	1.40	1.95	2.57	6.26	11.00	23.00

Source: Own model calculations; values expressed in percentage changes. Equivalent variation (EV) expressed in Euros. The above table lists the change in relevant outcomes of the supply counterfactual in relation to the baseline calibration.

In the new counterfactual equilibrium accommodations for 9000 households are provided at an average rent of 7.91 Euro, which lies above the city-wide average of 7.87 Euro. Table 5.8 shows the percentage changes in rents, floor space, rental income and welfare. As theory suggests, rental prices for all income types decrease slightly as housing supply increases and average floor space consequently also increases slightly. The change in

[28] More precisely, we choose parameters such that in the new equilibrium 9,000 households live in this area. Proposals by the political parties in support of the policy range from 6000 to 12000 flats.

[29] For example, the electoral program of Social Democrats suggests exclusive construction rights for state-owned housing associations.

rental income is small but negative. Thus, the negative effect on rental income from the reduction in rental prices marginally dominates the positive effect of the increase in total supply in the city. Overall welfare increases across all income groups. The partial redevelopment of the Tempelhofer Feld leads to a fall in rents across the city. This makes living in ZIP codes previously predominantly inhabited by richer households more affordable for less affluent households, which slightly reduces segregation. The dissimilarity index falls marginally from 0.2323 to 0.2313 after rezoning.

Welfare effects are largest for the lowest income group and the upper three income groups (see Figure 5.3). The highest income groups experience the largest reduction in prices, while the decrease in rental income is relatively small. The lowest income group experiences a reduction in rental prices similar to the other types, but benefits proportionately more because of their large expenditure share of housing. Further, these households do not suffer much from the reduction in rental income. The average amenity index falls across income types. This effect is partially mechanic, as the amenity index for the newly constructed housing on the Tempelhofer Feld derived from the adjacent ZIP code lies below the type-specific average values with a value of 10.92 for the lower income households and a value of 21.92 for Types four to seven.[30] Note, however, that unlike for the rental cap, the amenity index falls only slightly and to similar extents across income groups. Similar to the housing subsidy, we find that in the case of the rezoning policy, locational sorting does not adversely impact poor households.

5.5.6 Interpretation of Results

Our results suggest that the observed patterns of gentrification in Berlin had distributional consequences, as has been argued by various advocates and policy makers: richer households who own a home benefit, while poor households lose. According to our simulations, rent control cannot solve this redistribution from bottom to top. While we do find that rents decrease across the board, this does not help poor households, even

[30]Note that the different values for low and high income types stem from the endogenous amenity component which in the main calibration is only modelled for Type four to seven.

though the budget share of housing decreases with income. Rather, poor households' welfare falls. This fall is driven partly by the fact that the reduction in rents is largest in central high-amenity locations predominantly inhabited by the rich, and it is further reinforced by the fact that rent control increases segregation.

Alternative affordable housing policies such as rezoning policies and housing subsidies appear better suited to address the consequences of gentrification. The housing subsidy by construction directly redistributes income from rich to poor households. The rezoning measure of a partial residential development of the old Tempelhof airport suggests that an increase in housing supply reduces average housing prices across the city and across income groups, which leads to an increase in average welfare across income groups. Our simulations show that in contrast to rent control, these policies do not have adverse redistributive consequences, in part because there is no increase of segregation.

The following section presents a number of sensitivity analyses and model extensions. Most importantly, we extend the model by including additional adjacent regions, which may be important since clearly mobility is not restricted by city borders. We also analyze the consequences of absentee home-ownership. Moreover, we introduce a source of market failure by assuming an externality stemming from income segregation. This reflects the belief of some advocates that gentrification has adverse consequences by altering the composition of households in city quarters.

5.6 Model Extension and Sensitivity Analyses

This section includes several model extensions and sensitivity analyses, where we vary the levels of parameters on whose value we are less confident.

5.6.1 Including adjacent ZIP Codes

We first extend the model by including adjacent ZIP codes outside the administrative city limits in the state of Brandenburg. Including Berlin's peripheral regions is potentially important. As they lie outside of the city and state borders, the rules of the rental cap did not apply there, so analyzing the consequences of the policy should take account of the potential mobility between restricted and unrestricted regions.[31]

Obtaining price estimates for these ZIP codes based on the Berlin rental barometer and geo-coded housing features as before is not possible. We hence use the rough estimates of advertised rents listed in Investitionsbank (2019). Further, we adjust these estimates in order to account for the upward bias contained in advertised rents in order to mimic the mix of new and existing contract rents that reflect the price estimates based on the Berlin rental barometer.

Table 5.B.1 in Appendix B shows the new baseline calibration results for Berlin and the included regions in Brandenburg. In comparison to the baseline calibration without adjacent regions, average rental income is slightly higher across income groups in Berlin. Budget shares marginally increase and average floor space by income group slightly increases.

Table 5.9 shows the calibration results relative to the new baseline calibration.[32] Overall, the results for Berlin are similar to the counterfactual results without adjacent regions. As before, welfare decreases across all income groups and the welfare losses in relative terms are larger for lower income households.

In addition to our previous results, we can now assess to what extent Berlin's surroundings are affected by the rental cap. There is an increase in demand for unregulated housing outside Berlin among the five lowest income brackets.[33] In contrast, a very small fraction

[31] Since this is not relevant for the other counterfactuals, we only look at the rent control counterfactual in this Subsection.

[32] We calculate the respective outcomes for all ZIP codes in Berlin and all ZIP codes in Brandenburg seperately. Note that the total number of households in Berlin and in Brandenburg changes between baseline calibration and counterfactual, as households can move between these two regions.

[33] The expected welfare function described in (5.17) takes the sum over regions. It is therefore necessary

of households of the richer income groups move to Berlin. In sum, due to the higher demand in the unrestricted segment, prices in Berlin's adjacent regions increase for all income groups (see Figure 5.5). As prices increase, housing supply increases, however, average floor space per household decreases due to the influx in households moving to Berlin's adjacent regions.

Our counterfactual results are consistent with recent findings on the impact of locally restricted rent control measures. Mense et al. (2019) highlight an increase in investment and construction activity in unregulated segments of the housing market, as the demand for unregulated dwellings increases due to the shortage of supply in the regulated segment. Hahn et al. (2022) analyze the short-term effects of Berlin's rental cap. They find an increase in advertised rents in the unregulated market of Berlin's adjacent regions indicating substitution away from regulated housing supply in Berlin towards unregulated housing supply in the city's periphery.

to assume that Brandenburg and Berlin are two distinct regions, in order to calculate expected welfare changes for both based on (5.17).

Table 5.9: Counterfactual: Rental Cap in Berlin and adjacent Regions

Change in %	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
Berlin							
Sqm Price	-2.0396	-1.6098	-1.5749	-1.5625	-1.8096	-2.4464	-2.5743
Floor Space	-2.4502	-5.0086	-5.0205	-5.0081	-4.9232	-4.7909	-4.7771
Amenity Index	-0.9377	-0.2910	-0.1844	-0.1416	-0.0500	-0.0081	-0.0106
Number of HH	-1.0025	-0.2857	-0.1133	-0.0907	-0.0303	0.0389	0.0281
Welfare	-2.4714	-0.7763	-0.4961	-0.2833	-0.2368	-0.1655	-0.1954
EV	19.88	11.6090	10.4753	7.986	10.06	12.67	24.10
Percentage Point Difference in Welfare Change to Berlin-Only Counterfactual							
Welfare	0.1439	0.0926	0.0758	0.0270	0.0271	0.0243	0.0255
Brandenburg							
Sqm Price	0.2118	0.2068	0.1894	0.1642	0.1133	0.1040	0.0740
Floor Space	0.1261	-0.0056	-0.0255	0.0039	-0.0651	-0.1049	-0.1605
Amenity Index	-0.0011	-0.0012	-0.0010	-0.0005	-0.0007	0.0000	0.0005
Number of HH	10.8795	2.0237	0.7409	0.4182	0.0839	-0.0891	-0.0440
Welfare	-0.1377	-0.2745	-0.2442	-0.1743	-0.2078	-0.2103	-0.2364
EV	1.18	4.24	5.27	4.98	8.92	16.21	29.26

Source: Own model calculations; values expressed in percentage changes. Equivalent variation (EV) expressed in Euros. The above table lists the change in relevant outcomes of the rental cap counterfactual for Berlin and adjacent regions in Brandenburg in relation to the baseline calibration.

Figure 5.5: Price changes due to rental cap in Berlin and adjacent areas in Brandenburg

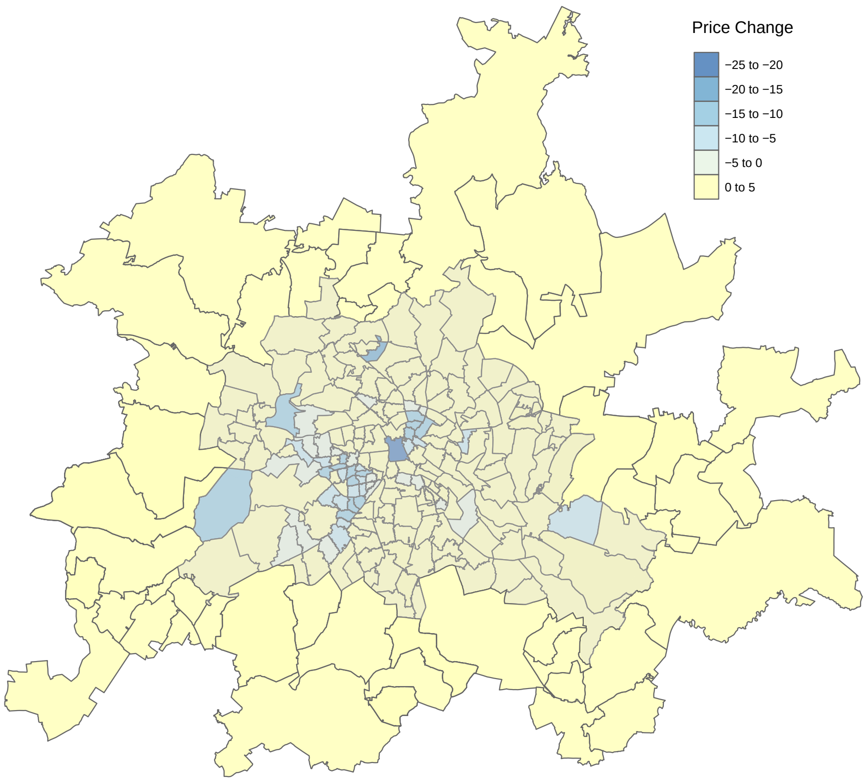


Figure 5.5 compares prices after the rental cap counterfactual with prices before. Price changes are denoted in percentage terms.

5.6.2 Choice of ζ^H

In this section we vary ζ^H, the parameter which governs the strength of the endogenous amenity that affects rich households' utility. This seems prudent, since we don't have any external information about this parameter. In order to find a unique equilibrium, ζ^H needs to be sufficiently small. We first rerun our counterfactual simulations, where we increase ζ^H from 0.01 to 0.1. Thus we weight the endogenous amenity component for richer households more heavily. Table 5.B.2 in Appendix B shows the results. The subsidy counterfactual largely remains unchanged: key outcomes such as floor space, rental price, amenity index and welfare are barely different from the previous simulations.

For the other counterfactuals we find slightly different results. First, for the gentrification counterfactual, the results hardly change for the lower three income groups, and all three groups experience a reduction in welfare. However, for the higher four income groups we find strong positive welfare effects. The explanation is intuitive: as shown above, gentrification leads to higher concentration of rich households in central city areas. With a higher value of ζ^H, this benefits the rich households even more. This is further supported by the substantially more pronounced percentage increase in the amenity index of the four highest income groups.

For the rent control counterfactual, we find slightly less negative welfare effects with a value of $\zeta^H = 0.1$ for the high income groups.[34] As shown in Section 5.5.3, the compositional effect stemming from rent control tends to favor rich households, and this effect is stronger with a larger ζ^H.

Under the rezoning counterfactual, amenity levels slightly fall. With a higher value of ζ^H, this in turn implies a lower increase in welfare, particularly for the high income households.

Lastly, we repeat our simulations for $\zeta^H = 0$. The results only change marginally for the rent control, subsidy and rezoning counterfactuals (see Tab. 5.B.2). For the gentrification counterfactual we find more pronounced changes. The positive welfare effects for the

[34]This does not hold for Type 7 households.

richest two household types now turn negative, which implies that gentrification has positive welfare effects for the rich only when we assume the presence of endogenous amenities.

In the next step, we look at an alternative rezoning policy where supply across all regions of the city is increased.

5.6.3 Alternative Rezoning Policy: Increasing Supply across the City

In Section 5.5.5, we found that the effect of allowing construction on the recreational area Tempelhofer Feld leads to increases in welfare across groups. However, since the area under consideration is small compared to the entire city, the welfare effects are obviously limited. In this section, we consider an alternative counterfactual. In particular, we assume a city-wide shock in the profitability of housing across all ZIP codes. An example might be the citywide lifting of restrictions for the construction of new housing. In our framework, Θ_k measures the underlying profitability of housing supply in a region. Increasing Θ_k can hence simulate an increase in this underlying profitability. In the next counterfactual we assume that profitability, and hence Θ_k is increased by five percent across all ZIP codes.

Table 5.10 shows the results of the counterfactual simulation. Intuitively, rents fall for all income groups as housing supply is expanded. Consequently, average floor space increases, average amenity consumption and ultimately welfare increases across all types. Interestingly, average welfare increases are most pronounced for households of the lower income groups, despite the fact that rental income increases, which predominantly benefits richer households. However, the poorer households benefit most from the reduction in rents, since their budget share is largest.

Table 5.10: Counterfactual: Rezoning across the City

Change in %	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
Sqm Price	-1.1463	-1.1523	-1.1538	-1.1523	-1.1466	-1.1455	-1.1376
Floor Space	0.0070	0.0031	0.0017	0.0018	0.0013	0.0006	0.0004
Amenity Index	0.0070	0.0043	0.0029	0.0022	0.0011	0.0007	0.0009
Rental Income	1.2827	1.2827	1.2827	1.2827	1.2827	1.2827	1.2827
Welfare	0.5033	0.3898	0.3291	0.2789	0.2568	0.2293	0.2224
EV	-4.02	-5.69	-6.79	-7.68	-10.66	-17.15	-26.73

Source: Own model calculations; values expressed in percentage changes. Equivalent variation (EV) expressed in Euros. The above table lists the change in relevant outcomes in relation to the baseline calibration.

5.6.4 Externality for Lower Incomes

Up until now, we have assumed that only rich households benefit from endogenous amenities through the number of other rich households in their neighborhood. Absent this endogenous amenity, the model assumes perfectly competitive housing markets, so the market outcome is necessarily efficient and policies such as rent control necessarily reduce welfare. One argument for such policies and against unfettered gentrification, however, is that poorer households benefit from living in neighborhoods with richer households,[35] while gentrification and segregation may directly lower poor households' welfare because of the disadvantageous effect on the mix of households.

We now assume that the endogenous amenity benefits rich and poor symmetrically, so we now assume that the parameter ζ^L, which governs the endogenous amenity effect for the poor, is positive as well. Table 5.B.3 in Appendix B shows the results of the four counterfactuals for different values of ζ^L. As above, in order to find a unique equilibrium ζ^L needs to be sufficiently small. Existing literature does not suggest values for this parameter, and the estimation of the parameter in a reduced form setting is difficult and

[35] For example see Chetty and Henderson (2018a); Guerrieri et al. (2013).

likely prone to endogeneity issues. We thus present results for $\zeta^L = 0.1$ in addition to our baseline calibration with $\zeta^L = 0$.

As before, in the gentrification counterfactual the highest two income groups experience a welfare gain and the lowest income group experiences a welfare loss. However, the welfare loss of the lowest income group is less pronounced than in the basic calibration without the cross-type externality. In addition, type two and three now experience a welfare increase. With the externality, lower income households benefit from a greater number of high income households in their ZIP code, which is sparked by gentrification (see Section 5.5.2). Further, they are more likely to select into areas with richer households. Overall, these effects lead to an increase in the average amenity consumption of the lowest three income groups and ultimately dampen the negative welfare consequences of gentrification.

In the rental cap counterfactual, average welfare changes remain negative across all income groups. However, compared to the basic counterfactual without the segregation externality, the decrease in welfare is less pronounced, particularly for the lowest income groups. With the externality, lower income households are more likely to select into or remain in an area with a greater number of richer households, which dampens the negative welfare effect for poorer households.

The results for the housing subsidy largely remain unchanged. For the rezoning counterfactual, the positive welfare effects for the lowest two income groups as well as type four are somewhat more pronounced. Conversely, for the remaining income groups the welfare gains are less strong (or basically unchanged). As shown before increasing the supply of affordable housing reduces prices across the city as well segregation (see Section 5.5.5), as housing becomes more affordable overall. With the externality, this pattern implies a smaller welfare increase for most household types, since the number of high income households is more spread out across the city. Only income type one and two experience a slight increase in welfare with the cross-type externality.

5.6.5 Absentee Landlords

Lastly, we set all the a^j parameters to zero, which implies that all housing in the city is owned by absentee landlords. Table 5.11 shows the amended welfare changes for all four main counterfactuals in relation to the baseline calibration. The effects of gentrification on welfare slightly change with absentee landowners. Welfare now decreases across all income groups, including households of the higher income groups. This in turn suggests that the positive effect found for the highest two income groups in the main specification mainly stems from an increase in rental income of homeowners. Note that, while the lowest income group still suffers the largest percentage welfare loss, the range of welfare changes is compressed relative to our main specification. Therefore, the regressive nature of gentrification seems to be largely due to the distributional implications of homeownership.

Looking at the rental cap counterfactual reveals some interesting differences to our main specification. In particular, while the lowest five income groups still experience welfare losses, the highest two income groups now experience a welfare gain. This is due to the fact that the rich do not suffer from the reduction in rental incomes under absentee ownership. Furthermore, rents are mainly reduced in areas inhabited by mostly richer households, so the rental cap would be even more regressive if the rich did not own any housing.

Table 5.11: Robustness Absentee Landowners: Δ in %Welfare

Change in %	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
Gentrification	-0.5082	-0.3548	-0.3076	-0.2500	-0.2895	-0.2467	-0.2974
Rental Cap	-2.9192	-0.8448	-0.3829	-0.1612	-0.0417	0.0451	0.0498
Housing Subsidy	2.3668	1.3665	0.9608	-0.7648	-0.5154	-0.2867	-0.1804
Rezoning	0.1416	0.1260	0.1572	0.1173	0.1430	0.1781	0.5521

Source: Own model calculations; values expressed in percentage changes. The above table lists the change in welfare of all counterfactuals in relation to the baseline calibration.

5.7 Conclusion

This paper uses a quantitative spatial equilibrium model to evaluate the distributional and welfare impacts of gentrification patterns and affordable housing policies in Berlin, Germany. We calibrate the model to key features of Berlin's housing market, in particular the recent gentrification of inner city locations and run a range of counterfactuals simulating affordable housing policies.

Overall, our results document patterns of gentrification in Berlin between the years 2013 and 2019. These patterns suggest distributional consequences: in particular, homeowners at the upper end of the income distribution benefit, while relatively poor renter households lose. According to our simulations, the rental cap does not address these issues. Instead, we find a reduction in housing prices but also a decrease in average housing consumption and rental income across the income distribution, which ultimately reduces welfare across all income groups. Most importantly, our results show that the welfare loss is strongest for the lowest income groups, which runs counter the aim of the policy.

Further, our results suggest that prices in adjacent regions of Berlin increase as a consequence of the rental cap. This effect is driven by an influx of households moving out of Berlin. The inclusion of a segregation externality suggests that the existence of the externality dampens the negative welfare effect of the rental cap on the lower income groups, as the rental cap appears to reduce segregation. However, for the parameter values used, the average welfare effect remains negative. Last, the results suggest that alternative affordable housing policies such as rezoning policies and housing subsidies are better suited to address the consequences of gentrification. The housing subsidy by construction directly redistributes income from households at the upper end of the income distribution to households at the lower end, which redistributes welfare towards the poor. The rezoning policies we consider increase housing supply which reduces average housing prices across the city and across income groups, leading to an increase in average housing consumption and ultimately an increase in average welfare across income groups.

Our results are highly topical, not only in a German context. As described in the introduction, rising housing prices and gentrification patterns are hotly debated topics, particularly in urban centres across the developed world. Affordable housing policies are a key topic on ballots in major urban agglomerations in both Europe and the United States. Directly combining the study of gentrification and affordable housing policies, as in our paper, is key for a better understanding of the mechanisms at play and ultimately the distributional and welfare consequences of different affordable housing policies in the context of growing urbanization. Importantly, our analysis shows that sorting patterns are key to fully understand the distributional consequences of affordable housing policies in the context of gentrification. Suppose that rich households sort into gentrifying locations, and local governments later introduce rent control measures in these locations. If this type of policy regulates the maximum absolute rent, then it may turn out, as in our analysis, that gentrification is in fact reinforced and poor households do not benefit from rent control, even if they do not own housing. The targeted housing subsidy we have studied presents an obvious alternative.[36]

In order to study a model with many locations and several household types, our model has made simplifying assumptions along other margins. For instance, we have assumed exogenous housing ownership, exogenous workplace locations, static equilibrium, and competitive housing markets. Extending the analysis along these directions may be a fruitful avenue for future research.[37]

[36]See also Favilukis et al. (2022) for policies targeting poor households.

