Essays in Empirical Population Economics

Von der Fakultät Wirtschaftswissenschaften
der Leuphana Universität Lüneburg

zur Erlangung des Grades
Doktor der Wirtschafts- und Sozialwissenschaften
(Dr. rer. pol.)

genehmigte Dissertation von

Inna Petrunyk
aus Chernoutsy (Ukraine)

Eingereicht am: 27.02.2020

Mündliche Verteidigung (Disputation): am 29.04.2020

Erstbetreuer und Erstgutachter: Prof. Dr. Christian Pfeifer
Zweitgutachter: Prof. Dr. Joachim Wagner
Drittgutachter: Prof. Dr. Boris Hirsch

Die einzelnen Beiträge des kumulativen Dissertationsvorhabens sind oder werden wie folgt veröffentlicht:

Prenatal Air Pollution Exposure and Neonatal Health. IZA Discussion Paper No.12467, 2019.
Life Satisfaction in Germany after Reunification: Additional Insights on the Pattern of Convergence. Journal of Economics and Statistics, 2016; 236 (2): 217-239.
Unemployment Benefits Duration and Labor Market Outcomes: Evidence from a Natural Experiment in Germany. IZA Discussion Paper No.11300, 2018.
How Older Workers Respond to Raised Early Retirement Age: Evidence from a Kink Design in Germany.

Veröffentlichungsjahr: 2020

Veröffentlicht im Onlineangebot der Universitätsbibliothek unter der URL:
http://www.leuphana.de/ub

Declaration

Inna Petrunyk
Steinrückweg 2
14197 Berlin
015203776638
petrunyk@leuphana.de

Hiermit erkläre ich, dass ich mich noch keiner Doktorprüfung unterzogen oder mich um Zulassung zu einer solchen beworben habe.

Ich versichere, dass die Dissertation mit dem Titel "Essays in Empirical Population Economics" noch keiner Fachvertreterin bzw. Fachvertreter vorgelegen hat, ich die Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und, dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

Ich versichere, dass ich die eingereichte Dissertation "Essays in Empirical Population Economics" selbstständig und ohne unerlaubte Hilfsmittel verfasst habe. Anderer als der von mir angegebenen Hilfsmittel und Schriften habe ich mich nicht bedient. Alle wörtlich oder sinngemäß anderen Schriften entnommenen Stellen habe ich kenntlich gemacht.

Lüneburg, 27.02.2020

Acknowledgements

I would like to thank my supervisor, Christian Pfeifer, for his guidance through each stage of my doctoral studies. With his continuous support I could complete one of the biggest projects in my life. I am also very grateful to him for the possibility to attend top conferences and workshops, which substantially contributed to my personal and professional development. I further thank all my colleagues at the Institute of Economics for their valuable insights and helpful comments on previous versions of my dissertation. I would like to acknowledge my coauthors Daniela Vuri and Alessandro Palma for the chance to work with and learn from them. I benefited a lot from this collaboration, which deepened my expertise in applied microeconomics and academic work in general.

I am indebted with my parents. Their unconditional support and belief in me encourage me in all my projects. My deep gratitude goes to my farther for his financial support throughout my previous studies on the way to my doctoral studies. A special thank goes to my husband for his patience during all these years. I appreciate his sincere attempt to understand my true motivation for engaging in this activity.

To my husband

Litterarum radices amarae, fructus dulces sunt

Contents

List of Figures

List of Tables

Chapter 1

Introduction

My dissertation embraces four empirical papers addressing socio-economic issues relevant to policy-makers and society as a whole. These papers cover important aspects of human life including health at birth, life satisfaction, unemployment periods and retirement decisions, and are intended to provide a contribution to the respective research areas. The analyses are carried out applying advanced econometric methods and are based on data sets consisting of survey data as well as administrative records.

The joint paper with Alessandro Palma and Daniela Vuri *"Prenatal Air Pollution Exposure and Neonatal Health"* in Chapter 2 investigates the causal impact of prenatal exposure to air pollution on neonatal health in Italy in the 2000s combining detailed information on mother's residential location from birth certificates with PM_{10} concentrations from air pollution monitors. Variation in local weekly rainfall is exploited as an instrumental variable for non-random air pollution exposure. Using quasi-experimental variation in rainfall shocks allows to identify the effect of PM_{10}, ruling out potential bias due to confounder pollutants. The paper estimates the effect of exposure for both the entire pregnancy period and separately for each trimester to test whether the neonatal health effects are driven by pollution exposure during a particular gestation period. This information enhances our understanding of the mechanisms at work and help prevent pregnant mothers from most dangerous exposure periods. Additionally, the effects of prenatal exposure to PM_{10} are estimated by maternal labor market status and maternal education level to understand how the pollution burden is shared across different population groups. This decomposition allows to identify possible mechanisms through which environmental inequality reinforces the negative impact of early-life exposure to air pollution. This study finds that average PM_{10} and days with PM_{10} level above the hazard limit reduce birth weight, gestational age, and measures of overall newborn health. Effects are largest for third trimester exposure and for low-income and less educated mothers. These findings imply that further policy efforts are needed to fully protect fetuses from the adverse effects of air pollution and to mitigate the environmental inequality of health at birth.