[37]Ahlfeldt et al. (2015) is a classic paper with endogenous residential and workplace locations. We follow Couture et al. (2019), however, in assuming workplace locations to be exogenously given. For a dynamic model with endogenous homeownership decisions – which has only two locations – see Favilukis et al. (2022).

Appendices

5.A Appendix A: Technical Appendix

Calculation of Hedonic Rent Price Index

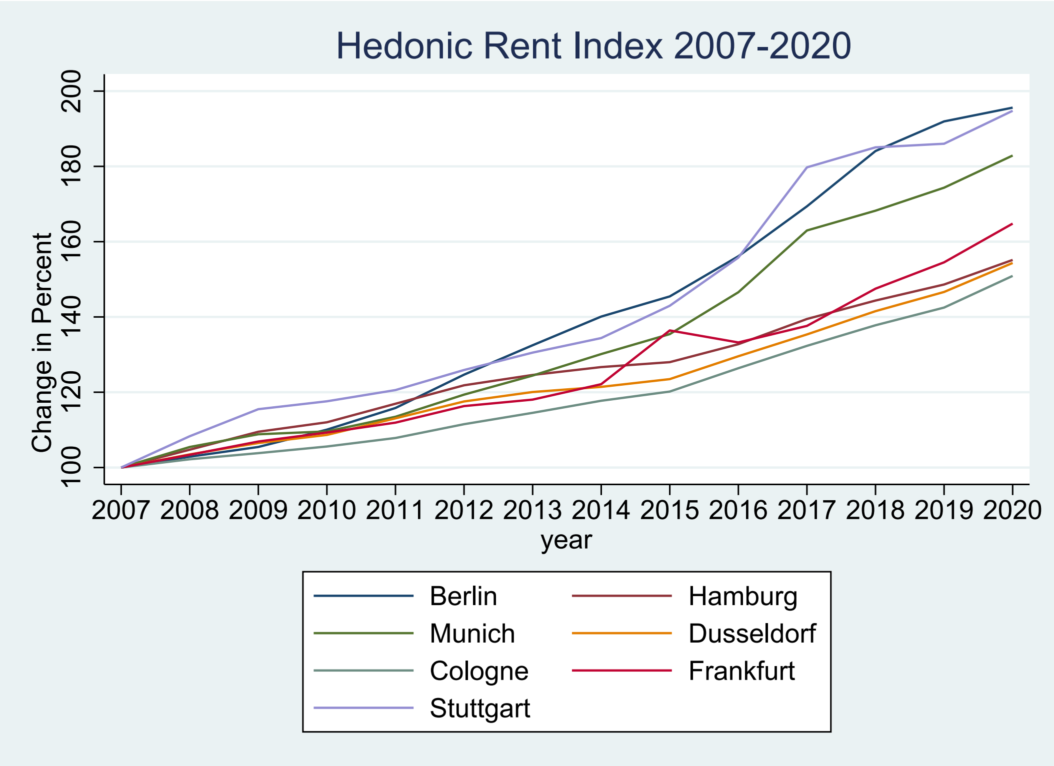
For the derivation of the hedonic rent price index we use geo-coded real estate data provided by the RWI. The data is based on online listings from the internet platform ImmobilienScout24 and includes both apartments and houses for sale and for rent (for a more detailed description see Breidenbach and Schaffner (2020)). In our derivation we focus on rental dwellings.[38] For each postal code and for each year between 2007 and 2020 we regress the logarithm of the rental price per square metre on a range of housing and quality characteristics such as the number of rooms, the apartment size, the age of the dwelling, the existence of a balcony, garden or cellar and the type of building the dwelling is situated in. Based on the intercepts we can then calculate a quality-adjusted hedonic rent price index for each ZIP code across Germany. For illustrative purposes Figure 5.A.1 solely focusses on the largest seven cities in Germany. For each city the year 2007 serves as the base year which is normalized to 100.

Calculation of Amenity Index

The data is collected through the open source platform Open Street Map. For each postal code we count the number of a range of amenities within the post code. More precisely, we collect the number of bars, nightclubs, restaurants, shops, supermarkets, ATMs, parks and doctors within each ZIP code. Based on these values we perform a principal component analysis (PCA). PCA allows us to calculate a single measure which can best predict amenities in each ZIP code. The first principal component of these amenities then serves as the basis of the exogenous part in the composite amenity index as depicted in Figure 5.B.2. Table 5.A.1 depicts the corresponding loadings.

[38] The corresponding datasets are Boelmann et al. (2020a,b); Boelmann and Schaffner (2019)

Figure 5.A.1: Development of hedonic transaction price rents



The figure depicts the development of a hedonic, quality–adjusted rent price index for new rental leases in the seven largest cities of Germany. The base year is 2007 for each city. The index uses the RWI RED dataset, which includes all advertised rents on Germany's largest internet platform *Immoscout24*. For a detailed description see of the data and the derivation of the hedonic price index see the Appendix.

Table 5.A.1: Principal Component Analysis and Loadings

Variable	Loadings
Bar Score	0.3905
Shop Score	0.4119
Station Score	0.1441
Supermarket Score	0.3320
Nightclub Score	0.3395
ATM Score	0.4178
Park Score	0.1849
Restaurant Score	0.4103
Doctor Score	0.2335
Proportion of Variance	0.5219

Source: Open Street Map Data, own calculations.

Estimation of shape parameters

We now describe the estimation of the income-specific shape parameters, ϵ^j. Rewriting (5.10), we get

$$\ln \pi_k^j = \ln B_k^j + \epsilon^j \ln\Big((A_k^j)^\beta (w^j + a^j R - p_k q_0)^{1-\beta} p_k^{-\alpha}\Big) + \ln\left(\sum_{\ell=1}^{K} B_\ell^j ((A_\ell^j)^\beta (w^j + a^j R - p_\ell q_0)^{1-\beta} p_\ell^{-\alpha})^{\epsilon^\ell}\right). \tag{5.A.1}$$

Letting s_k^j be the observed share of type-j households in ZIP code k, we then estimate

$$\ln s_k^j = \lambda^j + \epsilon^j \ln X_k^j + \mu_k^j, \tag{5.A.2}$$

where $X_k^j = \ln\Big((A_k^j)^\beta (w^j + a^j R - p_k q_0)^{1-\beta} p_k^{-\alpha}\Big)$ and the constant is

$$\lambda^j = \ln\left(\sum_{\ell=1}^{K} B_\ell^j ((A_\ell^j)^\beta (w^j + a^j R - p_\ell q_0)^{1-\beta} p_\ell^{-\alpha})^{\epsilon^j}\right).$$

Finally, μ_k^j is an error term which captures $\ln\Big(B_k^j\Big)$ together with other unobservables.

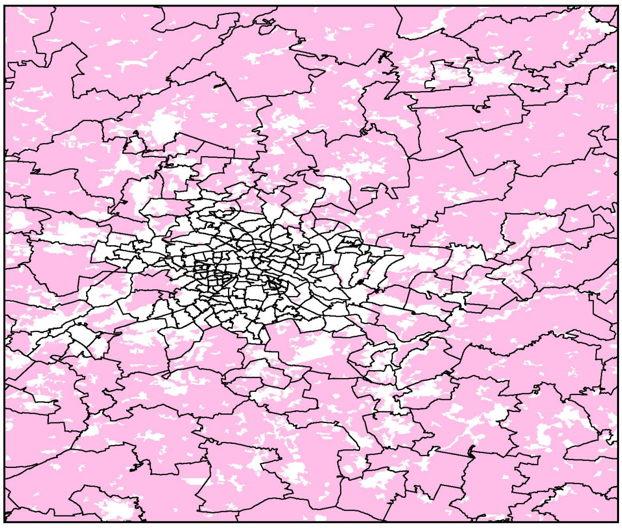
Table 5.A.2: Shape Parameter across Income Brackets: OLS Approach

Income Bracket	850	1100-1500	1500-2000	2000-2600	2600- 4000	4000-7500	9000
ϵ	4.9141	6.5097	6.0734	5.8911	3.530823	3.4030	4.3654

OLS estimates of the shape parameter are presented in Table 5.A.2. However, these estimates may be biased. For instance, housing prices could be influenced by B_k^j or other ZIP code specific unobservables that attract certain household types. Hence, we estimate equations (5.A.2) by IV regressions. To this end, we need variables that affect prices but are not correlated with the population shares. The instruments we use include a range of supply shifters which make housing supply in specific areas of the city more expensive through special regulations or physical constraints, but should be exogenous to the share of households living in the region. We now briefly discuss these instruments.

Our first instrument is the number of archeological sites and historical land monuments per hectare in each ZIP code.[39] Second, we compute the average distance from ground water to the surface in each ZIP code. Our third instrument is the share of land in each ZIP code that is part of a so-called preservation zone (*Erhaltungsgebiet*). All of the aforementioned instrumental variables are derived from openly available geo-coded data provided by the Fis-Broker. Last, we compute a ZIP code specific developable land index from satellite data.[40] Based on the landuse categories provided by the data we calculate the share of developable land in each ZIP code (see Figure 5.A.2) and construct an indicator variable equal to one if there is developable land within the ZIP code. In total, approximately 30.53 percent of ZIP codes in the city contain developable land.

Figure 5.A.2: Developable Land



Based on the Corine landcover data we derive the share of developable land. Figure 5.A.2 depicts the share of developable land in each ZIP code in pink.

All four variables shift the housing supply function. Building is subject to special regula-

[39]Land monuments are, for instance, ruins of old fortifications, burial mounds or estates. Archeological sites are excavations with findings ranging from pleistocene deposits to coinage, ruins or skeletons in more recent history. Also note that we standardize the variable by dividing it by its standard deviation.

[40]We use Corine Land Cover 2018 data to identify land that is potentially developable. The Corine Land Cover data is based on 100m resolution satelite data, which via remote sensing provides the basis for over 44 land cover and land use categories. Accordingly, we classified as developable all categories that could be used for supply, in principle: non-irrigated arable land; vineyards; fruit trees and berry plantations; pastures; complex cultivation patterns; land principally occupied by agriculture, with significant areas of natural vegetation; agro-forestry areas; broad-leaved forest; coniferous forest; mixed forest; natural grasslands; moors and heathland; transitional woodland-shrub; beaches, dunes, sands; bare rocks; sparsely vegetated areas. Note that we classify public urban green spaces such as the Tempelhofer Feld as non-developable, as public parks and recreational spaces require special rezoning legislation to be redeveloped and in itself are potential amenities that might attract specific household types.

tions in the presence of archeological sites or within preservation zones. A small distance to groundwater makes construction more difficult and hence more expensive. Last, the amount of developable land is a physical constraint on the supply of housing.

The exclusion restriction states that the instruments should not affect the type-specific population shares other than through their effect on housing prices. This likely holds for distance to groundwater, which should not affect type-specific population shares directly. It might be problematic if, for instance, landmarks and archeological sites are amenities that particularly attract high-income earners, or if non-developable land consists of parks or other green amenities which attract certain types of households. This is not especially likely for historical land monuments and archeological sites. Since these sites are underground, they do not directly influence households' decisions to locate in specific neighborhoods. In addition, since we have several instruments for each household type, we can implement specification tests checking for the set of instruments that are least likely to be endogenous. We do this by presenting Hausman-Wu tests for endogeneity and overidentification tests using Sargan statistics. These are presented at the bottom of Tab. 5.A.3. In addition we present first-stage F-statistics. In summary, the instruments seem to be strong, and pass the overidentification tests.[41]

Our IV estimates of the type-specific shape parameters are shown in Tab. 5.A.3. Overall, the IV coefficients are lower than the OLS estimates except for the highest income type. Throughout the main body of the paper, we use these parameters for our counterfactuals. For completeness, in Table 5.A.4 we present the welfare effects of our main counterfactuals using the shape parameter estimates obtained via OLS (see Tab.5.A.2). Qualitatively, the results are similar and quantitatively, the differences are mostly small.

[41]The F-statistic for Type 1 is just below 10, which might indicate a weak instrument. However, the usual problem that the first-stage coefficient becomes close to zero, which would inflate the IV coefficient, does not seem to occur here.

Table 5.A.3: Shape Parameter across Income Types: Instrumental Variable Approach

	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7
ϵ	4.793	4.542	3.374	4.654	3.915	2.849	1.753
Instrument	monuments	monuments pres. z.	pres. z.	land index	pres. z.	log monuments	monuments, pres. z., dist. groundw.
F-Stat	9.36	16.88	12.652	14.397	49.470	29.814	39.964
Wu-Haus. p-va.	0.9884	0.681	0.6787	0.8465	0.912	0.903	0.685
Sargan p-va.		0.186					0.227

Table 5.A.3 shows the estimated shape parameter for each household type, using an instrumental variable approach. Each instrument is listed below the respective shape parameter. Further, for each estimated shape parameter we show the first-stage F-Statistic as well as the p-values of Wu-Hausman and Sargan tests. The abbreviation *prev. z.* stands for preservation zones, i.e. areas with special building regulations. *Monuments* stands for the number of historical land monuments and archeological sites per hectar. *Land index* stands for the share of developable land in each postal code, and *dist. groundw.* stands for the distance to the groundwater level.

Table 5.A.4: Robustness OLS Shape Parameter: Δ in %Welfare

Change in %	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
Gentrification	-0.5524	-0.0205	-0.0088	-0.0751	-0.0267	0.0281	0.0243
Rental Cap	-2.5618	-0.8082	-0.5123	-0.2986	-0.2561	-0.1817	-0.2128
Housing Subsidy	2.3096	1.2638	0.8949	-0.7273	-0.4827	-0.2681	-0.1670
Rezoning	0.1398	0.0961	0.0948	0.0938	0.1516	0.1477	0.1926

Source: Own model calculations; values expressed in percentage changes. The above table lists the change in welfare of all counterfactuals in relation to the baseline calibration using the shape parameter estimates obtained via OLS.

5.B Appendix B: Additional Figures and Tables

Figure 5.B.1: Share of Households of the lower three Income Brackets

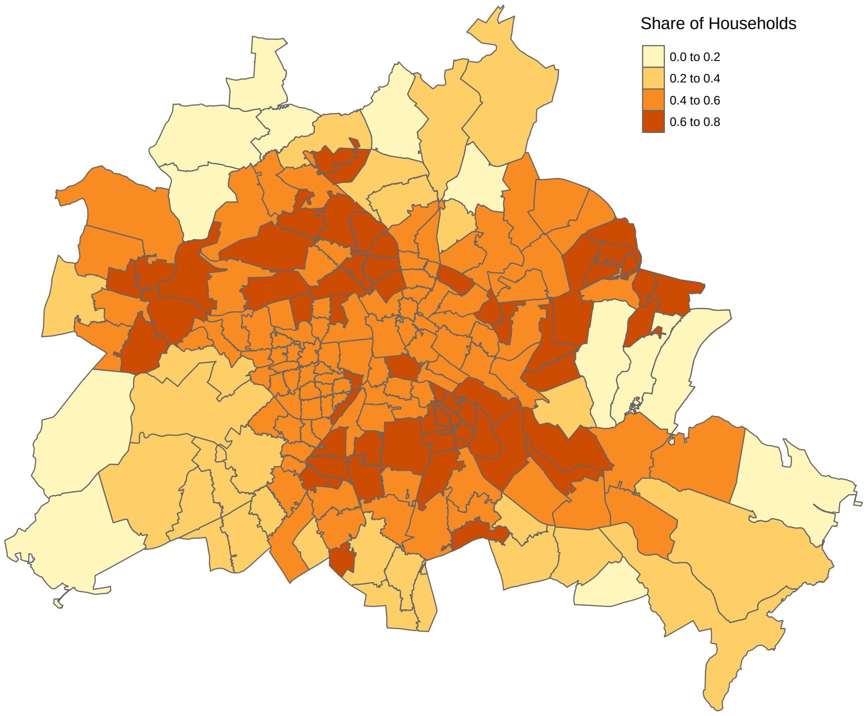
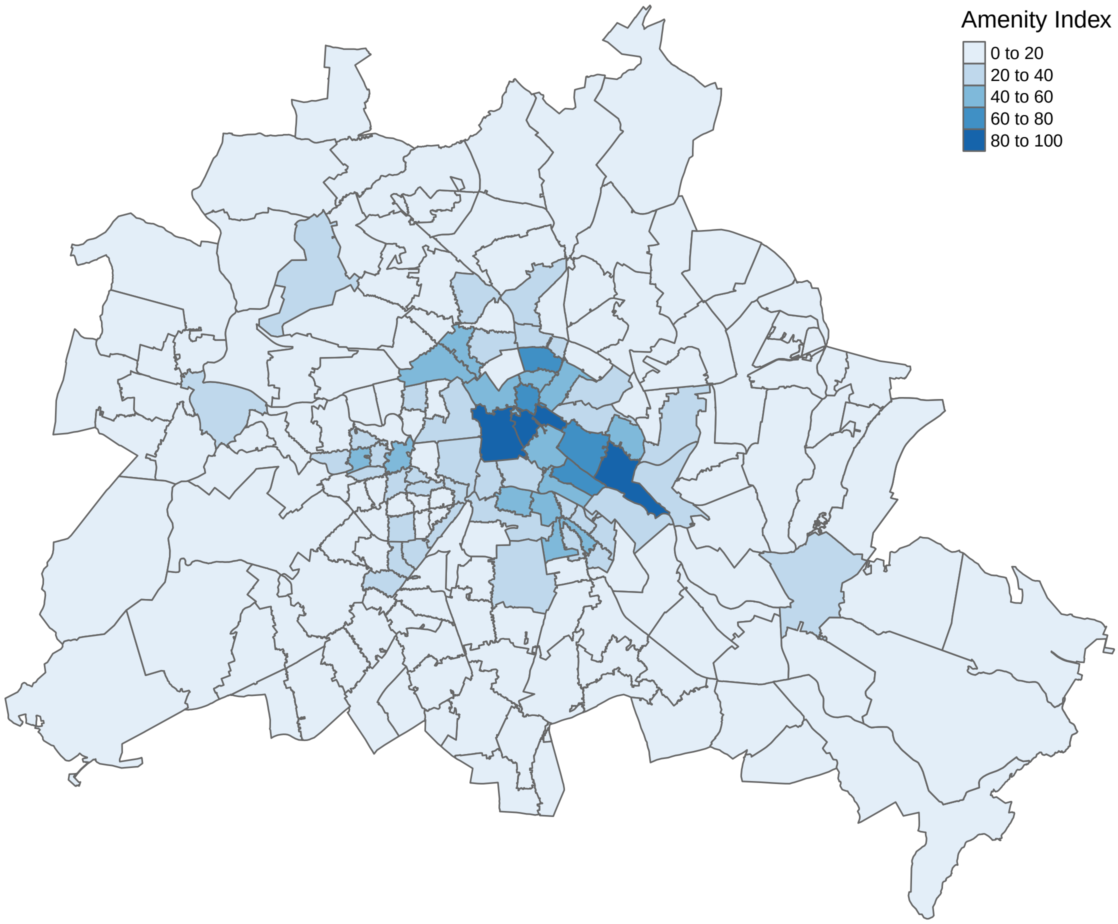


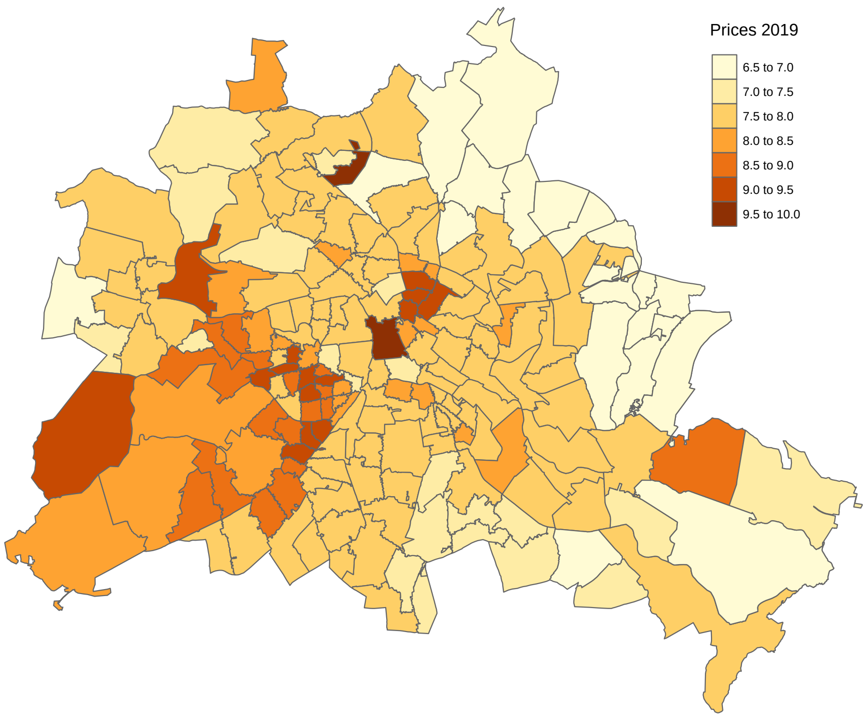
Figure 5.B.1 shows the spatial distribution of households in the lower three income brackets. More precisely, it shows the share of households in the lowest three brackets out of all households living in a given postal code.

Figure 5.B.2: Amenity Score



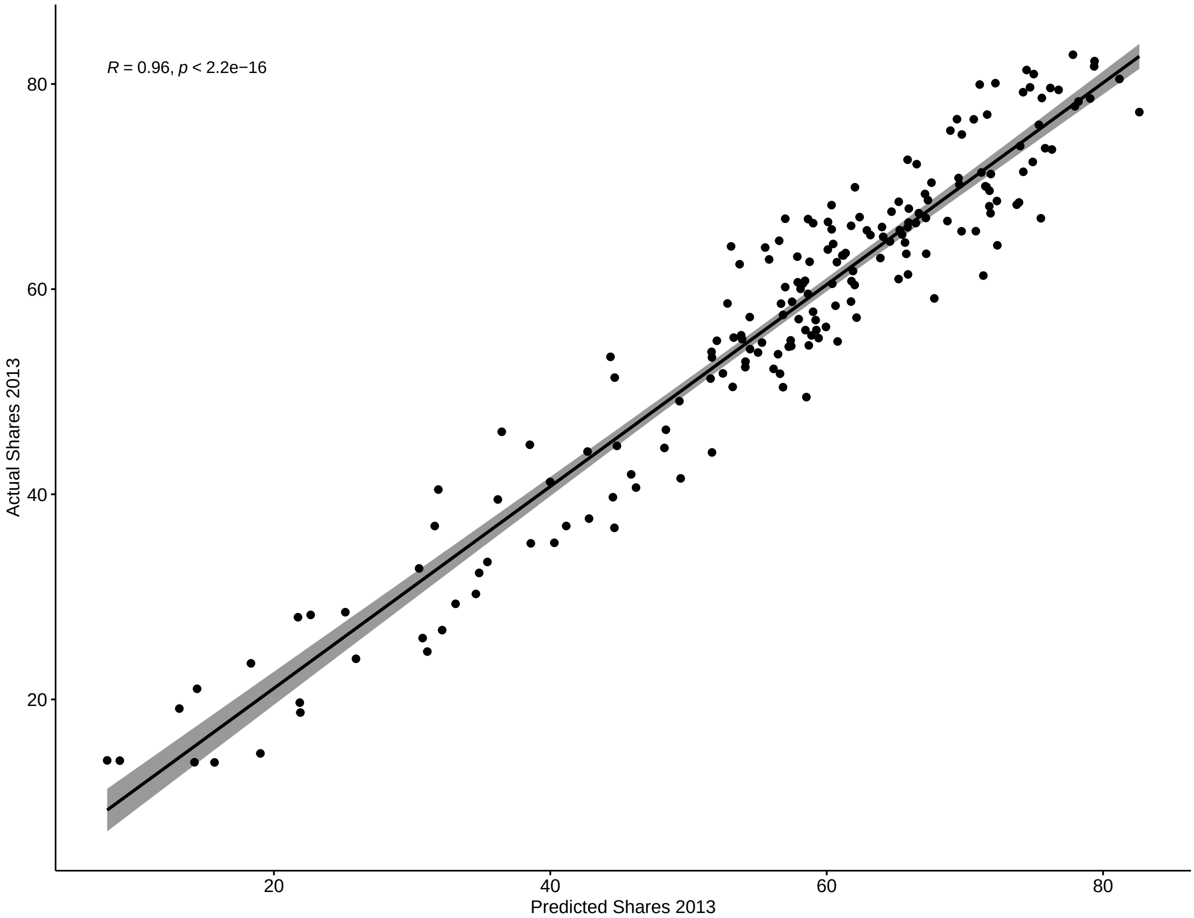
Note: The figure depicts the amenity index across ZIP codes in Berlin. The range of the amenity index is normalized to 0 (lowest)–100 (highest).

Figure 5.B.3: Rental Prices 2019



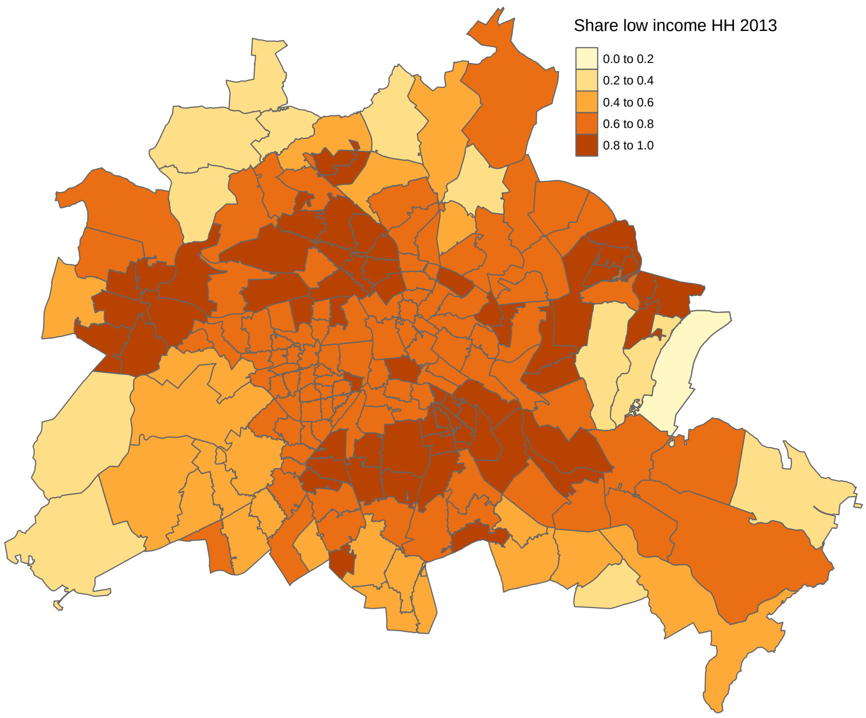
Note: The figure depicts the average square metre price of rental leases across Berlin ZIP codes. Rents include payments for utilities and bills but exclude heating payments.

Figure 5.B.4: Correlation Actual and Predicted Shares 2013



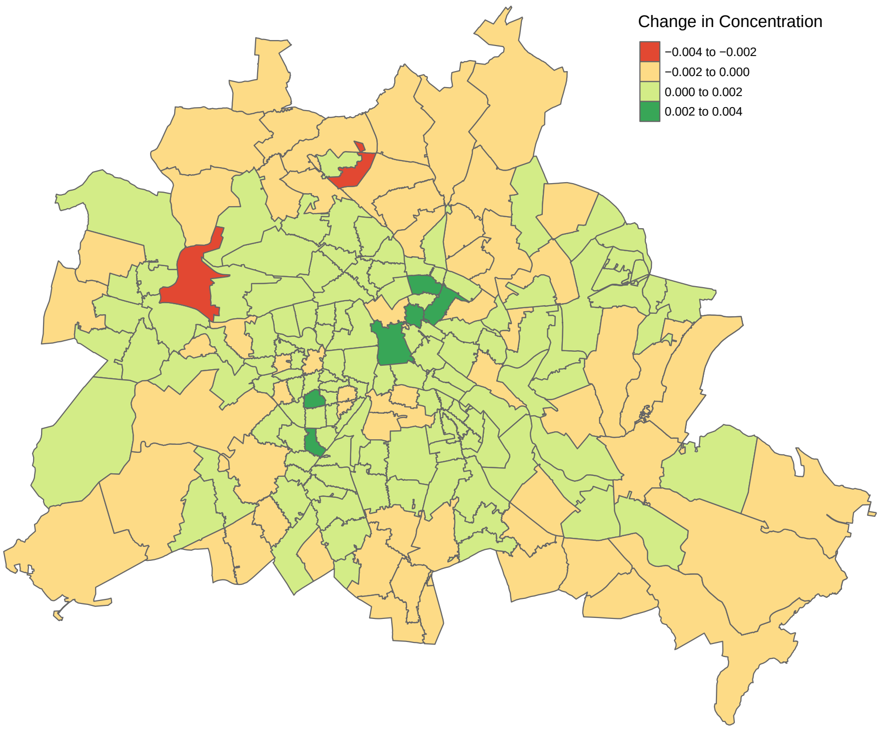
Note: The figure plots the actual share of households of the lowest three income brackets in each ZIP code for the year 2013 against the predicted shares of our model. The correlation coefficient R is equal to 0.96.

Figure 5.B.5: Share of Low Income Households per ZIP Code in 2013



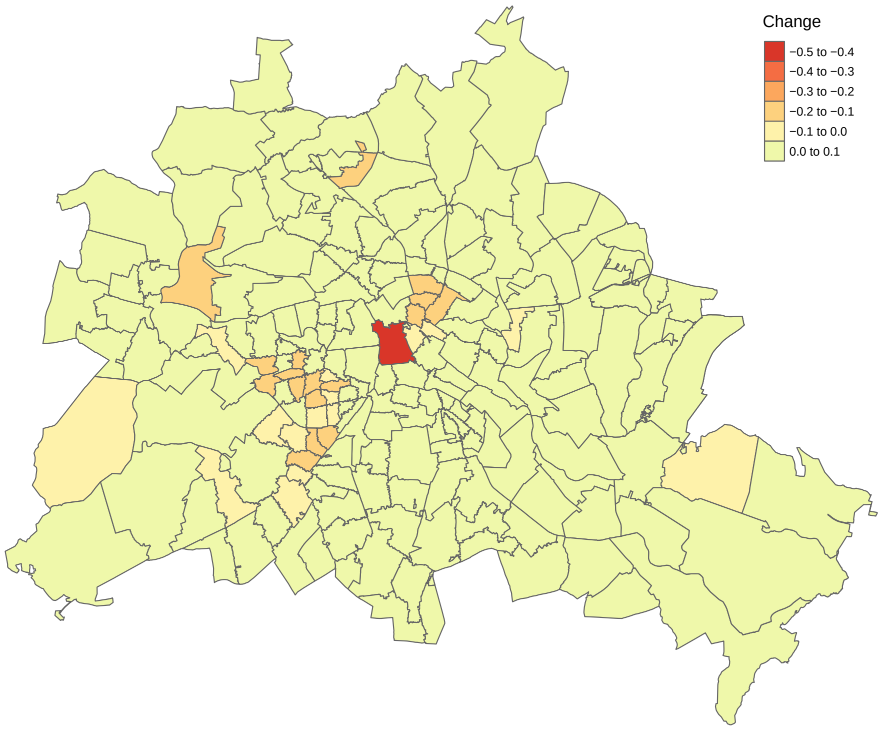
Note: The figure depicts the predicted share of households of the lowest three income brackets in each ZIP code for the year 2013.

Figure 5.B.6: Change in within ZIP Code Concentration after Rental Cap



Note: The figure depicts the change in the within ZIP code concentration, measured by the dissimilarity index, after the rental cap counterfactual.

Figure 5.B.7: Change in within ZIP Code Share of Low Income Households after Rental Cap



Note: The figure depicts the change in the within ZIP code share of households of the lowest three income group after the rental cap counterfactual.

Table 5.B.1: Baseline Calibration in Berlin and adjacent Regions

Change in %	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
Berlin							
Budget Share	0.40	0.32	0.28	0.26	0.23	0.21	0.20
Sqm Price	7.86	7.85	7.85	7.84	7.86	7.96	7.93
Floor Space	43.92	55.73	66.30	78.26	102.98	161.63	241.90
Amenity Index	21.43	20.21	19.96	29.92	29.78	30.92	29.34
Rental Income	13.37	89.01	113.90	105.57	211.33	392.35	732.98
Brandenburg							
Budget Share	0.36	0.28	0.25	0.24	0.21	0.19	0.18
Sqm Price	6.64	6.68	6.70	6.68	6.71	6.73	6.84
Floor Space	47.48	61.49	74.02	88.21	117.53	187.08	282.28
Amenity Index	15.42	15.34	15.13	25.05	24.49	23.88	23.04
Rental Income	21.23	141.35	180.88	167.65	335.60	623.07	1164.00

Source: Own model calculations. Table 5.B.1 shows the results of the basic calibration for a scenario including Berlin and adjacent regions in Brandenburg.

Table 5.B.2: Different Values of ζ^H

Change in %	ζ^H	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**	**Type 7**
				Gentrification Counterfactual				
Sqm Price	0	1.5252	1.6081	1.7270	1.8641	2.4403	2.5583	3.2505
	0.01	1.5252	1.6080	1.7270	1.8641	2.4405	2.5586	3.2510
	0.1	1.5283	1.6102	1.7282	1.8665	2.4405	2.5605	3.2526
Floor Space	0	-0.5584	-0.5465	-0.7135	-0.9879	-1.4734	-1.7285	-2.4173
	0.01	-0.5584	-0.5465	-0.7134	-0.9880	-1.4735	-1.7288	-2.4177
	0.1	-0.5596	-0.5476	-0.7141	-0.9900	-1.4727	-1.7296	-2.4184
Amenity Index	0	-0.0567	-0.0265	-0.0135	-0.0061	0.0200	0.0145	0.0157
	0.01	-0.0567	-0.0265	-0.0135	0.1579	0.1922	0.1913	0.2014
	0.1	-0.0571	-0.0269	-0.0137	1.6183	1.7308	1.7690	1.8521
Welfare	0	-0.5451	-0.0155	-0.0053	-0.1089	-0.0662	-0.0124	-0.0201
	0.01	-0.5451	-0.0155	-0.0053	-0.0734	-0.0289	0.0257	0.0197
	0.1	-0.5466	-0.0163	-0.0057	0.2456	0.3066	0.3682	0.3772
				Rental Cap Counterfactual				
Sqm Price	0	-1.9879	-1.5557	-1.5146	-1.5048	-1.7644	-2.4072	-2.5474
	0.01	-1.9879	-1.5556	-1.5144	-1.5049	-1.7643	-2.4070	-2.5469
	0.1	-1.9898	-1.5547	-1.5146	-1.5067	-1.7629	-2.4054	-2.5413
Floor Space	0	-1.0967	-3.5936	-3.8696	-3.9137	-4.3405	-5.4625	-5.3163
	0.01	-1.0973	-3.5945	-3.8701	-3.9099	-4.3408	-5.4621	-5.3158
	0.1	-1.1049	-3.6071	-3.8994	-3.8541	-4.3422	-5.4565	-5.3082
Amenity Index	0	-0.9402	-0.3072	-0.2435	-0.1095	-0.0429	-0.0005	0.0014
	0.01	-0.9403	-0.3070	-0.2434	-0.1126	-0.0430	-0.0018	-0.0029
	0.1	-0.9409	-0.3045	-0.2131	-0.1577	-0.0466	-0.0191	-0.0567
Welfare	0	-2.6149	-0.8692	-0.5721	-0.3110	-0.2642	-0.1899	-0.2208
	0.01	-2.6153	-0.8689	-0.5720	-0.3103	-0.2639	-0.1898	-0.2209
	0.1	-2.6199	-0.8646	-0.5557	-0.3017	-0.2614	-0.1886	-0.2238
				Subsidy Counterfactual				
Sqm Price	0	-5.4534	-4.4086	-3.7484	-0.0030	-0.0003	0.0017	0.0021
	0.01	-5.4534	-4.4086	-3.7484	-0.0031	-0.0003	0.0017	0.0020
	0.1	-5.4533	-4.4085	-3.7484	-0.0052	-0.0004	0.0014	0.0017
Floor Space	0	0.8603	0.6965	0.5924	-0.5446	-0.4199	-0.2715	-0.1840
	0.01	0.8603	0.6965	0.5924	-0.5445	-0.4199	-0.2715	-0.1840
	0.1	0.8602	0.6963	0.5923	-0.5433	-0.4199	-0.2713	-0.1838
Amenity Index	0	0.0135	0.0035	0.0013	-0.0017	-0.0009	-0.0003	-0.0002
	0.01	0.0135	0.0035	0.0013	-0.0018	-0.0009	-0.0004	-0.0002
	0.1	0.0136	0.0035	0.0014	-0.0027	-0.0009	-0.0009	-0.0007
Welfare	0	2.3096	1.2638	0.8949	-0.7273	-0.4827	-0.2680	-0.1669
	0.01	2.3096	1.2638	0.8949	-0.7273	-0.4827	-0.2680	-0.1669
	0.1	2.3096	1.2638	0.8950	-0.7272	-0.4827	-0.2681	-0.1670
				Rezoning Counterfactual				
Sqm Price	0	-0.1065	-0.1068	-0.1073	-0.1068	-0.1091	-0.1152	-0.1178
	0.01	-0.1064	-0.1068	-0.1073	-0.1068	-0.1091	-0.1153	-0.1176
	0.1	-0.1065	-0.1073	-0.1080	-0.1058	-0.1092	-0.1164	-0.1143
Floor Space	0	0.0419	0.0527	0.0604	0.0669	0.0755	0.0897	0.0944
	0.01	0.0419	0.0527	0.0604	0.0669	0.0755	0.0897	0.0942
	0.1	0.0420	0.0530	0.0608	0.0660	0.0754	0.0903	0.0907
Amenity Index	0	-0.0391	-0.0337	-0.0325	-0.0296	-0.0320	-0.0343	-0.0497
	0.01	-0.0389	-0.0336	-0.0324	-0.0296	-0.0319	-0.0345	-0.0470
	0.1	-0.0376	-0.0329	-0.0320	-0.0275	-0.0291	-0.0340	-0.0178
Welfare	0	0.1422	0.1257	0.1522	0.1168	0.1373	0.1735	0.4804
	0.01	0.1422	0.1256	0.1522	0.1132	0.1385	0.1737	0.4590
	0.1	0.1421	0.1255	0.1521	0.0807	0.1498	0.1749	0.3233

Source: Own model calculations. Table 5.B.2 shows main counterfactual results for different values of ζ^H.

Table 5.B.3: Different Values of ζ^L

Change in %	ζ^H	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7
				Gentrification Counterfactual				
Sqm Price	0	1.5252	1.6080	1.7270	1.8641	2.4405	2.5586	3.2510
	0.1	1.5301	1.6147	1.7272	1.8633	2.4403	2.5598	3.2509
Floor Space	0	-0.5584	-0.5465	-0.7134	-0.9880	-1.4735	-1.7288	-2.4177
	0.1	-0.5602	-0.5495	-0.7135	-0.9871	-1.4731	-1.7294	-2.4173
Amenity Index	0	-0.0567	-0.0265	-0.0135	0.1579	0.1922	0.1913	0.2014
	0.1	1.5434	1.5856	1.6026	0.1574	0.1919	0.1910	0.2013
Welfare	0	-0.5451	-0.0155	-0.0053	-0.0734	-0.0289	0.0257	0.0197
	0.1	-0.1934	0.3340	0.3458	-0.0733	-0.0289	0.0254	0.0196
				Rental Cap Counterfactual				
Sqm Price	0	-1.9879	-1.5556	-1.5144	-1.5049	-1.7643	-2.4070	-2.5469
	0.1	-1.9885	-1.5481	-1.5198	-1.5025	-1.7620	-2.4050	-2.5449
Floor Space	0	-1.0973	-3.5945	-3.8701	-3.9099	-4.3408	-5.4621	-5.3158
	0.1	-1.0926	-3.6290	-3.8163	-3.9106	-4.3386	-5.4560	-5.3095
Amenity Index	0	-0.9403	-0.3070	-0.2434	-0.1126	-0.0430	-0.0018	-0.0029
	0.1	-0.9831	-0.2969	-0.2564	-0.1116	-0.0422	-0.0013	-0.0027
Welfare	0	-2.6153	-0.8689	-0.5720	-0.3103	-0.2639	-0.1898	-0.2209
	0.1	-2.5413	-0.8669	-0.5392	-0.3085	-0.2625	-0.1883	-0.2197
				Subsidy Counterfactual				
Sqm Price	0	-5.4534	-4.4086	-3.7484	-0.0031	-0.0003	0.0017	0.0020
	0.1	-5.4528	-4.4090	-3.7478	-0.0031	-0.0003	0.0018	0.0021
Floor Space	0	0.8603	0.6965	0.5924	-0.5445	-0.4199	-0.2715	-0.1840
	0.1	0.8601	0.6967	0.5921	-0.5445	-0.4199	-0.2715	-0.1840
Amenity Index	0	0.0135	0.0035	0.0013	-0.0018	-0.0009	-0.0004	-0.0002
	0.1	0.0140	0.0032	0.0018	-0.0018	-0.0009	-0.0004	-0.0002
Welfare	0	2.3096	1.2638	0.8949	-0.7273	-0.4827	-0.2680	-0.1669
	0.1	2.3096	1.2639	0.8950	-0.7273	-0.4827	-0.2680	-0.1669
				Rezoning Counterfactual				
Sqm Price	0	-0.1064	-0.1068	-0.1073	-0.1068	-0.1091	-0.1153	-0.1176
	0.1	-0.1070	-0.1076	-0.1078	-0.1079	-0.1089	-0.1152	-0.1135
Floor Space	0	0.0419	0.0527	0.0604	0.0669	0.0755	0.0897	0.0942
	0.1	0.0420	0.0530	0.0606	0.0677	0.0754	0.0895	0.0904
Amenity Index	0	-0.0389	-0.0336	-0.0324	-0.0296	-0.0319	-0.0345	-0.0470
	0.1	-0.0288	-0.0261	-0.0274	-0.0293	-0.0317	-0.0343	-0.0468
Welfare	0	0.1422	0.1256	0.1522	0.1132	0.1385	0.1737	0.4590
	0.1	0.1435	0.1350	0.1190	0.1133	0.1384	0.1735	0.4583

Source: Own model calculations. Table 5.B.2 shows main counterfactual results for different values of ζ^L.