The joint paper with Christian Pfeifer *"Life Satisfaction in Germany After Reunification: Additional Insights on the Pattern of Convergence"* in Chapter 3 updates previous findings on the total East-West gap in overall life satisfaction and its trend by using data from the German Socio-Economic Panel for the years 1992 to 2013. Additionally, the effects are separately analyzed for men and women as well as for four birth cohorts. The results indicate that reported life satisfaction is, on average, significantly lower in East than in West German federal states and that part of the raw East-West gap is due to differences in household income and unemployment status. The conditional East-West gap decreased in the first years after the German reunification and remained quite stable and sizable since the mid-nineties. The results further indicate that gender differences are small. Finally, the East-West gap is significantly smaller and shows a trend towards convergence for younger birth cohorts.

The joint paper with Christian Pfeifer *"Unemployment Benefits Duration and Labor Market Outcomes: Evidence from a Natural Experiment in Germany"* in Chapter 4 explores the effects of a major reform of unemployment benefits in Germany on the labor market outcomes of individuals with some health impairment. The reform induced a substantial reduction in the potential duration of regular unemployment benefits for older workers. This work analyzes the reform in a wider framework of institutional interactions, which allows to distinguish between its intended and unintended effects. The results based on routine data collected by the German Statutory Pension Insurance and a Difference-in-Differences design provide causal evidence for a significant decrease in the number of days in unemployment benefits and increase in the number of days in employment. However, they also suggest a significant increase in the number of days in unemployment assistance, granted upon exhaustion of unemployment benefits. Transitions to unemployment assistance represent an unintended effect, limiting the success of a policy change that aims to increase labor supply via reductions in the generosity of the unemployment insurance system.

The single-authored paper *"How Older Workers Respond to Raised Early Retirement Age: Evidence from a Kink Design in Germany"* in Chapter 5 explores how an increase in the early retirement age affects labor force participation of older workers. The analysis is based on a social security reform in Germany, which raised the early retirement age over several birth cohorts to boost employment of older people and ultimately alleviate the burden on the public pension system. Detailed administrative data from the Federal Employment Agency allow to distinguish between employment and unemployment as well as disability pensions and retirement benefits claims. Using a Regression Kink design in a quasi-experimental framework, I show that the raised early retirement age had positive employment effects and negative effects on retirement benefits claims. The reform did not affect unemployment benefits or disability pensions claims. My results also show that some population groups are more sensitive to a reduction in retirement options and more likely to seek benefits from other government programs. In this respect, I find that workers in manufacturing sector respond to the raised early retirement age by claiming benefits from the disability insurance program designed to compensate for reduced earnings capacity due to severe health problems. The treatment heterogeneity analysis further suggests that high-wage workers are more likely to delay exits from employment, which is in line with incentives but might also indicate an increased inequality within the affected birth cohorts induced by the reform. Finally, women seem to rely on alternative sources of income such as retirement benefits for women, or spouse's or partner's income not observed in the data. All things considered, workers did not adjust to the increased early retirement age by substituting early retirement with other government programs but rather responded to the reform in line with the policy intent. At the same time, the findings point to heterogeneous behavioral responses across different population groups. This implies that raising the early retirement age is an effective policy tool to increase employment only among older people who have the real choice to delay employment exits. Therefore, reforms that raise statutory ages should ensure social support for workers only marginally attached to the labor market or not able to work longer due to potential health problems or other circumstances.

Chapter 2

Prenatal Air Pollution Exposure and Neonatal Health

Joint with: Alessandro Palma, Daniela Vuri. Authors' Contributions.

Available as: IZA DP No.12467.

Presented at:

European Economic Association **(EEA)**, Manchester (UK), 2019.

European Association of Labour Economists **(EALE)**, Uppsala (SE), 2019.

Data replication statement:

The raw birth data employed in this study are originally retrieved at individual level by the Italian Ministry of Health and constitute sensitive data according to the Italian Law 101/2018. These data were obtained under a strict confidentiality agreement with the Directorate General of Statistics of the Ministry of Health, which allows to use the data only at aggregated level and for the specific research purposes mentioned in the research project approved by the National Center for Prevention and Diseases Control (Scientific Research Program "The effect of air pollution on the Italian population. An analysis based on microdata.", grant no. E83C17000020001). Statistical programs with all steps of data preparation and cleaning are available. Unfortunately, the data can not be provided as this would represent a violation of the confidentiality agreement above mentioned. Nevertheless, in case of request, full assistance to anybody interested in getting the individual data from the Ministry of Health will be provided.

2.1 Introduction

Air pollution has far-reaching health effects, mostly for the old and very young (Anderson, 2009) but much less is known about the impacts of prenatal exposure. Particulate matter consisting of solid and liquid particles with a diameter of less than 10 μm, also known as PM_{10}, is a particularly harmful pollutant (WHO, 2013). Unlike other pollutants, it originates both from natural and anthropogenic sources such as volcanic ash, naturally suspended dust, emissions produced from a variety of industrial activities but also fuel combustion for vehicles and domestic heating for households. This makes PM_{10} a relatively widespread pollutant, with concentration levels decreasing at a slower pace (EEA, 2019). This paper uses regional and daily variation in rainfall as a source of quasi-experimental variation in prenatal air pollution exposure to identify its causal effect on neonatal health.