Chapter 6

Ticket to Paradise? The Effect of a Public Transport Subsidy on Air Quality.[1]

Abstract

This paper provides novel evidence on the impact of public transport subsidies on air pollution. We obtain causal estimates by leveraging a unique policy intervention in Germany that temporarily reduced nationwide prices for regional public transport to a monthly flat rate price of 9 Euros. Using DiD estimation strategies on air pollutant data, we show that this intervention causally reduced a benchmark air pollution index by more than eight percent and, after its termination, increased again. Our results illustrate that public transport subsidies - especially in the context of spatially constrained cities - offer a viable alternative for policymakers and city planers to improve air quality, which has been shown to crucially affect health outcomes.

[1]Co-authored with Philipp Schrauth.

6.1 Introduction

The UN's Sustainable Development Goals emphasize the importance of air quality within cities and explicitly list the improvement of urban air quality as a key measure to make agglomerations safer, healthier, more resilient and sustainable.[2] Key contributing factors to air pollution, particularly in cities, are car traffic and congestion. Indeed, research demonstrates that reduced congestion and traffic can lead to better air quality and subsequently improved health outcomes (Margaryan, 2021; Knittel et al., 2016; Currie and Walker, 2011). Frequently proposed measures to reduce automobile use and air pollution include the extension of public transport supply and the reduction of its cost, ultimately encouraging a modal switch towards publicly provided means of transportation. This paper leverages a unique policy intervention in Germany to causally estimate how lowering public transport prices impacts air quality.

In June 2022, Germany introduced the "9-Euro-Ticket" (9ET), which temporarily reduced regional public transportation fares to a flat rate price of 9 Euros per month from June until August. This policy intervention is unusual in the sense that it temporarily and relatively suddenly reduced public transport fares for a whole country. Further, the implied price drop was substantial - in Berlin, for example, the price of the monthly standard ticket, which normally costs 86 Euros, experienced a decrease of about 90%.[3] The aim of the ticket's introduction was twofold. First, as part of a set of relief measures, passed in the first half of 2022, mitigating rising costs of living was a central goal of the policy. Second, it was considered a potential means to promote environmentally friendly mobility by incentivising a reduction in car traffic and thereby carbon emissions and pollution (Spiegel Online, 2022).

A priori the effect of a public transport price reduction on air pollution is not clear. If individuals switch from car travel to public transport there might indeed be a fall in air pollution. However, if individuals predominantly replace walks and bike trips with public transport or seek to avoid crowded trains, there might be no effect at all. Conclusive and

[2] See Goal 11 at https://sdgs.un.org/goals.

[3] See `https://www.bvg.de/de/tickets-und-tarife/alle-tickets/zeitkarten/monatskarte`.

causal evidence directly studying the impact of a public transport subsidy does not exist. Our paper closes this gap by providing evidence that a price reduction indeed can reduce air pollution, at least for the short time-period considered in our scenario. Further, in contrast to previous research on the relation of public transport and air quality we provide evidence for a whole country, Germany, and not only selected cities.

In order to causally analyse whether the described substantial reduction in public transport prices indeed has an impact on air pollution, we adopt a difference-in-differences (DiD) approach. More precisely, the empirical setting compares changes in air pollution between months May and June in non-treatment years before 2022 to changes in pollution in the same two months in 2022. We construct a state of the art air quality index (AQI) and show that it decreases by more than eight percent as a consequence of the introduction of the ticket. After the end of the policy measure, air pollution increased again. Further, we document substantial effect heterogeneity and show that the effect is largest in urban areas, during work days and in areas with high levels of public transport provision. Ultimately, back-of-the-envelope calculations show that potential health benefits associated with a reduction in air pollution have the potential to exceed the costs of the intervention.

Our results are in line with recent findings that descriptively document a modal switch in response to the introduction of the subsidy. The 9ET itself was sold over 52 million times and survey evidence on individuals' mobility patterns clearly documents a strong increase in the use of public transport and a decrease in other modes of transportation in response to the ticket's introduction (DLR, 2022; VDV, 2022). For example, in a survey of 6,000 individuals by the German Public Transport Association, ten percent of 9ET owners said they avoided at least one of their daily car journeys by using the ticket (VDV, 2022).[4] Further, studies on the 9ET's impact on traffic flows support that car traffic was indeed reduced: following the introduction of the ticket, the level of congestion in many German cities decreased between May and June 2022, while the amount of train travel

[4]Based on calculations, there is also evidence showing that the 9ET led to significant reductions in CO_2 emissions, albeit the size of the effect differs depending on the respective study and the underlying assumptions (VDV, 2022; INFAS, 2023).

increased (Süddeutsche Zeitung, 2022; German Federal Office for Statistics, 2022). We supplement these findings with a case study of traffic developments for the city of Berlin supporting the notion that car travel decreased.

Expanding this descriptive evidence, our paper, to the best of our knowledge, is the first to causally show that cheaper public transport can imply positive externalities such as reduced air pollution. Thus, in the absence of effective first best road pricing systems that internalize the externalities of traffic, large subsidies on ticket prices might be a viable alternative solution.[5]

Our research adds to a large body of literature that shows a positive impact of public transit provision on air quality. Anderson (2014) and Bauernschuster et al. (2017) use unexpected public transport strikes as a quasi-natural experiment and find a temporary increase in congestion and in air pollution respectively. Other studies look at the effect of building or extending new subway lines, which have been shown to improve air quality (Chen and Whalley, 2012; Gendron-Carrier et al., 2022). For Germany, Lalive et al. (2018) find a positive effect of railway expansions on air quality. A second strand of literature develops and calibrates quantitative equilibrium models to quantify welfare effects of public transport investments. For instance, Parry and Small (2009) and Basso and Silva (2014) document significant welfare gains from subsidizing public transport. Borck (2019), in a counterfactual analysis, shows that subsidies reduce air pollution, albeit only modestly due to offsetting long term equilibrium effects such as residential relocation. In contrast, causal reduced form evidence on the effect of public transport pricing on air quality is, to the best of our knowledge, scarce and limited to the study of price increases. Yang and Tang (2018) using a synthetic control group approach paired with DiD analysis find that an increase in Beijing's public transport fare led to a short run increase in air pollution of approximately 16 percent. Fully understanding how public transport prices and in particular a price reduction via subsidized tickets impact

[5]There was substantial critical press coverage regarding the 9ET, especially that it was too much of a fiscal burden and led to over-crowded trains (see e.g. TAZ, 2022; Münchener Merkur, 2022). Indeed, there are potentially more efficient ways of addressing the external effects of traffic (such as low emission zones or congestion charges) other than strengthening public transport via price subsidies, which also provides cheaper tickets to those already using public transport regularly.

traffic patterns and consequently air quality, however, is crucial. Price adjustments, as opposed to infrastructural interventions, might offer an easier to implement measure for policymakers to reduce air pollution - especially in spatially constrained cities where extending existing networks or constructing new ones might not be an option.

Overall, our results may have important policy and health implications: they show that subsidizing public transportation might be a viable option to reduce air pollution particularly in cities, thereby contributing to the UN's sustainability goal of creating more resilient, safer and healthier urban agglomerations. For instance, reductions in air pollution have been shown to substantially reduce cardiovascular diagnoses (Margaryan, 2021) and improve infant health (Knittel et al., 2016).

The remainder of the paper is structured as followed. Section 6.2 provides the relevant institutional background. Section 6.3 introduces the different data sources used for estimation and presents the estimation approach. Section 6.4 presents and discusses the results and Section 6.5 concludes.

6.2 Institutional Background

The 2021 general election in Germany resulted in a new coalition government of Social Democrats (SPD), the Greens and the Liberals (FDP). The creation of a sustainable mobility sector was one of the topics highlighted in the new government coalition agreement.[6] Since 2022, and particularly the start of the war in Ukraine, living costs - especially in the form of energy prices - increased drastically. This prompted the government to pass a set of relief measures aiming to mitigate such costs and prices. The 9ET was part of a second set of such relief measures seeking to, amongst others, maintain affordable means of transportation. Additionally, in particular the co-governing Green party frequently emphasised the ticket's potential to reduce carbon emissions and provide a more sustainable alternative for the mobility sector, as stipulated by the coalition agreement (Spiegel Online, 2022).

[6] See Coalition Agreement 2021.

On February 23rd, 2022 a first relief package consisting of several relief measures such as temporary tax cuts was decided upon by the ruling coalition parties. After the start of the war in Ukraine one day later and the consequent steep increase in gas and oil prices, a second relief package was passed in the German Bundestag. This second set of relief measures included a reduction of the energy tax on fuel in order to help people commuting by cars. Figure 6.A.1 in the Appendix depicts the development of gasoline and diesel prices in 2022. The beginning of the war in Ukraine was followed by a steep increase of all types of fuel prices. After reaching a peak at the beginning of March, the price stabilized at moderately lower levels until June 1st. The energy tax implied a temporary drop in fuel prices. However, diesel regained pre-tax prices just after a few days. Overall, prices for all types of fuel remained at a relatively high level especially compared to pre-war times. Crucially, in our identification strategy presented in the following section we can control for daily fuel prices to account for these patterns and the tax cut.

In order to additionally compensate users of public transportation, the relief package also included the 9ET. For a total price of nine Euros a month, it allows its holders to use most types of public transportation like buses, subways, and regional trains all over Germany. The ticket was available from June 1st, 2022 until the end of August of 2022.[7] One of the aims of subsidizing public transportation was to foster its utilization and thereby to speed up sustainable transportation. Just before June 1st, about seven million tickets and by the end of June approximately 21 million tickets had been sold (Handelsblatt, 2022; Süddeutsche Zeitung, 2022). Factoring in the roughly 9 million regular subscribers to monthly or yearly tickets whose fare is automatically reduced to 9 Euros from June to August, more than thirty million 9ETs were in circulation by the end of June.

[7] See https://www.bundesfinanzministerium.de/....html for more information on the relief packages.

6.3 Data and Empirical Approach

In order to analyse the 9ET's impact on air pollution, we collect pollution measurements for four key pollutants and data on covariates that have been shown to influence air pollution, such as weather conditions and holidays. Crucially, we also collect data on fuel prices allowing us to fully account for the tax cut in fuel prices that was simultaneously introduced with the 9ET. The remainder of this section firstly presents our different data sources to then introduce our empirical approach.

6.3.1 Data

Air pollution data We use air pollution data for months May until September from 2018 to 2022. The data is provided as hourly measures by the Federal Environmental Agency (FDA) (Umweltbundesamt, 2022).[8] We observe whether a measuring station is located close to a street (traffic station) or in residential areas (background station). While the former provides information on air quality in relation to traffic, the latter rather indicates the general quality of the air in an area. We make use of this differentiation in our estimations: in our main specification we exclusively focus on air pollution concentrations measured by traffic stations, in order to test whether it is indeed the reduction in car traffic that reduced air pollution following the 9ET. In a heterogeneity check, we exclusively focus on background stations where we expect the potential effect of the ticket to be smaller.[9] Further, the data includes information on whether the station lies in rural, suburban or urban areas.[10]

In order to assess air quality we look at the pollutants nitrogen dioxide (NO_2), particu-

[8]The FDA API provides the data: https://www.umweltbundesamt.de/daten/luft/luftdaten/doc.

[9]In our data set, 411 ground-level stations measure the concentration of NO_2, 360 stations measure PM_{10} and 273 $PM_{2.5}$. One of the reasons why there are fewer measuring stations for $PM_{2.5}$ is that the coverage of this pollutant has only started relatively recently and the measuring net is still being extended.

[10]The representativeness of the measuring spots is required by European law, as described by the FDA: https://www.umweltbundesamt.de/themen/luftmessnetz-wo-wie-wird-gemessen. The requirement of representative placement of air quality measuring stations was examined and confirmed by an external testing organization (TÜV Rheinland, 2019). Moreover, in the data we find that the number of measuring stations is largely proportional to population between districts.

late matter with diameter less than 10 (PM_{10}) micrometers or smaller than 2.5 ($PM_{2.5}$) micrometers, as well as ozone (O_3). These we use to construct the AQI, commonly used by national and international authorities in order to assess pollution levels and provide information on potential health impacts at local levels. For the construction of the AQI, we follow a benchmark European air quality index (Van den Elshout et al., 2014), which is based on the core pollutants NO_2, PM_{10}, and $PM_{2.5}$ for traffic stations (our key outcome) and additionally O_3 for background stations. The index itself takes on values between 0 and 100, which is then further classified into four categories from very low pollution (index number 0-25) to high pollution (index numbers 75-100).[11]

Fuel prices We collect data on local fuel prices. Since 2013, fuel stations have been obliged to report each and every change of fuel prices (specifically for diesel and gasoline) in real time to the Market Transparency Office for fuels ("Markttransparenzstelle für Kraftstoffe") run by the German Cartel Office.[12] We aggregate all fuel prices to district level by taking the daily mean of all stations within each administrative entity.[13]

Meteorology When analysing air quality data, it is critical to control for current weather conditions. Wind and rain tend to improve air quality, since they clean the air e.g. from pollutants such as particulates. We use weather data aggregated to daily levels to control for mean temperature, mean wind speed and total precipitation levels. In order to do so, we acquired measurements from about 3.000 stations from the German Meteorological Service (DWD Climate Data Center (CDC), 2022). Since measurement stations are independently located from air quality measuring stations, we follow the approach by Auffhammer and Kellogg (2011) to match each air quality to the nearest weather station. Through this method, we are able to match about 99 percent of pollution observations to the weather variables of interest.

[11]For a technical description for the construction of the AQI see table 4 in Van den Elshout et al. (2014).

[12]For more information see https://www.bundeskartellamt.de/EN/Economicsectors/MineralOil/...

[13]The price data is provided on the following website: https://creativecommons.tankerkoenig.de/.

Further controls We control for a variety of potential economic and traffic-related variables. In particular, the German Federal Office for Statistics made available district level data of the total mileage of regional and local trains as well as per capita bus provision. We use this information in our heterogeneity analysis to check whether effects differ across districts with low and high levels of public transport provision.

Furthermore, we know whether a day falls on a weekend, a public holiday or on school vacations.[14] In order to be able to differentiate between effects in urban areas including their commuting zones, we resort to OECD definitions of functional urban areas (FUAs).[15] Those are constructed based on commuters' daily movements and they consist of a city centre to which people commute to as well as the surrounding commuting zone (Dijkstra et al., 2022). In Germany, there are 96 FUAs in total.

6.3.2 Empirical Approach

We use a "time-shifted" DiD design that was also recently used in Hall and Madsen (2022); Brodeur et al. (2021); Metcalfe et al. (2011). Unlike traditional difference-in-differences estimators, such approaches use two time dimensions instead of one, i.e., in addition to a post-treatment and a pre-treatment period as used in standard DiD approaches, there is also a treatment and a control period instead of the traditional treatment and control groups. In our main setting, we compare air quality in the treatment period (year 2022) and the control period (years 2018/19) for the month of June after the policy and the month of May before the policy. Thus, the approach essentially compares the difference in air quality in May between 2018/19 and 2022 with the difference in air pollution between June 2018/19 and 2022.

We explicitly exclude years 2020 and 2021, as due to changes of COVID 19 restrictions in May and June 2020 and 2021 respectively there might be confounding factors influ-

[14]Information about state-level holidays and vacations are retrieved from http://www.feiertage-api.de/ and https://ferien-api.de/ respectively.

[15]Shapefiles can be downloaded from https://www.oecd.org/regional/regional-statistics/functional-urban-areas.htm.

encing air pollution outcomes in these years.[16] We focus on working days outside of school holidays, as unusual traffic patterns are common during these. Further, there is more construction work during school holidays, which can directly affect air quality and traffic.[17] Equation 6.1 describes the approach in more detail.

$$p_{symd} = \alpha_s + \gamma_{yf} + \eta_{mf} + \zeta_d + \beta Post_m \cdot \kappa_{2022} + X'_{symd}\theta + \epsilon_{symd} \tag{6.1}$$

p_{symd} is the logarithm of the AQI measured at station s, in year y for month m, at day of the week d. We regress this outcome on station level fixed effects, α_s, controlling for all potential time invariant station-specific observables and unobservables such as the number of lanes the station is placed at or the distance to rail tracks. Further, we include federal state specific year fixed effects, γ_{yf}, to control for general differences across years in each federal state f and federal state specific month fixed effects η_{mf} to account for differences in pre-treatment and post-treatment months within each federal state.

We also control for day-of-the-week fixed effects, ζ_d. This allows us to only compare pollutant levels at the same day of the week with one another. This is necessary as there are pollution trends across days of the week and thus comparing pollutant levels on a Wednesday and a Friday across years might falsely pick up a difference in pollution between years that can simply be attributed to these trends.[18] We then include a dummy variable $Post_m$, which is equal to one for the post-policy month, i.e. June. Interacting the Post indicator with an indicator variable for the treatment year 2022, i.e. κ_{2022}, then gives the DiD estimate for coefficient β. This setup allows us to test whether there indeed is a change in air quality between May and June in 2022, i.e. the treatment year when the 9ET was introduced, that goes beyond the changes observed between these two months in previous years that were not subject to such a policy intervention.

[16]For example, in most federal states in-class teaching only restarted in late May/June in 2021 and mid and late May in 2020.

[17]In addition, school holidays in each year and in each federal state start at different times, which makes controlling for the described patterns difficult.

[18]For example, on Fridays there may be a higher tendency to work from home.

In addition, we control for a matrix of covariates, X', including linear and squared daily weather conditions in the vicinity of each pollution station such as wind speed, precipitation and temperatures, as these have been shown to influence pollution levels (see e.g. Auffhammer and Kellogg, 2011). Further, we control for daily average fuel prices and interact the fuel prices with the year fixed effects in order to account for different effects across years. Controlling for fuel prices is crucial as a fuel tax cut was introduced jointly with the 9ET (see Section 6.2). As detailed above, fuel prices fell only moderately in response to the tax cut compared to the yearly trend (see Figure 6.A.1).

In the robustness checks of the paper we also present specifications excluding the first days in June and the last days in May to account for the temporary fall in gasoline prices and potential anticipation effects, which could entail that individuals drove less before June in order to wait for a fall in gas and public transport prices. In our main specifications, the error term ϵ_{symd} is clustered at the district level.

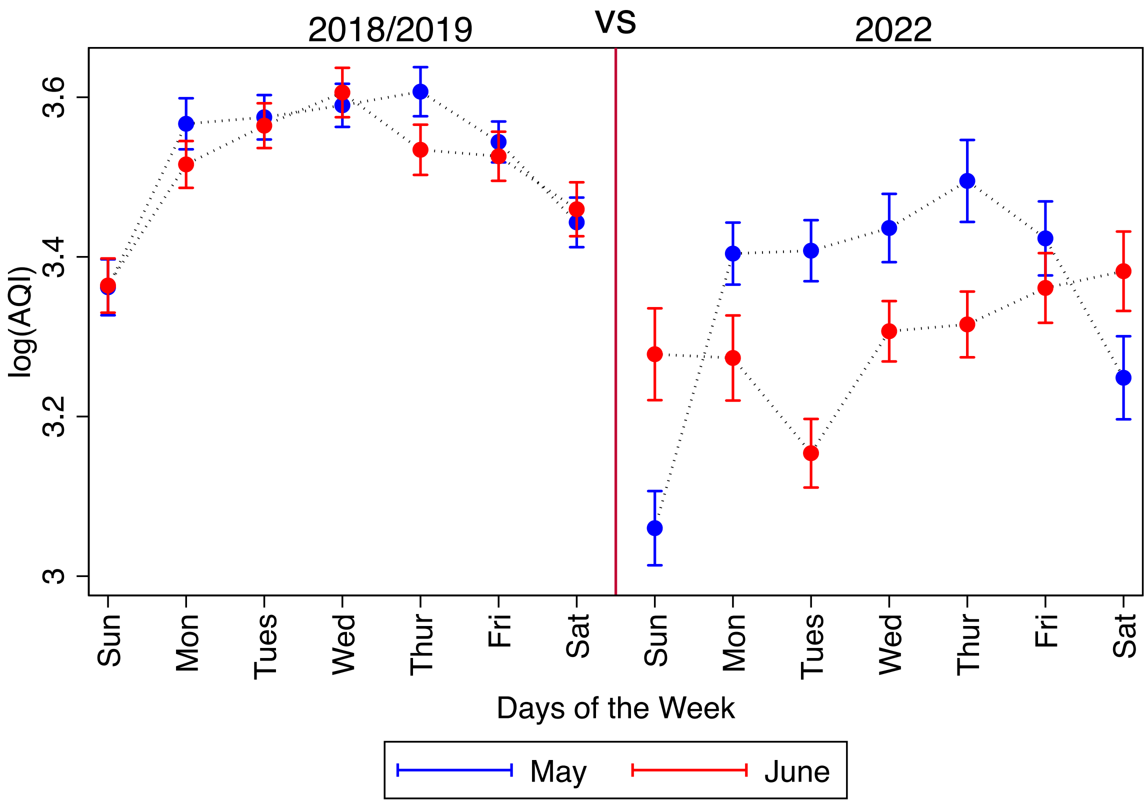
The DiD approach identifies the causal effect of the 9ET on air pollution as long as pollutant levels in pre-policy month May and post-policy month June would have developed parallely between the treatment year 2022 and the control years 2018-2019 in the absence of treatment and conditional on the observed covariates and fixed effects. In the following section we check this assumption by carefully examining the common pre-trend assumption, i.e. average air pollution levels in May for 2022 and May 2018/2019.