Studying the health impact of prenatal pollution exposure is important because the intra-uterine environment is a crucial determinant of infant's survival and health for the years to come (Barker et al., 1989; Elder et al., 2019). Given that health shocks can affect labor supply, productivity and cognition, air pollution can be viewed as an important externality able to modify production factors associated with economic growth (Bharadwaj et al., 2017; Currie and Moretti, 2007; Figlio et al., 2014; Isen et al., 2017). Therefore, it is of utmost importance to identify the risk factors for neonatal development to mitigate the negative effects of poor health at birth on future child and adult outcomes (Currie, 2009).

Analyzing the causal effect of prenatal exposure to air pollution is challenging for two main reasons. First, pollution is rarely randomly assigned across individuals, and individuals who live in highly polluted areas might be in worse health for reasons unrelated to pollution. Maternal preferences for clean air can covary with unobservable determinants of health, which can lead to bias in regression analysis. Moreover, heterogeneity across individuals in either health response to, or preference for, air pollution level implies that the individuals might self-select into locations on the basis of these unobservable differences. In both cases, the neonatal health effects of prenatal air pollution exposure might reflect the responses of various subpopulations or spurious correlations related to omitted variables. Recent research attempts to address the issue of non-random assignment using various econometric tools such as comparisons across siblings in a panel framework, mother fixed effects, or reduced form approaches, focusing mainly on infant mortality. Much less is known about the impact of particle pollution on neonatal health.

Second, given that air pollutants often originate from the same sources, they are highly correlated with each other, which complicates the identification of the effect of a single pollutant. Very few contributions have been able to pin down the health effects of single pollutants such as CO, NO_2 and O_3 (Deschenes et al., 2017; Hanna and Oliva, 2015; Lavaine and Neidell, 2017; Schlenker and Walker, 2015). Knittel et al. (2016) estimate the specific effects of PM_{10} on infant mortality while Deryugina et al. (2019) estimate the medical costs deriving from $PM_{2.5}$ exposure; both studies account for potential pollutants correlation. Nonetheless, we have very limited knowledge of studies that investigate the specific effect of PM on neonatal health in a causal setting. The goal of this article is to identify the mechanisms underlying the neonatal health effects of in-utero exposure to PM_{10}.

We make four major contributions to the existing literature in this area. First, we exploit quasi-experimental variation in local weekly rainfall to address endogeneity of air pollution. Rainfall has a strong ability to clean air from PM with no effect on other pollutants and is mostly idiosyncratic around a local mean. This property allows to specifically identify the effect of PM_{10} ruling out potential bias due to other confounder pollutants.

Second, we estimate the effects for both the entire pregnancy period and each trimester separately and test whether the neonatal health effects are driven by pollution exposure during a particular period of gestation. We construct precise measures of prenatal exposure to air pollution using administrative data from the Italian national registry of births which provide information on the newborn's exact date of birth and the geographic residence of the mother at the municipality level, among other characteristics, combined with PM_{10} concentrations at daily level obtained from monitoring stations, and granular weather information. This information enhances our understanding of the mechanisms at work and help prevent pregnant mothers from most dangerous exposure periods. In addition to consider a measure of average prenatal exposure to pollution, we also investigate the effect of pollution at the intensive margin, i.e. the effect of days of prenatal exposure to pollution levels beyond the current recommended limit, an important parameter of direct policy relevance which has been overlooked in the literature.

Third, while previous research mainly focused on birth weight and gestation as proxies for neonatal health, the richness of data from the birth certificates allows to extend the analysis to a broad range of neonatal health outcomes including intra-uterine growth restriction and Apgar index. These outcomes offer a complementary measure of health impacts not captured by commonly used measures such as birth weight, which has been increasingly considered a weak summary measure (Almond et al., 2005). For instance, a low Apgar score correlates with neonatal mortality in large populations (Casey et al., 2001; Li et al., 2013).

Finally, we analyze the effects of prenatal exposure to PM_{10} by maternal labor market status and maternal education level to understand how the pollution burden is shared across different population groups. This allows us to identify possible mechanisms through which environmental inequality reinforces the negative impact of early-life exposure to air pollution.

Our results show that changes in PM_{10} induced by rainfall variation have substantial negative effects on neonatal health outcomes, in particular weight, gestational age, preterm birth as well as overall health status of the newborn. For instance, ten additional units in the average PM_{10} level, corresponding to approximately one standard deviation, decrease birth weight and gestational age on average by about 0.5% and 0.2% respectively. A trimester-specific analysis reveals that exposure in the third trimester is the driving gestation window responsible for the detrimental neonatal health. Our baseline IV estimates are larger than OLS fixed-effects estimates, highlighting the importance of accounting for endogeneity in conventional estimators. The treatment heterogeneity analysis suggests that the adverse neonatal health effects are much larger in the subsamples of low-educated mothers and of unemployed mothers. Given that pregnant women with low socio-economic status can be more vulnerable to polluted air (Neidell, 2004) - due to differences in residence locations, access to health care or information on ambient air quality - our results imply that air pollution could amplify social disparities. These findings have direct policy implications since, despite the policy efforts so far obtained, PM_{10} concentrations exceeded the EU recommended limit values in large parts of Europe in recent years, overexposing about 20% of the population (EEA, 2019).

We present several sensitivity checks of the results that do not alter or even strengthen our conclusions. For example, we show that rainfall is not associated with anticipation or postponement of deliveries that ultimately might lead to a problem of self-selection. We also find no evidence that rainfall has a direct effect on mother's health, measured by the number of hospitalizations related to respiratory and nervous diagnoses. Moreover, we show that extending the sample to municipalities further from the monitoring stations (within 15 km) weakens our results, pointing to the importance of using detailed data on mother's location.

The reminder of the paper is as follows. Section 2.2 describes the relationship between air pollution and health and provides a brief review of the empirical findings. Section 2.3 presents the data and some descriptive evidence, while Section 2.4 illustrates the identification strategy and the estimation method. In Section 2.5 we report the results, investigate the treatment effect heterogeneity and discuss potential threats to identification. In Section 2.6 we discuss the issue of measurement error in pollution. Section 2.7 concludes.

2.2 Background

Air pollution is characterized by high spatial and temporal variability. It includes a large number of substances either directly emitted into the atmosphere such as PM, carbon monoxide (CO), sulphur dioxide (SO_2), and nitrogen dioxide (NO_2) or formed from chemical reactions in the presence of other pollutants such as ozone (O_3) (EEA, 2016). In this study we focus on PM_{10}, a particulate matter with less than 10 micrometers (μm) in aerodynamic diameter, which is considered one of the most serious hazards for human health at global level (WHO, 2013).

The adverse health effects of PM_{10} depend on the concentration and duration of exposure as well as on particles' deposition. Long-term exposure, possibly to high pollution levels, is likely to produce larger, more persistent and cumulative effects than short-term exposure. Further, the deeper the particles are deposited, the longer it takes to remove them from the human body. While there is general consensus on the mechanisms behind the health responses to fine particle inhalation among adults and children (Halliday et al., 2018; Xu et al., 2014), the biological pathway through which prenatal exposure to PM affects fetal health is more controversial. The dominant explanation is that maternal exposure to air pollution during pregnancy can affect foetuses because of its effect on maternal health. Inhaled fine particles that enter through the nose and throat can easily penetrate deep into the lungs and blood streams unfiltered. The processes responsible for adverse neonatal health are related to inflammation, oxidative stress, endocrine disruption, and insufficient oxygen transport across the placenta, to which the immature fetal cardiovascular and respiratory systems are particularly sensitive (Whyatt and Perera, 1995).[1] The resulting prenatal exposure can increase the risk of preterm birth, low birth weight or very low birth weight, linked to shortened length of gestation and/or intra-uterine growth restriction. A further potential mechanism has been suggested by a recent experimental study, which shows that PM accumulates on the placenta of exposed mothers, affecting directly the foetus (Bové et al., 2019).

The adverse health effects of extreme pollution events are well established in the epidemiological literature. Nevertheless, capturing the causal effects of prenatal air pollution on health is challenging because maternal exposure to pollution is likely to be non-random and systematically correlated with other determinants of birth outcomes. Ignoring these factors might lead to biased estimates. Several economic studies have addressed the non-random assignment of pollution relying on comparisons across siblings in a panel framework (Currie et al., 2009), mother fixed effects (Currie and Schwandt, 2016), or reduced form approaches. The reduced-form studies exploit exogenous shocks in air quality as natural experiments such as economic recessions, environmental disasters, regulations of allowed pollution levels, implementation of congestion tax or other policy changes (Chay and Greenstone, 2003; Currie et al., 2015; Currie and Walker, 2011; Lavaine and Neidell, 2017; Simeonova et al., 2018; Yang et al., 2017). However, most of these studies focus on infant mortality or children health, while less is known about neonatal health.[2] An exception is the study by Currie et al. (2009), who analyze the impact of prenatal exposure to carbon monoxide (CO), ozone (O_3), and PM_{10}, though the estimated effects for PM_{10} are much less robust than the ones obtained for CO.

[1]Epidemiological research suggests that a potential mechanism responsible for the association between prenatal air pollution and fetal health is a decline in the mitochondrial content of the placenta essential to the nourishment, growth, and development of the foetus. For a review of epidemiological literature on this topic, see Barrett (2016).

[2]More recently, a new wave of studies has examined the impact of pollution on other aspects of human life. Ebenstein et al. (2016) study the effect of elevated levels of $PM_{2.5}$ on student test scores of Israeli students; Sager (2016) documents the existence of a relationship between pollution and road safety in the UK; Lichter et al. (2017) show that variation in pollution affects professional soccer players in Germany. Finally, Isen et al. (2017) find a significant relationship between pollution exposure in the year of birth and later-life outcomes such as labor force participation and earnings at age 30, using the Clean Air Act as a source of exogenous variation in TSPs.

2.3 Data

2.3.1 Birth Data

The main dataset used in this study comes from the birth certificates (Standard Certificates of Live Births, henceforth SCLB) obtained from the Italian Ministry of Health, collected for the entire population of mothers who delivered both in public and private hospitals between 2002 and 2008 (about 500,000 births per year). The SCLB provides information on newborns' and mothers' characteristics, such as the newborn's date of birth and the geographic residence of the mother at the municipality level,[3] but also information on hospital of delivery, sex of newborn, pluriparity, and presence of neonatal pediatrician at delivery, as well as several measures of neonatal health. Background information on the mother includes demographic and labor market information, childbearing history and prenatal care.