In the main analyses, we primarily focus on the months May and June. However, in order to present time patterns of the subsidy's impact, we subsequently provide estimates including June, July and August as part of the post-policy period and also estimate the effect between August and September, when the measure ended.

6.4 Results

Descriptive Results First, we compare average values for a given day of the week in 2018/2019 and 2022 for months May and June. Figure 6.1 shows the day of the week

Figure 6.1: Day of the week averages for log(AQI).



Note: Day of the week averages for log(AQI) and 99 % confidence intervals in June and May for treatment (2022) vs. control year (2018/2019), excluding school and public holidays. Average working day values and differences for May and June before and after the reform are: Monday to Friday 2018/19: June= 3.55, May= 3.58, Diff= −0.03, p-value= 0.0002; 2022: June= 3.29, May= 3.43, Diff= −0.14, p-value= 0.0000.The p-values are derived from a two-sided t-test on the difference between both values. Corresponding values for weekends are 2018/19: June= 3.411, May= 3.400, Diff= −0.011, p-value= 0.39; 2022: June= 3.331, May= 3.169, Diff= −0.162, p-value= 0.0000.

specific average for the logarithm of the AQI for the control years, i.e. 2018/19 (left), and the treatment period (right), i.e. the year 2022. The graph illustrates that when assessing air quality trends it is crucial to only compare pollutant levels at the same day of the week with one another, as there are clear general patterns in the average pollutant levels for a certain day of the week. For example, there generally appears to be a greater level of pollution in the middle of the week and lower levels over weekends.

Figure 6.1 shows that in the years 2018/19 log AQI values followed a relatively parallel trajectory across the days of the week in May and June. Further, the log index values were relatively similar in size. For the year 2022, there is an obvious divergence in the week day trajectory of the index values for days in May and June. In particular, during the week there appears to be a sharp decline in the log of the AQI in 2022, which suggests an improvement in air quality. This is further confirmed by comparing the average log values of the AQI for working days, i.e. Monday to Friday, in May and June in the year 2022, as presented in the table notes: average air pollution is 0.14 log points smaller in June than in May 2022 - a notable difference compared to 2018/19, where the average pollution on working days in June was 0.03 log points lower than in May.

In contrast, weekend average values in June in comparison to May seem to increase in treatment year 2022. However, this effect seems to be driven by factors that we control for in the main analyses, as we fail to find significant air quality changes for weekends in our regression approach.

Figure 6.1 can also be interpreted as direct support for our identification assumption: during the pre-treatment period, i.e., in May, log AQI values for treatment (2022) and control (2018/2019) years behaved relatively similarly and followed a parallel trajectory throughout the week, albeit at different levels. To illustrate this more clearly, Figure 6.A.2 in the Appendix shows average week day values in May separately for 2022 and 2018/19. Log AQI values for both periods show similar patterns with clear peaks on Thursdays. Overall, air quality levels follow a parallel trajectory prior to the introduction of the 9ET in June 2022, implying that in the absence of the introduction of the public transport subsidy, air quality in June 2022 likely would have developed parallelly to air quality in June 2018/19.

Main Results: Table 6.1 presents the results using log(AQI) as an outcome variable. In all specifications we control for the full set of fixed effects and covariates described above. Specification (1) shows the results for the months May and June. We find a negative point estimate for the interaction term suggesting that the average AQI fell by more than eight percent. The estimated effect is significant at the one percent level. Since an index value of 0 indicates the best possible air quality, the result suggests an overall average improvement in air quality. In specification (2) and (3), we include, apart

Table 6.1: Main Results: log(AQI)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interaction	-0.0877***	-0.0833***	-0.0630***	-0.0089	-0.0125	-0.0230	0.0665**
	(0.0207)	(0.0200)	(0.0157)	(0.0172)	(0.0170)	(0.0168)	(0.0314)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observed Years	18,19,22	18,19,22	18,19,22	18,19	18,19	18,19	18,19,22
Included Months	May-June	May-July	May-August	May-June	May-July	May-August	August-Sep
Observations	21926	26485	30737	14523	17093	20413	15054

Note: Source: own calculations. Standard errors in parentheses and clustered at district level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the log of the AQI as outcome variable. Column (1) shows the results for a basic DiD approach only including months May and June. Columns (2)-(3) augment the estimation with subsequently adding additional months. Columns (4) - (6) implement placebo tests where year 2019 is used as the new policy date. Column (7) shows the effect of abolishing the 9ET. All specifications control for the full set of covariates and fixed effects.

from the month of June, the months July and August respectively. Point estimates in both specifications remain negative and in case of specification (2) relatively similar to the main specification. When including August, there is a clear decrease in absolute size of the point estimate, suggesting a fall in the 9ET's impact on pollution. One potential explanation here might be a switch back from public transport towards other means of transportation after initially trying public means of transportation in June.

In columns (4)-(6) we repeat the estimations for the years 2018 and 2019 only. Crucially, we treat year 2019 as if the 9ET was introduced in June 2019. This allows us to run a placebo test, explicitly testing whether our results are driven by general differences in air pollution between May and June between years. The point estimates are negative and insignificant in each specification, thereby supporting the notion that there is no systematic difference between months May and June in untreated years previous to the actual introduction of the ticket.

Lastly, in column (7) we analyse whether the abolishment of the ticket resulted in an increase in air pollution using August as the control month and September as the treatment month.[19] The positive point estimate confirms that air pollution was significantly

[19] August and September coincide with the end of school holidays in many federal states. Thus there are fewer observations stemming from non-holiday days, which we explicitly use in our estimation.

lower in months the 9ET was in place compared to times with regular public transport pricing.[20]

All in all the results indicate a substantial fall in the AQI of more than eight percent and suggest a decrease in the effect over time.

Mechanisms and Heterogeneous Results In a next step, we split our sample along several dimensions to analyse heterogeneous effects and potential mechanisms. Table 6.2 depicts the results. All specifications control for the full set of covariates and fixed effects and use observations from months May and June. First, in columns (1)-(4) we split our sample into the core of a functional urban area (FUA), the total FUA and non-FUA (rural) areas.[21] The effect is driven by a substantial reduction in pollution in core areas. One potential explanation is that most jobs are situated in core areas. If more individuals commute to these core areas by public transport there will be less traffic and congestion in these areas implying a fall in pollution concentration levels. Another possible explanation is that there is more supply of public transport in these areas, which would then allow for an easier switch to public means of transportation. In the Online Appendix 6.B, we lay out these arguments in more detail by providing more supportive evidence for the underlying mechanisms driving our results. In particular, we analyse the 9ET's effect on traffic in a large metropolitan area and document a substantial fall in motorized traffic. Further, we provide more information on public transport ridership following the introduction of the 9ET.

In order to generally assess the role of public transport supply, we split our sample into districts with a relatively high and low level of public transport infrastructure (see columns (5)-(6)). We can do so by using the aforementioned data on public transport kilometres per population for districts in Germany. More precisely, we split our sample

[20]Using August as the treatment group and September as the control group or vice versa might imply that our estimation will also pick up potential spillover effects, undermining our identification strategy. For example, if the introduction of the ticket managed to incentivize lasting behavioural change, some individuals might have permanently switched towards public transport. This in turn would impact air pollution in September as well as August 2022, ultimately violating the Stable Unit Treatment Value Assumption (SUTVA) commonly needed in DiD approaches.

[21]The total FUA includes the core as well as the commuting zone of a FUA.

Table 6.2: Heterogeneous Results

	(1) City cores	(2) Outside city cores	(3) Urban areas	(4) Non-urban areas	(5) High pub. trans. share	(6) Low pub. trans. share	(7) Weekend	(8) Weekday	(9) Background stations
Interaction	-0.1095***	-0.0062	-0.0985***	-0.0172	-0.1019**	-0.0836***	0.0101	-0.1244***	-0.0241**
	(0.0208)	(0.0441)	(0.0206)	(0.0690)	(0.0418)	(0.0242)	(0.0391)	(0.0218)	(0.0113)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18001	3925	19894	2032	5554	16372	6211	15715	42855

Note: Source: own calculations. Standard errors in parentheses and clustered at district level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the log of the AQI as outcome variable. Column (1) shows the results for stations from the core of functional urban areas, column(2) for stations positioned outside the core. Specification (3) just includes observations from functional urban areas and specification (4) from outside these areas. (5) solely uses measurements from stations in districts in the highest 25 percentile of public transport supply and (6) in districts in the lowest 75 percentiles of public transport kilometres per person. (7) and (8) split the sample into weekends and weekdays and (9) solely includes measurements from background stations. All specifications control for the full set of covariates and fixed effects.

into districts in the highest 25 percentiles of public transport kilometres per person and the lowest 75 percentiles. The results show that for both groups we find significant effects. As expected the effect size is more pronounced for the group with higher levels of public transportation.[22]

Next, in columns (7)-(8) we confirm what the visual examination above indicated: the effect is driven exclusively by weekdays. For weekends we find a positive, albeit small and insignificant effect when including the full set of covariates and fixed effects.

Lastly, in our data we can differentiate between air quality measured by stations directly next to roads and background measurements in quieter areas. If there indeed is less traffic due to more individuals using public transport, the effect at stations directly exposed to traffic should be stronger. In our main specifications we have so far exclusively focused on traffic stations. In column (9) of Table 6.2 we repeat our estimation for background stations. Only using background stations we would expect a smaller effect size, as these stations are not directly positioned next to streets. The results confirm this notion, as the point estimates remains negative and significant, albeit relatively small in absolute

[22]We would have liked to differentiate treatment intensity by accounting for local price changes to provide a more comprehensive picture. However, due to the decentralized supply structure of public transport services in Germany, gathering the respective data from each local provider was beyond our capabilities.

size.

Robustness In the Appendix we repeat our main estimation using AQI values instead of log AQI values as outcome variable. Table 6.A.1 in the appendix shows that for our main specification using May and July, the index value drops by more than 2.3 points. Next, we repeat our main estimation allowing for different standard error clusters, i.e. at the station specific level and at the station-year level. Tables 6.A.2 and 6.A.3 show the results. Inference and statistical significance do not change across these different clusters. We also repeat our main estimation this time including station specific year and day of the week fixed effects in order to account for potential differences in yearly trends and weekday differences across stations. Table 6.A.4 in the Online Appendix shows the results. All point estimates remain highly significant and our main inference does not change.[23]

Further, as detailed in Section 6.2, in addition to the 9ET, a tax cut on gasoline prices was implemented from June, 1st. The corresponding law, however, was already passed on May 19th, 2022. Car owners thus might have waited for June 1st in order to purchase new fuel and drive again. As a consequence pollution levels directly before June 1st might have been slightly dampened and heightened at the beginning of June, leading to a lower estimated decrease in pollutants after the introduction of the 9ET. In order to analyse these potential anticipation effects in the run up to June 1st, we repeat our main estimation for the log(AQI) in a leave-one-out exercise. Here we successively drop the first and last two days in May and June respectively. The results, presented in Table 6.A.5 in the Appendix, indicate no substantial difference in the point estimates for the variable of interest, which remain very similar in size across specifications.

Lastly, we use each pollutant (and its logarithm) that contributes to the AQI as a outcome variable separately in order to check that it is not simply the construction method of the AQI that drives the results and to ease the interpretation of results in the para-

[23]For specifications (1) and (2) point estimates become larger than in the main approach, and for the specification additionally including August the point estimate becomes smaller. Overall, the inference, however, does not change.

graph below. Table 6.A.6 in the Appendix shows the results for PM_{10}, $PM_{2.5}$ and NO_2 respectively. All point estimates are significant and negative supporting the results in the main specification using the AQI as the key measure for air quality.

Implications Overall, our results document a decrease in air pollution in response to the 9ET, which might impact further outcomes such as individuals' health. Indeed, the negative relationship between air pollution and health hazards is a well established fact (Anderson, 2009). Generally, AQI values below 50 are categorized as an overall "good" air quality (Van den Elshout et al., 2014). While the average AQI value in Germany is about 35, and thus an eight percent increase still leaves us within the range of good air quality, there are nevertheless numerous studies that find significant health effects even at moderate levels of air quality deterioration. For instance, the introduction of low emission zones (LEZ) in Germany lowered PM_{10} concentrations by about 3 percent or approximately 0.8 $\mu g/m^3$, which consequently reduced the number of patients with cardiovascular diagnoses by approximately 2-3 percent (Margaryan, 2021). According to back-of-the-envelope calculations, this resulted in economic welfare gains of more than 4.4 billion Euros (Margaryan, 2021). Assuming the relationship between cardiovascular health problems and PM_{10} to be linear, this would imply an even larger effect of the 9ET, as PM_{10} was reduced by more than 1.5 $\mu g/m^3$ following the introduction of the ticket (see Table 6.A.6).

Further research suggests a significant impact of pollution reductions on infant health and mortality, even at levels below common thresholds of concern (Simeonova et al., 2021; Knittel et al., 2016; Currie et al., 2011). For example, PM_{10} reductions of 1 $\mu g/m^3$ have been shown to imply ten saved lives per 100,000 births (Knittel et al., 2016). Taking the value of a statistical life of 1.72 million Euro in Germany, this would imply saved costs of about 200 million Euros per year, given total births of 800,000 and our finding of a 1.5 $\mu g/m^3$ reduction in PM_{10}.

While such back-of-the-envelope calculations are based on highly simplified assumptions, they nevertheless indicate that positive health implications alone could have the potential to amortise the costs associated with the ticket of approximately 2.5 billion Euro.[24] This is especially the case since the exemplary calculations conducted above ignore a range of further health hazards caused by pollution, e.g. pulmonary diseases, lung damages, respiratory distress, or birth defects (for an extensive list of potential health effects only caused by particulate pollution, see Pope III and Dockery, 2006).

6.5 Conclusion

In this paper, we provide novel causal evidence on the effect of a large scale public transport subsidy on air pollution. The policy we study is unique in the sense that it reduced public transport fares for a whole country in some areas by as much as 90 percent. To the best of our knowledge, we are the first to causally study the effects of a public transport subsidy on air pollution, thereby expanding previous literature, which has so far predominantly focused on price increases (Yang and Tang, 2018), calibrated quantitative models (Borck, 2019) or leveraged natural experiments such as unexpected public transit strikes (Bauernschuster et al., 2017) to study the general relationship between public transport and air pollution.

As our key finding, we show that pollution levels fall in response to the policy intervention and increase again after the 9ET was abolished. In particular, the AQI decreases by more than eight percent. Further, we document effect heterogeneity showing that the effects are largest during the week, i.e. when individuals commute to and from work. We also show that the effects are more pronounced in urban areas and regions with a well developed public transport network.

Our results are relevant for policymakers and researchers alike. First, our findings suggest that subsidizing public transportation can indeed incentivize a modal switch, which sparks a decrease in pollution levels and potentially other outcomes not studied in this paper such as carbon emissions. Further, the results echo findings of quantitative equilibrium models e.g. by Borck (2019), who also finds a decrease in pollution in response to lifting

[24]The ticket's cost were estimated based on foregone ticket revenues (Die Bundesregierung, 2022).

public transport fares all together. Our effect sizes lie in between the relatively small effect of public transport on air pollution documented in general equilibrium models by Borck (2019) and relatively large effects found in other papers studying price changes in public transport fares such as Yang and Tang (2018). The differences in effect sizes is plausible, as we estimate the short-term impact of a temporary subsidy and can hence not account for long term equilibrium effects, which might work in offsetting directions.[25] Further, we focus on pollution measurements across Germany, whereas other papers have singled out large urban agglomerations where a stronger relationship between public transport and air pollution seems plausible, and indeed is also shown in our heterogeneity analyses. Additionally, our back-of-the-envelope calculations show that if the 9ET was to be permanent and behavioural adjustments in the modal split were to remain, the costs of the ticket in theory could be amortised by positive health effects and their related economic burden. The analysis of permanent price changes may provide information on longer-term and also potential relocation effects as well as a test for the external validity of our results.

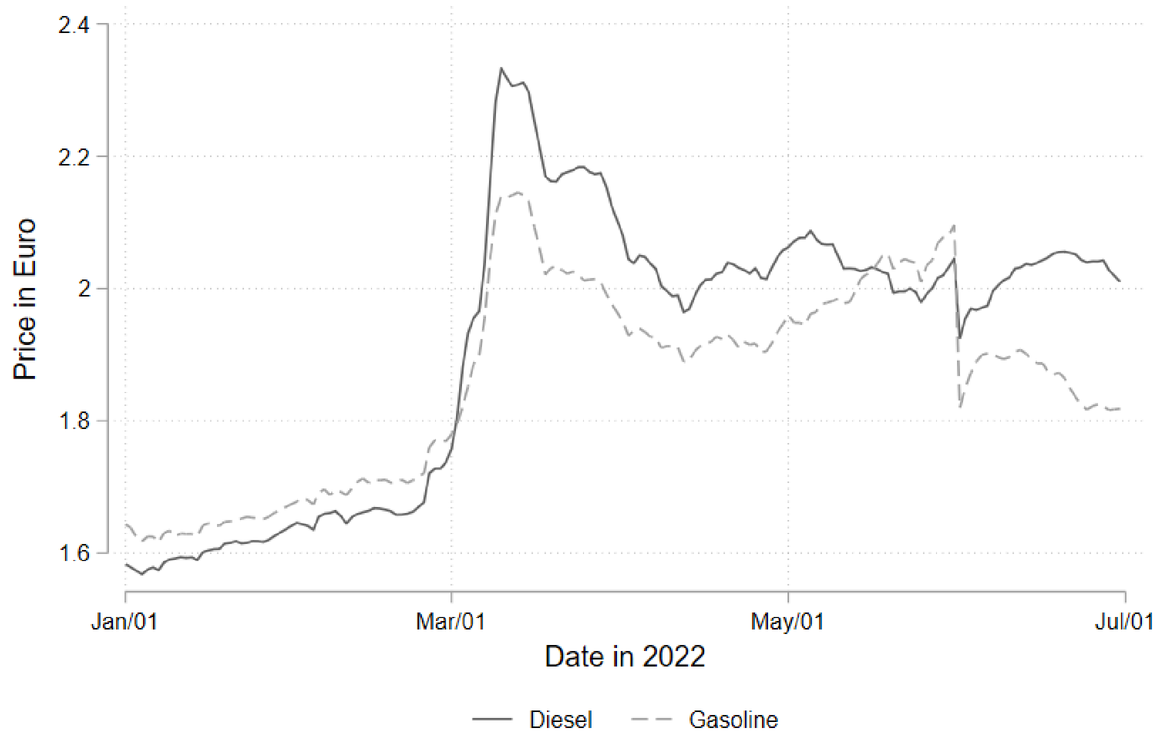
Overall, the findings thus indicate that subsidizing public transport might be a viable option to reduce pollution in the short term and given results by Borck (2019) potentially also in the long term. Thereby, a policy intervention such as the 9-Euro-ticket may indeed contribute to the UN's sustainability goal of creating more resilient, safer and healthier urban agglomerations.

[25] In particular, (Borck, 2019), in his model, emphasizes relocation mechanisms and an increase in housing consumption and residential emissions.