The main outcomes of interest are measures of gestation (gestational age and preterm birth) and measures of weight at birth (birthweight, low birthweight and very low birth weight). Gestational age measures gestation duration in days.[4] Preterm birth (PTB, henceforth) is coded as a dummy equal to one if a baby is born alive before 37 completed weeks of gestation and zero otherwise. Birth weight is measured in grams, while low birth weight and very low birth weight (LBW and VLBW, respectively henceforth) are coded as dummies equal to one if weight at birth is less than 1,500 and 2,500 grams, respectively, and zero otherwise. Additionally, we employ intra-uterine growth restriction (IUGR, henceforth) as an outcome, coded as a dummy equal to one if reduced in-utero growth for a given gestational age has been diagnosed and zero otherwise. Finally, we use the Apgar score measured five minutes after birth to construct a dummy equal to one if the Apgar score is less than nine and zero otherwise (Low APGAR, henceforth).[5]

We restrict our sample to mothers aged between 15 and 45. We consider only singleton births and newborns with gestational age between 26 (who have completed the second term of pregnancy) and 44 weeks and birth weight between 500 and 6,500 grams. We drop missing values in the relevant variables and year 2002 due to an insufficient number of installed stations monitoring PM_{10} concentration levels (see next section for details on this

[3]In Italy, municipality is the finest administrative unit, with an average area of only 22 km^2. The Italian geographical administrative system is organized in regions, provinces and municipalities corresponding, respectively, to the NUTS-2, NUTS-3 and NUTS-4 Eurostat regional breakdown.

[4]Gestational age refers to the length of pregnancy after the first day of the last menstrual period and is reported in weeks. The estimation of gestational age is generally based on the last menstrual period, ultrasound or physical examination, but birth certificates do not report the exact method. The date of onset of the last menstrual period serves as a proxy for the date of conception and is calculated by subtracting the number of gestation days from the birth date.

[5]The Apgar score is a summary measure of a newborn's physical condition based on appearance, pulse, grimace, activity, and respiration and determines need for special medical care. It ranges from zero to ten and is a good predictor of survival and neurological problems at one year of age (Apgar, 1966). An Apgar score lower than nine is considered a critical threshold, below which the newborn's health might be compromised.

point). Using information on the exact birth date, the gestational age of each newborn and the maternal municipality of residence at delivery, we first expand the repeated cross-sections of birth records to a municipality-day panel. Then, we collapse birth data by mother's municipality of residence × week of child's birth to ease the computational burden and to account for the identifying variation occurring at a higher level of aggregation, i.e. the municipality. It leads to a total of 860,473 municipality by week-of-birth cells. Since we do not know exact mothers' locations within municipality, we implicitly assume that mothers living in the same municipality are exposed to the same air pollution level. We finally add information on the average gross income per capita in the municipality of mother's residence as a proxy for income (data from Agenzia delle Entrate).

2.3.2 Environmental data

We measure air pollution using data from the European Air Quality Database (Airbase), which collects information on 24h average of PM_{10} concentrations, corresponding to national ambient air quality standards, registered by monitoring stations.[6] The number of monitoring stations varies across space and time, as some municipalities have installed stations after the introduction of more stringent regulation on air quality. Only few of them operate continuously. Given concerns about the endogeneity of monitor "births" and "deaths" (Bharadwaj et al., 2017), we use data only from monitors that have more than 90% of readings in the period of study. We further exclude year 2002 because there were too few monitoring stations. This restriction leaves us with a sample of 109 monitoring stations with valid records from 2003 to 2008 for a total of 59 municipalities.[7] For municipality with more than one monitoring station like in big cities, we impute to the municipality the average of pollution concentration levels registered in all the monitoring stations belonging to that municipality. Figure 2.1 shows the geographical distribution of the selected municipalities, mainly clustered in the North, and for each municipality, the average PM_{10} concentration over the period of analysis. It shows that the municipalities in the North are more polluted than those in the South and some of them have values close to the EU annual concentration limit of 40 mcg/m^3.

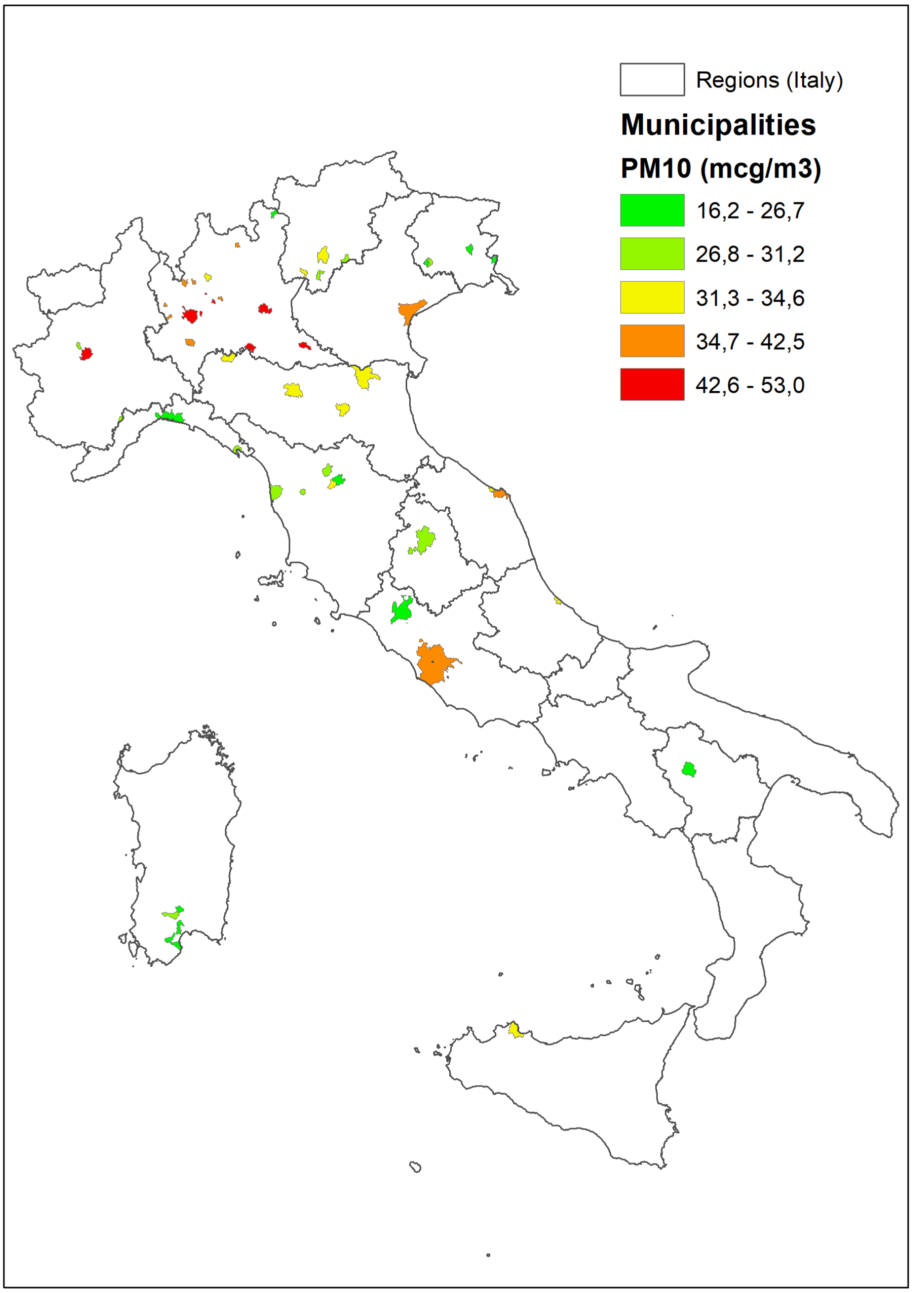
Because weather, particularly temperature, can potentially impact both pollution formation and child health (Deschênes et al., 2009), we include data on temperature and precipitations (Gridded Agro-Meteorological Data - CGMS), available on a daily basis for the whole Italian territory. We select the daily maximum and minimum temperatures, expressed in Celsius degrees (°C), since mothers are negatively affected by temperature

[6] The Airbase database is maintained by the European Environmental Agency (EEA) through the European topic center on Air Pollution and Climate Change mitigation. It contains air quality data delivered annually under the 97/101/EC Council Decision, establishing a reciprocal exchange of information and data from networks and individual stations measuring ambient air pollution within the member states.

[7] In Section 2.6 we consider the extension to municipalities whose centroid falls within a radius of 15 km from the monitors' geographical coordinates to assess the robustness of our findings to potential measurement errors in pollution.

extremes (Deschênes and Greenstone, 2011).