Appendices

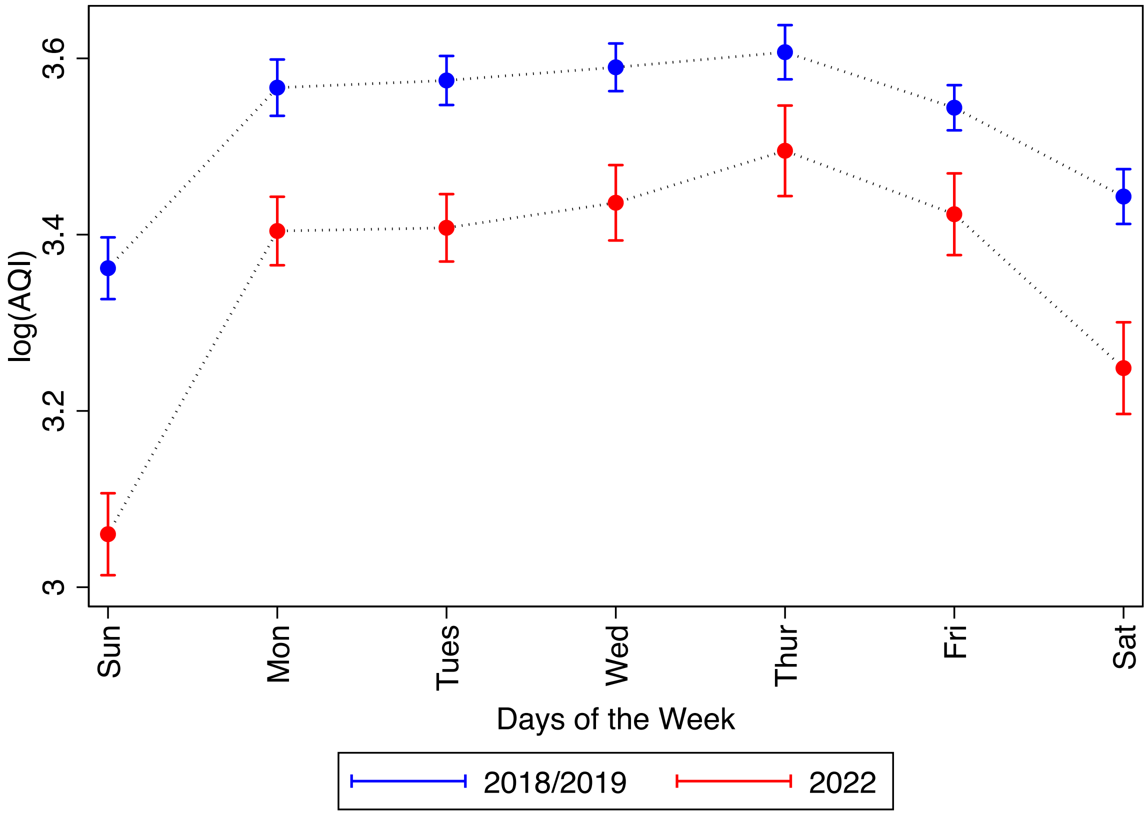
6.A Online Appendix: Tables and Figures

Figure 6.A.1: Development of average gasoline and diesel prices in Germany, 2022



Note: The graph shows the development of gasoline and diesel prices between January, 2022 and July, 2022. Gasoline and diesel data are retrieved from https://creativecommons.tanker koenig.de/. Gasoline price is the average of E5 and E10 prices.

Figure 6.A.2: Day of the week averages for log(AQI) in May 2018/2019 and 2022.



Note: Day of the week averages for log(AQI) and 99 % confidence intervals in pre-policy period of May for treatment (2022) vs. control year (2018/2019), excluding school and public holidays.

Table 6.A.1: Main Results: AQI Absolute Values

	(1)	(2)	(3)	(4)
Interaction	-2.3421***	-2.3171***	-1.9269***	2.0520*
	(0.6743)	(0.6388)	(0.5184)	(1.0836)
Covariates	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes
Observed Years	18,19,22	18,19,22	18,19,22	18,19,22
Included Months	May-June	May-July	May-August	August-Sep
Observations	21926	26485	30737	15054

Note: Source: own calculations. Standard errors in parentheses and clustered at district level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the AQI as outcome variable. Column (1) shows the results for a DiD approach only including months May and June. Columns (2)-(3) augment the estimation with subsequently adding additional months. Column (4) shows the effect of abolishing the 9ET. All specifications control for the full set of covariates and fixed effects.

Table 6.A.2: Results for SE Station Clusters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interaction	-0.0877***	-0.0833***	-0.0630***	-0.0089	-0.0125	-0.0230	0.0665**
	(0.0202)	(0.0197)	(0.0160)	(0.0167)	(0.0165)	(0.0163)	(0.0291)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observed Years	18,19,22	18,19,22	18,19,22	18,19	18,19	18,19	18,19,22
Included Months	May-June	May-July	May-August	May-June	May-July	May-August	August-Sep
Observations	21926	26485	30737	14523	17093	20413	15054

Note: Source: Own calculations. Standard errors in parentheses and clustered at station level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the log of the AQI as outcome variable. Column (1) shows the results for a basic DiD approach only including months May and June. Columns (2)-(4) augment the estimation with subsequently adding additional months. Columns (4) - (6) implement placebo tests where year 2019 is used as the new policy date. Column (7) shows the effect of abolishing the 9ET. All specifications control for the full set of covariates and fixed effects.

Table 6.A.3: Results for SE Station-Year Clusters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interaction	-0.0877***	-0.0833***	-0.0630***	-0.0089	-0.0125	-0.0230	0.0665**
	(0.0189)	(0.0178)	(0.0152)	(0.0151)	(0.0149)	(0.0146)	(0.0279)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observed Years	18,19,22	18,19,22	18,19,22	18,19	18,19	18,19	18,19,22
Included Months	May-June	May-July	May-August	May-June	May-July	May-August	August-Sep
Observations	21926	26485	30737	14523	17093	20413	15054

Note: Own calculations. Standard errors in parentheses and clustered at station-year level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the log of the AQI as outcome variable. Column (1) shows the results for a basic DiD approach only including months May and June. Columns (2)-(4) augment the estimation with subsequently adding additional months. Columns (4) - (6) implement placebo tests where year 2019 is used as the new policy date. Column (7) shows the effect of abolishing the 9ET. All specifications control for the full set of covariates and fixed effects.

Table 6.A.4: Station Specific Day and Year FX: log(AQI)

	(1)	(2)	(3)
Interaction	-0.1225***	-0.0897***	-0.0496***
	(0.0249)	(0.0205)	(0.0146)
Covariates	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Station FE	Yes	Yes	Yes
Observed Years	18,19,22	18,19,22	18,19,22
Included Months	May-June	May-July	May-August
Observations	21926	26485	30737

Note: Source: own calculations. Standard errors in parentheses and clustered at district level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the log of the AQI as outcome variable. Column (1) shows the results for a basic DiD approach only including months May and June. Columns (2)-(3) augment the estimation with subsequently adding additional months.

Table 6.A.5: Leave-One-Out Results

	(1)	(2)	(3)
Interaction	-0.0877***	-0.0876***	-0.0829***
	(0.0207)	(0.0213)	(0.0211)
Covariates	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Station FE	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes
Observations	21926	21660	21395

Note: Standard errors in parentheses and clustered at district level, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Outcome variable is the log of the air quality index. (1) leaves out no observations before and after the introduction, (2) the first day on each side and (3) the first two days on each side. All specifications control for the full set of covariates and fixed effects.

Table 6.A.6: Results for Different Air Pollutants

	(1) $\log(PM_{10})$	(2) $\log(PM_{2.5})$	(3) $\log(NO_2)$	(4) PM_{10}	(5) $PM_{2.5}$	(6) NO_2
Interaction	-0.1018***	-0.1300***	-0.1066***	-1.5706***	-0.9243**	-2.5090***
	(0.0251)	(0.0400)	(0.0204)	(0.4303)	(0.3513)	(0.6609)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Station x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Station FE	Yes	Yes	Yes	Yes	Yes	Yes
Station x Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18186	11118	21043	18186	11118	21043

Note: Own calculations. Standard errors in parentheses and clustered at district level, significance levels* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns (1)-(3) show the logged outcomes for different pollutants separately, while columns (4)-(6) repeat the exercise using absolute values. All specifications control for the full set of covariates and fixed effects.

6.B Online Appendix: Mechanisms

In order to support the claim that the improvement in air quality indeed comes from a modal shift of commuters leaving their cars in the carport, we provide further evidence of that causal chain by re-estimating Equation 6.1, but now use vehicle volume as outcome. We concentrate on the examination of traffic in the city of Berlin. The reason for considering this particular city is that it publicly shares very granular traffic monitoring data measured hourly at specific road segments for a representative set of streets.[26]

Results are shown in Table 6.B.1. The outcomes support the conjecture that there has been a reduction in car utilization. Column 1 only includes the month of June as treated unit, while columns 2 and 3 subsequently add the following months, when the 9ET was still in place.[27] The development of the point estimate over time moreover is in line with our findings that the improvement in air quality was largest in the first month the 9ET existed, but subsequently diminished. Reports about lower congestion levels in 23 out of 26 cities, based on data provided by the enterprise *TomTom* that supplies navigational systems for cars, additionally supports this notion of lower car use in urban areas (Süddeutsche Zeitung, 2022).

Since Berlin is a large metropolitan area with a well-developed public transportation network, the outcomes are likely larger than in other parts of Germany. Nevertheless, this would also be consistent with our finding that air quality improved primarily in urban areas.[28] Indeed, performing the air quality estimations only with Berlin in the sample also yields a significant higher point estimate compared to our main findings.[29]

The question remains to what extent those drivers actually turned to using trains due to now lower prices. According to the association, which gathers all German transport

[26] The data can be obtained at https://api.viz.berlin.de/daten/verkehrsdetektion. We used the street-specific measuring cross section for the years 2018, 2019 and 2022.

[27] The graphical assessment of the common trends assumption supports the empirical approach also for the traffic data used here. The graph is available upon request.

[28] Due to the fact that such detailed data do not exist for all of Germany, we limit this part of the analysis to Berlin. We acknowledge that these outcomes thereby provide rather suggestive evidence of the likely mechanism.

[29] Results are available upon request.

Table 6.B.1: Traffic in Berlin

	(1)	(2)	(3)
Interaction	-0.1533***	-0.1242***	-0.0942***
	(0.0064)	(0.0056)	(0.0042)
Covariates	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Station FE	Yes	Yes	Yes
Observed Years	18,19,22	18,19,22	18,19,22
Included Months	May-June	May-July	May-August
Observations	559158	587818	778327

Note: Source: own calculations. Standard errors in parentheses and robust, significance levels * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table displays regression results using the log of traffic volume as outcome variable. Column (1) shows the results for a basic DiD approach only including months May and June. Columns (2)-(3) augment the estimation with subsequently adding additional months. All specifications control for the full set of covariates and fixed effects.

companies under one roof (*Verband Deutscher Verkehrsunternehmen*, VDV), the 9ET was very well received. Based on a large-scale survey, the VDV found in a final report that the 9ET was sold about 52 million times and resulted in about 1.8 million new passengers that had not used public transport priorly. Furthermore, 43% of respondents saw the ticket as an incentive to leave the car behind and one out of ten journeys made through the ticket replaced a car ride (VDV, 2022).[30]

[30]Unfortunately, the VDV does not provide passenger data on a monthly basis such that we were not able to assess ridership data ourselves.

Chapter 7

Conclusion

This dissertation studies and evaluates policy tools designed to address two key societal challenges for developed countries in the 21st century: ***Ageing Societies*** and ***Living in Urban Agglomerations***. More generally, it analyses individual behaviour and its implications for welfare in response to these challenges. To this end, the dissertation uses state-of-the-art micro-econometric tools and economic theory such as methods for causal inference and quantitative modeling, which ultimately allows to isolate causal responses and simulate counterfactual policy scenarios. In this final chapter, I discuss the relevance as well as limitations of the analyses and highlight potential avenues for future research.

Ageing Societies

Chapter 2 and ***Chapter 3*** of this dissertation focus on issues related to the ageing of society and, in particular, assess the forward-looking effects of pension reforms that raise the official retirement age on a range of outcomes such as on-the-job training and employment behaviour. Previous research has largely focused on examining the direct effects of pension reforms, that is, the effects between the old and new retirement ages. By focusing on individuals' behaviour prior to reaching the shifted retirement threshold, the two chapters aim to contribute to a small but growing literature that examines the response of relatively young and middle-aged individuals to pension reforms. This is important because adjustments in individuals' behaviour before retirement are poorly studied but may affect the long-term sustainability of pay-as-you-go pension systems.

Both chapters clearly show that some individuals fare better than others in adjusting to increases in the work life. In particular, individuals with a low level of education and those who are employed in physically intense jobs do not respond to pension reforms by increasing training and their labour supply. This has important policy implications, as pension reforms may unintentionally exacerbate inequalities in old age - something that policymakers could address through additional labour market policies.

The mechanisms that lead to these differential responses and the reasons why some individuals, such as those with low levels of education, are unable to respond to prolonged working lives, require further research. Additionally, the few papers focusing on forward-looking pension reform effects including the two chapters in this dissertation study effects for individuals in their 40s and 50s. However, younger individuals may also adjust their behaviour in response to a shift in their work horizon. In particular, different behavioural responses and mechanisms may be at play for individuals in their 30s. Thus, future research should seek to study causal responses of younger individuals to pension reforms.

Living in Urban Agglomerations

The second part of the dissertation deals with the challenges of living in densely populated urban areas. In particular, two chapters deal with the housing market and one with air pollution.

Chapter 4 and ***Chapter 5*** study housing markets. What both chapters have in common is that they focus on long-term outcomes in the context of housing markets, whereas most previous research in this area has focused on short- and medium-term outcomes.

First, ***Chapter 4*** studies the formation of house price expectations and as a key novel finding documents that individual house price expectations show no evidence of momentum effects in the long run. In other words, individuals in "hot" and "cold" markets have the same long term price expectations. All in all, the findings are generally consistent with empirically observed and theoretically formulated models of housing cycles that document short-run momentum effects and long-run mean reversion, see e.g. Gao et al. (2009).

One limitation of the paper is that it focuses on hypothetical questions and outcomes, such as a portfolio choice question. Future research should focus on directly linking the elicited subjective beliefs with actual individual behaviour such as purchasing and moving decisions, which are highly relevant for actual house price developments and thus may also inform policies seeking to provide affordable housing.

Affordable housing is the key topic of ***Chapter 5***, which studies the long term welfare effects of a range of policies such as a rental cap, a housing subsidy and rezoning policies. Most importantly, it shows that the calibrated rental cap for Berlin does not succeed in mitigating unequal distributional consequences of gentrification between high and low income households. Instead, we find a reduction in housing prices but also a decrease in average housing consumption and rental income across the income distribution, which ultimately reduces welfare across all income groups. In particular, the welfare loss is strongest for the lowest income groups, which runs counter the aim of the policy. The paper shows that spatial sorting patterns, as studied in our model, are key to better understanding the distributional effects of housing policy.

Future research in this area should focus on expanding quantitative spatial sorting models that, like our model, use granular geo-spatial data but could extend and relax some of the assumptions we have made. For example, incorporating dynamic processes and endogenous homeownership decisions could be a fruitful avenue for future research.

Last, ***Chapter 6*** shows that a large reduction in public transport prices substantially reduces air pollution with effects particularly pronounced in urban areas and areas with a well-connected public transport network.

Consequently, a reduction in prices may provide a viable tool for policy-makers to incentivize people to use public transportation, especially in spatially limited urban agglomerations.

It should also be noted that the analysed transport subsidy was temporary and valid for a relatively short period of time. Studying the medium and long term effects of more permanent and moderate subsidies such as the announced 49-Euro-Ticket[1] may therefore be an interesting avenue for future research. For a subsidy to work in the long run, the supply of public transport, and in particular the frequency of trains, can play a key role in creating incentives for individuals to switch from private motorized transport to public transport. A closer look at the interplay between public transport supply and pricing could therefore be another fruitful starting point for future research.

Concluding Remarks

In summary, all chapters of the dissertation emphasize the importance of using, as Angel Gurría puts it, "reliable tools" to measure, quantify, and evaluate the main challenges and policies facing societies in the developed economies of the 21st century, and aim to gain a better understanding of the complexity of these challenges to ultimately help design public policies that can bear "the best fruits". Finally, it should be noted that the challenges and policies examined in this dissertation apply predominantly to relatively developed countries such as OECD member states. However, the same principle of using reliable instruments naturally applies to developing countries, which face different challenges than those considered here.

[1]see https://www.bundesregierung.de/breg-de/aktuelles/deutschlandticket-2134074.

List of Tables

List of Figures

List of Abbreviations

9ET	9-Euro-Ticket
API	Application Programming Interface
AQI	Air Quality Index
CBD	Central Business District
CDU	Christlich Demokratische Union
DiD	Difference-in-Differences
DWD	Deutscher Wetterdienst
ESDB	European Soil Database
EU	European Union
FEA	Federal Environmental Agency
FDP	Freie Demokratische Partei
FE	Fixed Effects
FUA	Functional Urban Area
IV	Instrumental Variable
LEZ	Low Emission Zones

NO Nitrogen Monoxide

NO_2 Nitrogen Dioxide

NO_X Nitrogen Oxides

OECD Organisation for Economic Co-operation and Development

OLS Ordinary Least Squares

OSM Open Street Maps

O_3 Ozone

PM_{10} Particulate matters with aerodynamic diameter of 10 micrometer or less

$PM_{2.5}$ Particulate matters with aerodynamic diameter of 2.5 micrometer or less

SO_2 Sulphur Dioxide

SPD Sozialdemokratische Partei Deutschlands

$\mu g/m^3$ Microgram per cubic meter

WHO World Health Organization

RDD Regression Discontinuity Design

RCT Randomized Controlled Trial

SAVE Saving and old-age provision in Germany Survey

SIAB Sample of Integrated Employment Biographies

IAB Institute for Employment Research

UN United Nations

SUTVA Stable Unit Treatment Value Assumption

SE Standard Errors

German Summary

In der vorliegenden kumulativen Dissertation werden wirtschafttheoretische Methoden und modernste mikroökonometrische Instrumente und Evaluierungsmethoden eingesetzt, um öffentliche Politikmaßnahmen und ihre Auswirkungen auf das soziale Wohlergehen und individuelles Verhalten zu analysieren. Die Arbeit konzentriert sich insbesondere auf die Analyse von Maßnahmen in zwei verschiedenen Bereichen, die grundlegende gesellschaftliche Herausforderungen im 21. Jahrhundert darstellen: die Alterung der Gesellschaft und das Leben in dicht besiedelten städtischen Ballungsräumen. Gemeinsam prägen diese Bereiche wichtige finanzielle Entscheidungen im Leben eines Menschen, wirken sich auf den Wohlstand aus und sind treibende Kräfte hinter vielen der Herausforderungen, mit denen sich entwickelte Länder derzeit konfrontiert sehen. Die Evaluation und Analyse der Auswirkungen öffentlicher Maßnahmen, die zur Bewältigung dieser Herausforderungen entwickelt wurden, helfen daher, ein besseres Verständnis von etwaigen Lösungsansätzen dieser fundamentalen und komplexen gesellschaftlichen Probleme zu schaffen. Neben der Einleitung unterteilt sich die Arbeit in fünf Hauptkapitel, welche in sich geschlossen sind und eigenständige Forschungsbeiträge darstellen. Im Folgenden werden zuerst die grundlegenden Herausforderungen in den Bereichen *Alterung der Gesellschaft* und *Zusammenleben in urbanen Ballungsräumen* skizziert um anschließend die einzelnen Kapitel der Arbeit vorzustellen.

Alterung der Gesellschaft Die Alterung der Gesellschaft stellt eine Herausforderung für die langfristige Haushaltsstabilität vieler Industrieländer dar. Genauer gesagt haben die steigende Lebenserwartung und das hohe Durchschnittsalter der Bevölkerung in den meisten OECD-Mitgliedsländern die fiskalische Belastung der öffentlichen Rentensysteme erheblich erhöht. So kamen in den 1980er Jahren in der OECD auf 10 Personen im erwerbsfähigen Alter zwei Personen, die älter als 65 Jahre waren. Diese Zahl wird bis 2060 stark steigen und in einigen Mitgliedsstaaten sogar bei etwa sechs Personen liegen (OECD, 2019).

Als Reaktion darauf haben viele Mitgliedstaaten eine Reihe von Rentenreformen durchgeführt, die darauf abzielen, ihre umlagefinanzierten Rentensysteme zu entlasten, indem sie die Zahl der LeistungsempfängerInnen verringern und die Zahl der SteuerzahlerInnen erhöhen. Insbesondere das gesetzliche Renteneintrittsalter wurde in den letzten drei Jahrzehnten erheblich angehoben. Dies wiederum kann das Beschäftigungsverhalten und die Humankapitalentscheidungen junger und mittelalter Menschen verändern, die nun mit einem längeren Arbeitshorizont konfrontiert sind - ein Forschungsbereich, der in der wirtschaftswissenschaftlichen Forschung relativ wenig untersucht wurde und in dieser Arbeit in den Kapiteln 2 und 3 betrachtet wird.

Zusammenleben in urbanen Ballungsräumen In Deutschland und in den OECD-Mitgliedstaaten leben etwa 75 % aller Einwohner in Städten mit mehr als 50.000 Einwohnern (OECD, 2020). Ein großer Teil der Forschung, insbesondere auf dem Gebiet der Stadtökonomie, hat sich mit den Vorteilen des Lebens in Städten befasst und dabei unter anderem eine positive Korrelation zwischen Bevölkerungsdichte, Produktivität und Löhnen herausgestellt. Das Leben in großen und dicht besiedelten Ballungsräumen bringt jedoch auch potenzielle gesellschaftliche Herausforderungen mit sich. Diese Dissertation konzentriert sich auf zwei davon: erschwinglichen Wohnraum und Luftverschmutzung. Kapitel 4-5 fokusieren sich auf den Wohnungsmarkt und analysieren unter anderem die Verteilungswirkung von Gentrifizierung sowie Politikmaßnahmen wie Mietpreisregulierungen. Kapitel 6 fokusiert sich auf Aspekte der (urbanen) Luftverschmutzung und analysiert die Auswirkung einer Preissubventionierung des öffentliche Nahverkehrs auf die Luftverschmutzung.