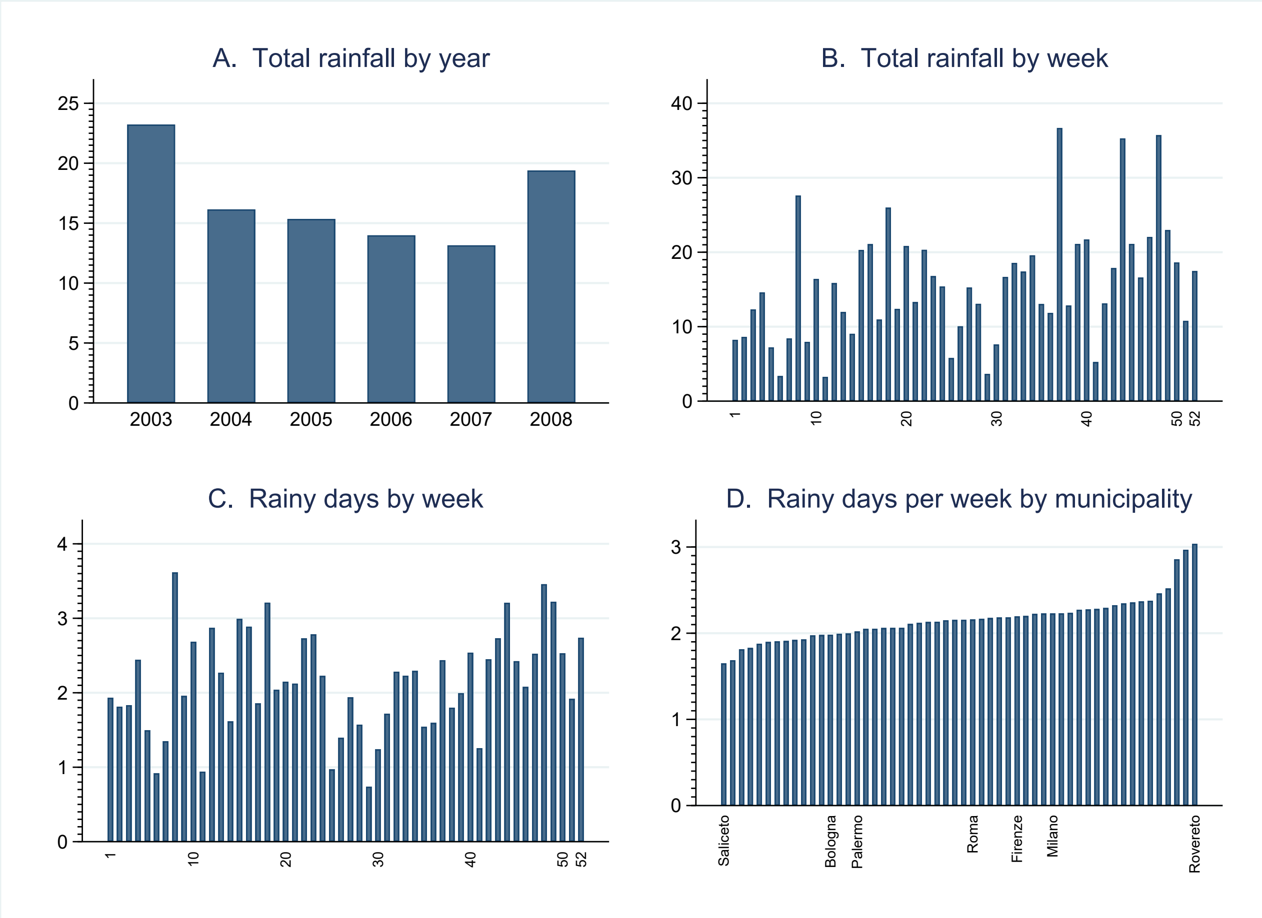
Figure 2.1: Monitoring Stations and Annual Average PM_{10} Concentration



Notes: Data are from the European Air Quality Database (Airbase), maintained by the European Environmental Agency through the European topic center on Air Pollution and Climate Change mitigation.

In addition, we compute the daily precipitation expressed in millimeters (mm) of rain by municipality. Figure 2.2 shows total precipitations by year (Panel A), week-of-year (Panel B), total number of rainy days by week-of-year (Panel C) and per week by municipality (Panel D). It turns out that in years 2003 and 2008 it rained the most (Panel A), while autumn and winter are the most rainy seasons (Panel B). With respect to the rainfall distribution, the total number of rainy days per week ranged from almost 4 at the end of February to less than 1 in August (Panel C), and from slightly above 1.5 days per week in the municipality of Saliceto in the North-West of Italy to 3 days per week in Rovereto located in the North-East (Panel D).

Figure 2.2: Rain Distribution



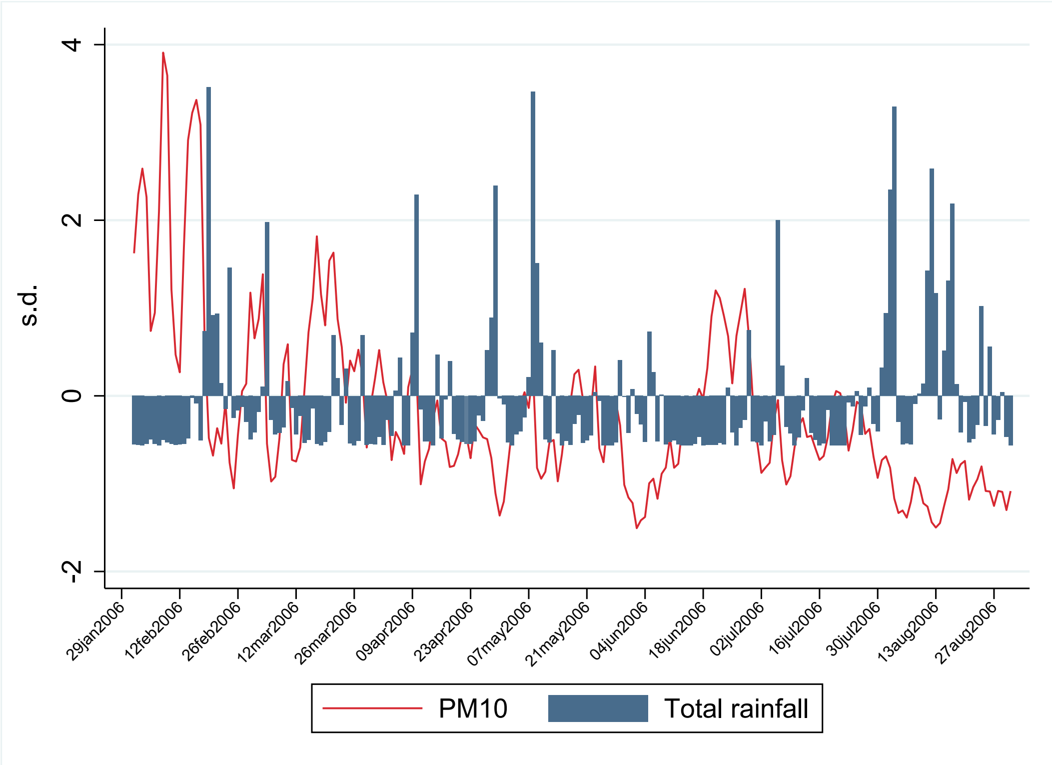
Notes: The figure shows the rain distribution across years (Panel A) and across weeks of the year (Panel B). Panels C and D show, respectively, the number of rainy days across weeks and across municipalities. Data are from the Gridded Agro-Meteorological Data (CGMS), which contain daily meteorological parameters from weather stations interpolated on a 25x25 km grid over the whole Italian territory.