Kapitelübersicht

Im Folgenden werden die einzelnen Kapitel und ihr jeweiliger Forschungsfokus, die Hauptergebnisse and ihre Methodik forgestellt.

Kapitel 2 - Working Life and Human Capital Investment: Causal Evidence from a Pension Reform Kapitel 2 befasst sich mit einem potenziellem Zusammenhang von Humankapitaltheorie und Rentenreformen. Die Humankapitaltheorie, die auf Ben-Porath (1967); Becker (1962) zurückgeht, besagt, dass Individuen umso mehr in ihr Humankapital investieren und z.B. an Weiterbildungsmaßnahmen teilnehmen sollten, je länger der Zeitraum ist, in dem sie Renditen auf ihre Investitionen erzielen können. Das Kapitel argumentiert, dass die Anhebung des gesetzlichen Renteneintrittsalters den Zeitraum verlängert, in dem Individuen Renditen auf ihr Humankapital erwirtschaften können. Rentenreformen, die eine Verschiebung des Renteneintrittsalters ankündigen, bevor die betroffenen Personen tatsächlich in den Ruhestand gehen, können daher zu einem Anstieg der Weiterbildungsmaßnahmen bei den derzeit noch Beschäftigten führen, wodurch sich die Beschäftigungsfähigkeit und die Fähigkeit des Einzelnen, länger zu arbeiten, verbessern können.

Um zu untersuchen, ob dies tatsächlich der Fall ist, wird in diesem Beitrag ein Quasi-Experiment betrachtet, das durch eine kohortenspezifische Rentenreform in Deutschland ausgelöst wurde, durch die das Vorruhestandsalter für Frauen, die nach 1951 geboren wurden, um drei Jahre erhöht wurde. Konkret verwendet die Studie Regression Discontinuity Designs (RDD) und Difference-in-Discontinuities-Ansätze, um Frauen, die Ende 1951 geboren wurden, mit Frauen zu vergleichen, die Anfang 1952 geboren wurden. Auf diese Weise werden Frauen verglichen, die sich in Bezug auf ihr Alter und andere Merkmale abgesehen von ihrem gesetzlichen Renteneintrittsalter stark ähneln, was es ermöglicht, kausale Schätzungen durchzuführen.

Das Hauptergebnis ist, dass eine Verlängerung der Lebensarbeitszeit kausal die Investitionen in Humankapital erhöht: Anhand von Mikrozensusdaten wird gezeigt, dass die

Wahrscheinlichkeit, an einer Weiterbildung teilzunehmen, um etwa 2,5-4,2 Prozentpunkte steigt. Je nach Spezifikation entsprechen die Punktschätzungen einem relativen Anstieg von etwa 20-30 %, was darauf hindeutet, dass eine Verlängerung der Lebensarbeitszeit beträchtliche Auswirkungen auf die Ausbildungsbeteiligung hat. Interessanterweise werden die Effekte ausschließlich von Frauen mit einem hohen Bildungsabschluss getrieben, was impliziert, dass es starke Unterschiede gibt, wie Individuen sich an eine längere Lebensarbeitszeit anpassen können.

Kapitel 3 - Working Longer, Working Stronger? The Forward-Looking Effects of Increasing the Retirement Age on (Un)employment Behaviour Kapitel 3 untersucht ebenfalls die vorausschauenden Reaktionen von Personen mittleren Alters auf Rentenreformen. Anstelle von Fort-und Weiterbildungen werden jedoch individuelle Beschäftigungsanpassungen auf eine Rentenreform untersucht, durch die das reguläre Renteneintrittsalter in Deutschland schrittweise von 65 auf 67 Jahre angehoben wurde.

Unter Verwendung von Differenz-in-Differenzen-Methoden (DiD) vergleicht das Papier Kohorten, die unterschiedlich von der Reform betroffen sind und zeigt anhand von Verwaltungs- und Umfragedaten, dass Individuen ihre Ruhestandserwartungen und ihr Beschäftigungsverhalten als Reaktion auf eine exogene Erhöhung ihres gesetzlichen Arbeitshorizonts anpassen.

Im Einzelnen dokumentiert das Papier vier Hauptergebnisse. Erstens wird gezeigt, dass selbst kleine Unterschiede in der gesetzlichen Lebensarbeitszeit, die durch Rentenreformen verursacht werden, zu einer Verschiebung der Länge des subjektiv erwarteten Arbeitshorizonts der Individuen führen. Zweitens erhöht ein längerer Arbeitshorizont die Beschäftigungswahrscheinlichkeit von Personen mittleren Alters erheblich und wirkt sich negativ auf die Wahrscheinlichkeit aus, in einem bestimmten Jahr als arbeitslos gemeldet zu sein. Drittens liefert die Studie Hinweise darauf, dass Personen mit einem längeren Arbeitshorizont intensiver nach Arbeit suchen und mehr Stellenangebote von den Arbeitsämtern erhalten. Viertens wird gezeigt, dass die dokumentierten Beschäftigungseffekte von Frauen ausgehen und überwiegend in Berufen mit geringerer körperlicher Intensität auftreten, d. h. in Berufen, die im höheren Alter ausgeübt werden können.

Chapter 4 - House Price Expectations Kapitel 4 ist die erste Studie, die sich mit Themen aus dem Bereich der Stadtökonomie befasst, insbesondere mit dem Wohnungsmarkt. Die Studie analysiert die subjektiven Erwartungen von Hauspreisentwicklung und trägt damit zu einem besseren Verständnis der Funktionsweise der Wohnungsmärkte bei, was eine Voraussetzung für die Gestaltung einer wirksamen Wohnungspolitik und das Verständnis ihrer Auswirkungen ist.

Insbesondere werden in diesem Kapitel die kurz-, mittel- und langfristigen Preiserwartungen auf den Wohnungsmärkten anhand von Umfragedaten der SOEP-Innovationsstichprobe untersucht. Der Hauptbeitrag besteht darin, dass Hauspreiserwartungen über einen Zeithorizont von dreißig Jahren untersucht werden, während sich frühere Forschungsarbeiten auf einen viel kürzeren Zeitrahmen konzentriert haben, siehe z. B. Armona et al. (2019). Langfristige Hauspreiserwartungen sind besonders interessant, da sie Kauf-und Investitionsentscheidung beinflussen können.

Als Hauptergebnis wird gezeigt, dass Preiserwartungen langfristig keine Anzeichen für Momentum-Effekte aufweisen.[2] Ebenfalls wird gezeigt, dass individuelle Erfahrungen, wie Preisanstiege in der unmittelbaren Nachbarschaft und Informationen über aggriegierte historische Preisentwicklungen Hauspreiserwartungen positiv beeinflussen können.

Kapitel 5 - Gentrification and Affordable Housing Policies In diesem Papier wird ein quantitatives räumliches Wirtschaftsmodell entwickelt, um die Auswirkungen von Gentrifizierung und verschiedene politische Maßnahmen zu bewerten, die den Verteilungseffekten von Gentrifizierung auf die Wohlfahrt heterogener Haushalte entgegenwirken. Das Modell ist auf die wichtigsten wirtschaftlichen Variablen in Berlin, Deutschland, kalibriert.

[2]Momentum-Effekte beschreiben hier eine Situation, in der Hauspreise und -erwartungen steigen, wenn Hauspreise in früheren Perioden ebenfallsgestiegen sind. Armona et al. (2019) dokumentieren Hauspreiserwartungen, die mit kurzfristigen Momentum-Effekten konsistent sind.

In der Hauptanalyse werden kontrafaktische Simulationen durchgeführt, um die Verteilungseffekte verschiedener Maßnahmen für erschwinglichen Wohnraum zu bewerten. Eine davon basiert auf einer Mietpreisregulierung, die von der Berliner Landesregierung eingeführt (und später nach einem Urteil des Verfassungsgerichts wieder aufgehoben) wurde. Mit dieser Maßnahme wurden im Wesentlichen Höchstmieten für alle bestehenden und neuen Mietverträge mit Ausnahme von neu gebauten Wohnungen festgelegt. Die in der Studie durchgeführten Simulationen zeigen, dass die Mietpreisregulierung die Wohlfahrt für alle Haushalte verringert. Mehr noch, die Regulierung verringert nicht nur die Wohlfahrt für alle Haushaltstypen, sondern die Wohlfahrtsverringerung ist relativ gesehen für die ärmsten Haushalte am größten. Die Maßnahme ist also insofern regressiv, als sie arme Haushalte am meisten trifft.

Weitere Analysen untersuchen alternative Politikmaßnahmen wie beispielsweise eine gezielte Wohnbauförderung für einkommensschwache Haushalte und zwei Maßnahmen, die im Wesentlichen den Wohnungsbau (in einem Teil oder in der gesamten Stadt) erhöhen. Die Simulationsergebnisse zeigen, dass sowohl die Wohnbauförderung als auch der Wohnungsneubau die Wohlfahrt der ärmsten Haushalte am stärksten erhöhen. Die Subvention bewirkt eine Umverteilung von den reichen zu den armen Haushalten, und diese Effekte werden teilweise durch eine Verringerung der Segregation noch verstärkt. Der Wohnungsneubau erhöht das Angebot und senkt damit die Mieten für alle Haushalte. Ebenfalls führt dieser auch zu einer leichten Verringerung der Segregation in der Stadt und, je nach räumlicher Verteilung des neuen Wohnraums, zu einem größeren relativen Wohlfahrtszuwachs für arme Haushalte.

Chapter 6 - Ticket to paradise? The effect of a public transport subsidy on air quality Das Kapitel 6 verwendet die vorübergehende Einführung des 9-Euro-Tickets, welches die monatlichen Preise für den öffentlichen Nahverkehr in Deutschland um bis zu 90 % senkte, als Quelle exogener Variation. Dies ermöglicht es, eine kausale Schätzung des Einflusses günstigerer Preise im öffentlichen Nahverkehr auf die Luftqualität durchzuführen. Insbesondere können günstigere Preise für öffentliche Verkehrsmittel einen Anreiz für den Einzelnen darstellen, vom motorisierten Individualverkehr, einer Hauptquelle der

Luftverschmutzung, auf öffentliche Verkehrsmittel umzusteigen.

Als empirischen Ansatz verwendet die Studie ein DiD-Design und dokumentiert als Ergebnis eine Verbesserung der Luftqualität. Dieser positive Effekt ist am größten in städtischen Gebieten, an Werktagen und dort, wo das Angebot an öffentlichen Verkehrsmitteln gut ausgebaut ist. Weitere Ergebnisse zeigen, dass der Effekt im Laufe der Zeit leicht abnimmt und die Luftverschmutzung nach dem Auslaufen der Maßnahme im September 2022 wieder zunimmt.

Bibliography

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

Almagro, M. and T. Dominguez-Iino (2021): "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," *Working Paper*.

Almenberg, J. and A. Dreber (2015): "Gender, stock market participation and financial literacy," *Economics Letters*, 137, 140–142.

Anderson, H. R. (2009): "Air pollution and mortality: A history," *Atmospheric Environment*, 43, 142–152.

Anderson, M. L. (2014): "Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion," *American Economic Review*, 104, 2763–96.

Armona, L., A. Fuster, and B. Zafar (2019): "Home Price Expectations and Behaviour: Evidence from a Randomized Information Experiment," *The Review of Economic Studies*, 86, 1371–1410.

Arnott, R. (1995): "Time for Revisionism on Rent Control?" *Journal of Economic Perspectives*, 9, 99–120.

Atalay, K. and G. Barrett (2016): "Pension Incentives and the Retirement Decisions of Couples," IZA Discussion Papers 10013, Institute for the Study of Labor (IZA).

Atalay, K. and G. F. Barrett (2015): "The impact of age pension eligibility age on retirement and program dependence: Evidence from an Australian experiment," *Review of Economics and Statistics*, 97, 71–87.

Auffhammer, M. and R. Kellogg (2011): "Clearing the air? The effects of gasoline content regulation on air quality," *American Economic Review*, 101, 2687–2722.

Autor, D. H., C. J.Palmer, and P. A. Pathak (2014): "Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge, Massachusetts," *Journal of Political Economy*, 122, 661–717.

Bailey, M., C. Ruiqing, T. Kuchler, and J. Stroebel (2018): "The Economic Effects of Social Networks: Evidence from the Housing Market," *Journal of Political Economy*, 126, 2224–2276.

Bao, H. and T. Gong, C.M. Yamato (2016): "Endowment effect and housing decisions," *International Journal of Strategic Property Management*, 5, 341–353.

Baranov, V. and H.-P. Kohler (2018): "The Impact of AIDS Treatment on Savings and Human Capital Investment in Malawi," *American Economic Journal: Applied Economics*, 10, 266–306.

Barrett, A. and P. J. O'Connell (2001): "Does Training Generally Work? The Returns to in-Company Training," *ILR Review*, 54, 647–662.

Basso, L. J. and H. E. Silva (2014): "Efficiency and substitutability of transit subsidies and other urban transport policies," *American Economic Journal: Economic Policy*, 6, 1–33.

Bauer, A. B. and R. Eichenberger (2017): "Endogenous aging: How statutory retirement age drives human and social capital," CREMA Working Paper Series 2017-02, Center for Research in Economics, Management and the Arts.

Bauernschuster, S., T. Hener, and H. Rainer (2017): "When labor disputes bring cities to a standstill: The impact of public transit strikes on traffic, accidents, air pollution, and health," *American Economic Journal: Economic Policy*, 9, 1–37.

Baum-Snow, N. and D. Hartley (2020): "Accounting for central neighborhood change, 1980–2010," *Journal of Urban Economics*, 117, 103228.

Becker, G. (1962): "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, 70.

Ben-Porath, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75, 352–365.

Bertoni, M., G. Brunello, and G. Mazzarella (2018): "Does postponing minimum retirement age improve healthy behaviors before retirement? Evidence from middle-aged Italian workers," *Journal of Health Economics*, 58, 215–227.

Binswanger, J. and M. Salm (2017): "Does everyone use probabilities? The role of cognitive skills," *European Economic Review*, 98, 73–85.

Blekesaune, M. and P. E. Solem (2005): "Working Conditions and Early Retirement: A Prospective Study of Retirement Behavior," *Research on Aging*, 27, 3–30.

Bloom, D., M. Kuhn, and K. Prettner (2019): "Health and growth," *Oxford Research Encyclopedia on Economics and Finance.*

Blundell, R., M. C. Dias, D. A. Goll, and C. Meghir (2019): "Wages, Experience and Training of Women over the Lifecycle," *NBER Working Paper 25776.*

BMF (2017): "Abschlussbericht, Spending Review (Zyklus 2016/2017) zum Politikbereich Wohnungswesen," *Bundesministerium der Finanzen: Spending Reviews.*

Boelmann, B., R. Bude, L. Klick, and S. Schaffner (2020a): "Real Estate data for Germany (RWI-GEO-RED)," *RWI-GEO-RED: RWI Real Estate Data (Scientific Use File)-apartments for rents. Version: 2 RWI - Leibniz Institute for Economic Research. Dataset. Dataset. http://doi.org/10.7807/immo:red:wm:suf:v2.*

Boelmann, B., R. Bude, L. Klick, S. Schaffner, and R. et al (2020b): "Real Estate data for Germany (RWI-GEO-RED)," *RWI-GEO-RED: RWI Real Estate Data (Scientific Use File)-houses for rents. Version: 2 RWI - Leibniz Institute for Economic Research. http://doi.org/10.7807/immo:red:hm:suf:v2.*

Boelmann, B. and S. Schaffner (2019): "FDZ Data description: Real-Estate Data for Germany (RWI-GEO-REDv1) - Advertisements on the Internet Platform ImmobilienScout24." *RWI Projektberichte*.

Boersch-Supan, A. and C. Wilke (2004): "The German Public Pension System: How it Was, How it Will Be," *NBER Working Paper 10525*.

Bonin, Holger; Euwals, R. (2005): "Why are Labor Force Participation Rates of East German Women so High?" *Applied Economics Quarterly*, 51, 359–386.

Borck, R. (2019): "Public transport and urban pollution," *Regional Science and Urban Economics*, 77, 356–366.

Börsch-Supan, A., M. Coppola, L. Essig, A. Eymann, and D. W. Schunk (2008): "The German SAVE study: design and results," .

Bracke, P. (2013): "How long do housing cycles last? A duration analysis for 19 OECD countries," *Journal of Housing Economics*, 22, 213–230.

Breidenbach, P. and S. Schaffner (2020): "Real Estate data for Germany (RWI-GEO-RED)," *German Economic Review*, 21, 401–4016.

Breunig, C., I. Grabova, P. Haan, F. Weinhardt, and G. Weizsäcker (2021a): "Long-run expectations of households," *Journal of Behavioral and Experimental Finance*, 31, 100535.

Breunig, C., S. Huck, T. Schmidt, and G. Weizsäcker (2021b): "The Standard Portfolio Choice Problem in Germany," *The Economic Journal*, 131, 2413–2446.

Brodeur, A., A. E. Clark, S. Fleche, and N. Powdthavee (2021): "COVID-19, lockdowns and well-being: Evidence from Google Trends," *Journal of Public Economics*, 193, 104346.

Brueckner, J. K., J.-F. Thisse, and Y. Zenou (1999): "Why Is Central Paris Rich and Downtown Detroit Poor? An Amenity-Based Theory," *European Economic Review*, 43, 91–107.

Brummet, Q. and D. Reed (2019): "The Effects of Gentrification on the Well-Being and Opportunity of Original Resident Adults and Children," .

Brunello, G. and S. Comi (2015): "The side effect of pension reforms on the training of older workers. Evidence from Italy," *The Journal of the Economics of Ageing*, 6, 113–122.

Bundesamt für Statistik (2019): "Eigentümerquote nach Bundesländern," Accessed: 2020-09-23.

Caliendo, M., K. Tatsiramos, and A. Uhlendorff (2013): "Benefit Duration, Unemployment Duration and Job Match Quality: A Regression Discontinuity Approach," *Journal of Applied Econometrics*, 28, 604–627.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2021): "Difference-in-Differences with a Continuous Treatment," *arXiv:2107.02637*.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2018a): "RDROBUST: Stata module to provide robust data-driven inference in the regression-discontinuity design," Statistical Software Components, Boston College Department of Economics.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

——— (2018b): "Manipulation testing based on density discontinuity," *The Stata Journal*, 18, 234–261.

Capozza, D. R., P. H. Hendershott, C. Mack, and C. J. Mayer (2002): "Determinants of Real House Price Dynamics," Working Paper 9262, National Bureau of Economic Research.

Card, D., J. Kluve, and A. Weber (2010): "Active Labour Market Policy Evaluations: A Meta-Analysis," *The Economic Journal*, 120, 452–477.

Carta, F. and M. de Philippis (2021): "Working Horizon and Labour Supply: The Effect of Raising the Full Retirement Age on Middle-Aged Individuals," *Bank of Italy Temi di Discussione (Working Paper) No. 1314*.

Case, K. E. and R. J. Shiller (1989): "The Efficiency of the Market for Single-Family Homes," *The American Economic Review*, 79, 125–137.

——— (2003): "Is There a Bubble in the Housing Market?" *Brookings Papers on Economic Activity*, 2003, 299–342.

Case, K. E., R. J. Shiller, and A. Thompson (2012): "What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets," Working Paper 18400, National Bureau of Economic Research.

Cattaneo, M. D. and R. Titiunik (2022): "Regression Discontinuity Designs," *Annual Review of Economics*, 14, 821–851.

Cervellati, M. and U. Sunde (2013): "Life Expectancy, Schooling, and Lifetime Labor Supply: Theory and Evidence Revisited," *Econometrica*, 81, 2055–2086.

Chan, S., S. Dastrup, and I. G. Ellen (2016): "Do homeowners mark to market? a comparison of self-reported and estimated market home values during the housing boom and bust," *Real Estate Economics*, 44, 627–657.

Charness, G. and U. Gneezy (2012): "Strong Evidence for Gender Differences in Risk Taking," *Journal of Economic Behavior and Organization*, 83, 50–58.

Chen, Y. and A. Whalley (2012): "Green infrastructure: The effects of urban rail transit on air quality," *American Economic Journal: Economic Policy*, 4, 58–97.

Chetty, R. and N. Henderson (2018a): "The Impacts of Neighborhoods on Intergenera- tional Mobility I: Childhood Exposure Effects," *Quarterly Journal of Economics*, 133, 1107–1162.

Chirikos, T. N. and G. Nestel (1991): "Occupational Differences in the Ability of Men to Delay Retirement," *The Journal of Human Resources*, 26, 1–26.

Couture, V., C. Gaubert, J. Handbury, and E. Hurst (2019): "Income Growth and the Distributional Effects of Urban Spatial Sorting," in *NBER WORKING PAPER SERIES*.

Couture, V. and J. Handbury (2020): "Urban revival in America," *Journal of Urban Economics*, 119, 103267.

Currie, J., S. H. Ray, and M. Neidell (2011): "Quasi-experimental studies suggest that lowering air pollution levels benefits infants' and children's health," *Health Affairs*, 30, 2391–2399.

Currie, J. and R. Walker (2011): "Traffic congestion and infant health: Evidence from E-ZPass," *American Economic Journal: Applied Economics*, 3, 65–90.

Dauth, W. and J. Eppelsheimer (2020): "Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: a guide," *Journal of Labour Market Research*, 54.

Diamond, R. (2016): "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill:1980-2000," *American Economic Review*, 106, 479–524.

Diamond, R., T. McQuade, and F. Qian (2019): "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco," *American Economic Review*, 109, 3365–3394.

Die Bundesregierung (2022): "9-Euro-Ticket 52 Millionen Mal verkauft," Available online: https://www.bundesregierung.de/breg-de/suche/faq-9-euro-ticket-2028756, last access: October 2022.

Dijkstra, L., H. Poelman, and P. Veneri (2022): "The EU-OECD definition of a functional urban area," Available online: https://www.oecd-ilibrary.org/docserver/..., last access: October 2022.

DLR (2022): “Wie hat das 9-Euro-Ticket unsere Mobilität verändert?” Available online: https://www.dlr.de/content/de/artikel/news/2022/03/20220825_wie-hat-das-9-euro-ticket-unsere-mobilitaet-veraendert.html, last access: October 2022.

Dlugosz, S., G. Stephan, and R. A. Wilke (2009): “Verkürzte Bezugsdauern für Arbeitslosengeld - Deutliche Effekte auf die Eintritte in Arbeitslosigkeit,” *IAB Kurzbericht No. 30/2009*.

Dlugosz, S., R. A. Wilke, and G. Stephan (2014): “Fixing the Leak: Unemployment Incidence before and after a Major Reform of Unemployment Benefits in Germany,” *German Economic Review*, 15, 329–352.

Duggan, M., P. Singleton, and J. Song (2007): “Aching to retire? The rise in the full retirement age and its impact on the social security disability rolls,” *Journal of Public Economics*, 91, 1327–1350.

DWD Climate Data Center (CDC) (2022): “Daily station data,” Available online: https://cdc.dwd.de/portal/, last access: October 2022.

Döring, C. and K. Ulbricht (2018): “Gentrification Hotspots and Displacement in Berlin. A Quantitative Analysis. In: Helbrecht.I (eds) Gentrification and Resistance,” *Gentrification and Resistance. Springer VS, Wiesbaden*.

Engels, B., J. Geyer, and P. Haan (2017): “Pension incentives and early retirement,” *Labour Economics*, 47, 216–231.

Fan, X., A. Seshadri, and C. Taber (2017): “Understanding Earnings, Labor Supply, and Retirement Decisions,” *Michigan Retirement Research Center Working Paper 367*.

Favilukis, J., P. Mabille, and S. Van Nieuwerburgh (2022): “Affordable Housing and City Welfare,” *Review of Economic Studies*, forthcoming.

Fitzpatrick, M. D. and T. J. Moore (2018): “The Mortality Effects of Retirement: Evidencefrom Social Security Eligibility at Age 62.” *Journal of Public Economics*, 157, 121 – 137.

Foster, A. D. and M. R. Rosenzweig (1995): "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 103, 1176–1209.

Frazis, H. and M. A. Loewenstein (2005): "Reexamining the Returns to Training:Functional Form, Magnitude, and Interpretation," *Journal of Human Resource*, 40, 453–476.

French, E. (2005a): "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *The Review of Economic Studies*, 72, 395–427.

——— (2005b): "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *The Review of Economic Studies*, 72, pp. 395–427.

French, E. and J. B. Jones (2011): "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, 79, 693–732.

French, E., A. S. Lindner, C. O'Dea, and T. A. Zawisza (2022): "Labour Supply and the Pension-Contribution Link," *NBER Working Paper Series*.

Frimmel, W. (2021): "Later Retirement and the Labor Market Re-Integration of Elderly Unemployed Workers," *Journal of the Economics of Aging*, 19.

Gao, A., Z. Lin, and C. F. Na (2009): "Housing market dynamics: Evidence of mean reversion and downward rigidity," *Journal of Housing Economics*, 18, 256–266.

Gehrsitz, M. (2017): "The effect of low emission zones on air pollution and infant health," *Journal of Environmental Economics and Management*, 83, 121–144.

Gendron-Carrier, N., M. Gonzalez-Navarro, S. Polloni, and M. A. Turner (2022): "Subways and urban air pollution," *American Economic Journal: Applied Economics*, 14, 164–96.

Genovese, D. and C. Mayer (1997): "Personal Experiences and Expectations about Aggregate Outcomes," *American Economic Review*, 87, 255–269.

German Federal Office for Statistics (2022): “9-Euro-Ticket: Mobilität steigt deutlich auf kurzen Distanzen im Schienenverkehr,” Available online: www.destatis.de, last access: October 2022.

Gewerkschaftsbund, D. D. (2016): “Arbeitsfähig bis zur Rente? Wie die Beschäftigten ihre Perspektiven einschätzen,” *DGB-Index Gute Arbeit kompakt*.

Geyer, J., P. Haan, A. Hammerschmid, and M. Peters (2020): “Labor Market and Distributional Effects of an Increase in the Retirement Age,” *Labour Economics*, 65, 106–120.

Geyer, J. and C. Welteke (2021): “Closing Routes to Retirement: How Do People Respond?” *Journal of Human Recources*, 56, 277–308.

Glaeser, E. (2011): *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*, Penguin.

Glaeser, E. L. and J. Gyourko (2006): “Housing Dynamics,” Working Paper 12787, National Bureau of Economic Research.

Glaeser, E. L. and E. F. Luttmer (2003): “The Misallocation of Housing Under Rent Control,” *American Economic Review*, 93, 1027–1046.

Goebel, J., M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019): “The German Socio-Economic Panel Study (SOEP),” *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics*, 239, 345–360.

Gohl, N., E. Kurz, P. Haan, and F. Weinhardt. (2021): “Working Life and Human Capital Investment: Causal Evidence from Pension Reform,” *Centre for Economic Performance Discussion Paper No. 1753*.

Gohl, N., A. Schrenker, and A. Steuernagel (2023): “Working Longer: Causal Effects on Career Trajectories of Raising the Statutory Retirement Age,” *Mimeo*.

Goodman, J., T. Ittner, J.B. Yamato, and K. Yokotani (1992): “The accuracy of home owners’ estimates of house value,” *Journal of Housing Economics*, 2, 339–357.

Green, R. K., S. Malpezzi, and S. K. Mayo (2005): "Metropolitan-specific estimates of the price elasticity of supply of housing, and their sources," *The American Economic Review*, 95, 334–339.

Grohmann, A., L. Menkhoff, C. Merkle, and R. Schmacker (2019): "Earn More Tomorrow: Overconfident Income Expectations and Consumer Indebtedness," *SFB Rationality and Competition, Discussion Paper No. 152*.

Guerrieri, V., D. Hartley, and E. Hurst (2013): "Endogenous gentrification and housing price dynamics," *Journal of Public Economics*, 100, 45–60.

Gurría, A. (2011): "First Session on Measuring Progress," Introductory Remarks at first Session on Measuring Progress at OECD's 50th Anniversary Forum [Accessed: 2023 02 14].

Gustman, A. L. and T. L. Steinmeier (1986): "A Structural Retirement Model," *Econometrica*, 54, 555–584.

Haan, P. and V. Prowse (2014): "Longevity, life-cycle behavior and pension reform ," *Journal of Econometrics*, 178, Part 3, 582 – 601.

Hahn, A. M., K. A. Kholodilin, and S. R. Waltl (2022): "Forward to the Past: Short-Term Effects of the Rent Freeze in Berlin ," *DIW Working Paper 1999*.

Hairault, J.-O., T. Sopraseuth, and F. Langot (2010): "Distance to Retirement and Older Workers' Employment: The Case for Delaying the Retirement Age," *Journal of the European Economic Association*, 8, 1034–1076.

Hall, J. D. and J. M. Madsen (2022): "Can behavioral interventions be too salient? Evidence from traffic safety messages," *Science*, 376, eabm3427.

Handelsblatt (2022): ""Es gibt einen Run auf das 9-Euro-Ticket" – Sieben Millionen Tickets verkauft," Available online: https://www.handelsblatt.com/politik/deutschland/sonderticket-es-gibt-einen-run-auf-das-9-euro-ticket-sieben-millionen-tickets-verkauft/28387430.html, last access: October 2022.

Heblich, S., S. J. Redding, and D. M. Sturm (2020): "The making of the modern metropolis: Evidence from London," *The Quarterly Journal of Economics*, 135, 2059–2133.

Hwang, J. and J. Lin (2016): "What Have We Learned About the Causes of Recent Gentrification?" *Cityscape*, 18, 9–26.

Inderbitzin, L., S. Staubli, and J. Zweimüller (2016): "Extended unemployment benefits and early retirement: Program complementarity and program substitution." *American Economic Journal: Economic Policy*, 8, 253–288.

INFAS (2023): "Alles wie vorher? Die Verkehrswende zwischen 9-Euro-Ticket und alten Herausforderungen," Available online: https://www.infas.de/wp-content/uploads/2023/02/infas_Mobilitaetsreport_07_7647-2.pdf, mobilitätsreport 07, Ausgabe Februar 2023. last access: April 2023.

Investitionsbank, B. (2019): "IBB Berlin Wohnungsmarktbericht 2019," *Investitionsbank Berlin: Annual Housing Market Report*.

Jayachandran, S. and A. Lleras-Muney (2009): "Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines," *The Quarterly Journal of Economics*, 124, 349–397.

Jofre Monseny, J., R. Martínez Mazza, and M. Segú (2022): "Effectiveness and Supply Effects of High-Coverage Rent Control Policies," IEB Working Paper 2022/02.

Kahneman, D., J. Knetsch, and H. Richard (1990): "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 98, 1325–1348.

Kholodilin, K., A. Mense, and C. Michelsen (2016): "Market Break or Simply Fake? Empirics on the Causal Effects of Rent Controls in Germany ," *DIW Discussion Paper 1584*.

Kholodilin, K. A., F. A. López, D. R. Blanco, and P. G. Arbués (2022): "Lessons from an Aborted Second-Generation Rent Control in Catalonia," DIW Berlin Discussion Paper No. 2008.

Killingsworth, M. R. (1982): "'Learning by Doing' and 'Investment in Training': A Synthesis of Two 'Rival' Models of the Life Cycle," *The Review of Economic Studies*, 49, 263–271.

Kindermann, F., J. Le Blanc, M. Piazzesi, and M. Schneider (2021): "Learning about Housing Cost: Survey Evidence from the German House Price Boom," CEPR Discussion Papers 16223, C.E.P.R. Discussion Papers.

Knittel, C. R., D. L. Miller, and N. J. Sanders (2016): "Caution, drivers! Children present: Traffic, pollution, and infant health," *Review of Economics and Statistics*, 98, 350–366.

Knoll, K., M. Schularick, and T. Steger (2017): "No Price Like Home: Global House Prices, 1870–2012," *American Economic Review*, 107, 331–353.

Kolesar, M. and C. Rothe (2018): "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, 108, 2277–2304.

Köller, O., M. Hasselhorn, F. Hesse, K. Maaz, J. Schrader, H. Solga, and C. Spiess (2017): *Das Bildungswesen in Deutschland. Bestand und Potenziale.*

Kroll, L. E. (2011): "Konstruktion und Validierung eines allgemeinen Index für die Arbeitsbelastung in beruflichen Tätigkeiten auf Basis von ISCO-88 und KldB-92," *Methoden, Daten, Analysen (mda)*, 5, 63–90.

Kuchler, T. and B. Zafar (2019): "Personal Experiences and Expectations about Aggregate Outcomes," *The Journal of Finance*, 74, 2491–2542.

Kuhn, A., J.-P. Wuellrich, and J. Zweimueller (2010): "Fatal Attraction? Access to Early Retirement and Mortality." *IZA Discussion Papers 5160,.*

Lalive, R., S. Luechinger, and A. Schmutzler (2018): "Does expanding regional train service reduce air pollution?" *Journal of Environmental Economics and Management*, 92, 744–764.

Lalive, R. and S. Staubli (2015): "How Does Raising Women's Full Retirement Age Affect Labor Supply, Income, and Mortality?" Working Paper 14-09, National Bureau of Economic Research.

Lamorgese, A. and D. Pellegrino (2019): "Loss aversion in housing price appraisals among Italian homeowners," *Banca d'Italia Working Papers*.

Leuven, E. (2005): "The Economics of Private Sector Training: A Survey of the Literature," *Journal of Economic Surveys*, 19, 91–111.

Lichter, A. and A. Schiprowski (2021): "Benefit duration, job search behavior and re-employment," *Journal of Public Economics*, 193, 104326.

Lusardi, A. (2008): "Household Saving Behavior: The Role of Financial Literacy, Information, and Financial Education Programs," Working Paper 13824, National Bureau of Economic Research.

Manoli, D. S. and A. Weber (2016): "The Effects of the Early Retirement Age on Retirement Decisions," Working Paper 22561, National Bureau of Economic Research.

Manski, C. F. (2004): "Measuring Expectations," *Econometrica*, 72, 1329–1376.

Margaryan, S. (2021): "Low emission zones and population health," *Journal of Health Economics*, 76, 102402.

Mastrobuoni, G. (2009): "Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities," *Journal of Public Economics*, 93, 1224–1233.

Mazzonna, F. and F. Peracchi (2017): "Unhealthy Retirement?" *The Journal of Human Resources*, 52, 128–151.

MENSE, A. (2020): "The Impact of New Housing Supply on the Distribution of Rents," *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2020: Gender Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg.*

MENSE, A., C. MICHELSEN, AND K. A. KHOLODILIN (2019): "The Effects of Second-Generation Rent Control on Land Values," *AEA Papers and Proceedings*, 109, 385–388.

METCALFE, R., N. POWDTHAVEE, AND P. DOLAN (2011): "Destruction and Distress: Using a Quasi-Experiment to Show the Effects of the September 11 Attacks on Mental Well-Being in the United Kingdom," *The Economic Journal*, 121, F81–F103.

MONTE, F., S. J. REDDING, AND E. ROSSI-HANSBERG (2018): "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 108, 3855–90.

MONTIZAAN, R., F. COERVERS, AND A. DE GRIP (2010): "The effects of pension rights and retirement age on training participation: Evidence from a natural experiment," *Labour Economics*, 17, 240–247.

MÜNCHENER MERKUR (2022): "DB-Experte zieht bittere Bilanz zum 9-Euro-Ticket: "Ich kenne Bahnfahrer, die auf das Auto umgestiegen sind"," Available online: https://www.merkur.de/wirtschaft/9-euro-ticket-db-umfragen-deutsche-bahn-auto-bahnfahrer-news-zr-91757638.html, last access: April 2023.

NIU, G. AND A. VAN SOEST (2014): "House Proce Expectations," *IZA Discussion Papers 8536.*

NIVOLA, P. S. (1999): "Are Europe's Cities Better?" *The Public Interest*, 137, 73–84.

OECD (2019): *Pensions at a Glance 2019*, OECD Publishing, Paris.

——— (2020): *Regions and Cities at a Glance 2020*, OECD Publishing, Paris.

OSTER, E., I. SHOULSON, AND E. R. DORSEY (2013): "Limited Life Expectancy, Human Capital and Health Investments," *American Economic Review*, 103, 1977–2002.

Parry, I. W. and K. A. Small (2009): "Should urban transit subsidies be reduced?" *American Economic Review*, 99, 700–724.

Paulus, W. and B. Matthes (2013): "The German Classification of Occupations 2010 – Structure, Coding and Conversion Table," *FDZ-Methodenreport 08/2013*.

Picchio, M. and J. C. van Ours (2012): "Retaining Through Training; Even for Older Workers," *Economics of Education Review*, 32, 29–48.

Pischke, J.-S. (2001): "Continuous Training in Germany," *Journal of Population Economics*, 14, 523—-548.

Pope III, C. A. and D. W. Dockery (2006): "Health effects of fine particulate air pollution: Lines that connect," *Journal of the Air & Waste Management Association*, 56, 709–742.

RDC of the Federal Statistical Office and Statistical Offices of the Laender (2015): "Microcensus, survey years 1996-2015, own calculations," .

Richter, D. and J. Schupp (2015): "The SOEP Innovation Sample (SOEP IS)," *Schmollers Jahrbuch: Journal of Applied Social Science Studies*, 135, 389–400.

Ruhose, J., S. L. Thomsen, and I. Weilage (2019): "The Benefits of Adult Learning: Work-Related Training, Social Capital, and Earnings," *Economics of Education Review*, 72, 166–186.

Rust, J. and C. Phelan (1997): "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, 65, 781–832.

Sagner, P., M. Stockhausen, and M. Voigtländer (2020): "Wohnen – die neue soziale Frage?" IW-Analyse 136, Institut der deutschen Wirtschaft, Köln.

Sagner, P. and M. Voigtländer (2022): "Supply side effects of the Berlin rent freeze," *International Journal of Housing Policy*, 0, 1–20.

Saiz, A. (2010): "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 125, 1253–1296.

Seibold, A. (2021): "Reference Points for Retirement Behavior: Evidence from German Pension Discontinuities," *American Economic Review*, 111, 1126–65.

Seiler, M. J., V. J. Seiler, S. Traub, and D. Harrison (2008): "Regret Aversion and False Reference Points in Residential Real Estate," *Journal of Real Estate Research*, 30, 461–474.

Shiller, R. J. (2015): *Irrational Exuberance*, no. 10421 in Economics Books, Princeton University Press.

Simeonova, E., J. Currie, P. Nilsson, and R. Walker (2021): "Congestion pricing, air pollution, and children's health," *Journal of Human Resources*, 56, 971–996.

Soares, R. R. (2005): "Mortality Reductions, Educational Attainment, and Fertility Choice," *American Economic Review*, 95, 580–601.

Spiegel Online (2022): "Grünenchefin Lang will Folgeregelung für 9-Euro-Ticket ausloten," Available online: https://www.spiegel.de/politik/deutschland/gruenen-chefin..., last access: October 2022.

Staubli, S. and J. Zweimueller (2013): "Does raising the early retirement age increase employment of older workers?" *Journal of Public Economics*, 108, 17–32.

Süddeutsche Zeitung (2022): "Weniger Stau durch Neun-Euro-Ticket," Available online: `https://www.sueddeutsche.de/wirtschaft/9-euro-ticket-stau-tomtom-1.5612144`, last access: October 2022.

TAZ (2022): "Bloß nicht verlängern," Available online: `https://taz.de/Debatte-um-das-9-Euro-Ticket/!5864045/`, last access: April 2023.

The Guardian (2022): "Free rail travel scheme begins in Spain to cut commuters' costs," Available online: https://www.theguardian.com/world/2022/sep/01/free-rail-travel-scheme-begins-in-spain-to-cut-commuters-costs..., last access: March 2023.

TÜV Rheinland (2019): "Begutachtung der Positionierung verkehrsnaher Probenahmestellen zur Messung der NO2-Konzentration an ausgewählten Standorten - Endbericht," Available online: https://www.bmuv.de/download/tuev-begutachtung-der-positionierung-verkehrsnaher-probenahmestellen-zur-messung-der-no2-konzentrationen-an-ausgewaehlten-standorten, last access: April 2023.

Umweltbundesamt (2022): "Air quality data," Available online: https://www.umweltbundesamt.de/en/data/air/air-data, last access: October 2022.

Van den Elshout, S., K. Léger, and H. Heich (2014): "CAQI common air quality index—update with PM2.5 and sensitivity analysis," *Science of the total environment*, 488, 461–468.

van Ours, J. (2004): "The Locking-in Effect of Subsidized Jobs," *Journal of Comparative Economics*, 32, 37–52.

VDV (2022): "Bilanz zum 9-Euro-Ticket ," Available online: https://www.vdv.de/bilanz-9-euro-ticket.aspx, last access: October 2022.

Vermeer, N., M. Mastrogiacomo, and A. Van Soest (2016): "Demanding occupations and the retirement age," *Labour Economics*, 43, 159–170, health and the Labour Market.

Vigdor, J. L. (2002): "Does Gentrification Harm the Poor? [with Comments]," *Brookings-Wharton Papers on Urban Affairs*, 133–182.

Voigtlaender, M. (2010): "Why is the German Homeownership Rate so Low?" *Housing Studies*, 24, 355–372.

Wheaton, W. C. (1999): "Real Estate "Cycles": Some Fundamentals," *Real Estate Economics*, 27, 209–230.

Yang, Z. and M. Tang (2018): "Does the increase of public transit fares deteriorate air quality in Beijing?" *Transportation Research Part D: Transport and Environment*, 63, 49–57.

Zweimueller, J. and R. Winter-Ebmer (2000): "Firm-specific Training: Consequences for Job Mobility," IZA Discussion Papers 138, Institute for the Study of Labor (IZA).

Zwick, T., M. Bruns, J. Geyer, and S. Lorenz (2022): "Early retirement of employees in demanding jobs: Evidence from a German pension reform," *The Journal of the Economics of Ageing*, 22, 100387.