Figure 2.3 links PM_{10} pollution and rainfall at daily level over a period of six months (February-August) in 2006, and shows a well-defined negative association between them: when it rains, the level of PM_{10} drops and viceversa.[8] Indeed, recent findings in atmospheric chemistry have shown that rainfall fluctuations are able to affect pollution dispersion and accumulation (Yoo et al., 2014), and even small amounts of rainfall can have strong effects on PM concentrations (Ouyang et al., 2015). Due to its chemical composition, PM strongly depends on atmospheric conditions and in some scenarios it is possible to have an even stronger dependence on meteorological conditions than on anthropogenic emissions

[8]The choice of this particular time window is only to improve exposition. The patterns, not shown here, are very similar for other time periods.

(Barmpadimos et al., 2011; Wang et al., 2015). For example, He et al. (2017) find that meteorological conditions are the primary factor driving the day-to-day variations in pollutant concentrations (including PM_{10}), explaining more than 70% of the variance in daily average pollution concentrations in China. In our empirical analysis, we use this weekly variation in rainfall to predict how pollution impacts some areas more than others in a given week.

Figure 2.3: Daily Precipitations and Average PM_{10} in 2006



Notes: The figure plots daily precipitation and PM_{10} expressed in standard deviations during 2006 averaged on all municipalities. Source: own elaboration based on the AirBase database and the Gridded Agro-Meteorological Data.

As for the SCLB data, we collapse the environmental data by municipality × week of the year for a total of 15,445 cells (59 municipalities with monitoring stations by an average of about 262 weeks over the period 2003-2008).[9] As a final step, we match the birth data with the environmental data, which leads to a final sample of 12,260 cells (54 municipalities x 227 weeks, on average). Each cell is made of mothers who live in the same municipality and give birth in the same week of the year. Since we use concentration values directly reported by monitoring stations to measure air pollution, pregnant women not living next to monitoring stations might be exposed to pollution levels other than those actually registered by the monitors. This might potentially generate a mismatch between the detected pollution level and the assigned one. The issue of misclassification will be addresses in our

[9]Given the unbalanced nature of monitoring stations data, the number of weeks varies across municipalities.

sensitivity analysis (see Section 2.6). We assign mothers' exposure to air pollution on a predetermined gestation length of 40 full weeks (expected duration). Assigning pollution exposure based on the expected duration of gestation instead of actual duration allows us to mitigate potential endogeneity due to the fact that actual gestation length represents one of the outcomes under study. Summary statistics of the baseline sample are presented in Appendix Table 2.A1.

Panel A of Appendix Table 2.A1 describes a healthy population of newborns, with an average birth weight of almost 3.3 kg, in line with the main international clinical standards (WHO, 2006). Good neonatal health is also reflected in a small portion of newborns with LBW (5%), VLBW (1%), and with low Apgar score (3%). Also the length of gestation is 273 days which corresponds to 39 weeks of gestation. Looking at the pollution variables in Panel B, we observe that the average PM_{10} level during the whole pregnancy is almost 35 mcg/m^3. The number of days with PM_{10} levels above the limit is about 54 days during pregnancy. This means that any mother during her pregnancy experiences on average more than once per week off-limit days.

Although our baseline sample represents only a fraction of the Italian newborns (about 13%), this restriction does not introduce a problem of sample selection. Indeed, Appendix Table 2.A2 compares our sample of analysis restricted to municipalities with monitoring stations (column 1) to the unrestricted sample (column 3). It turns out that the estimation sample does not substantially differ from the entire newborn population and therefore it is plausibly not affected by selection. The only differences are in the fractions of foreign, highly educated and employed mothers, which are higher in the estimation sample than in the unrestricted one. This is probably due to the fact that most of the selected municipalities are located in the North of Italy, where foreign people are mainly located and female employment is higher.

2.4 Econometric Framework

To investigate the relationship between prenatal exposure to PM_{10} and neonatal health, we first estimate the following fixed-effects model:

$$Y_{mt} = \beta PM_{10,mt} + \mathbf{X}'_{mt}\delta + \mathbf{W}'_{mt}\lambda + \gamma I_{my} + \mu_m + \theta_t + \upsilon_{mt} \tag{2.1}$$

where Y_{mt} is one of the seven outcomes of interest (listed in Section 2.3.1) for mothers giving birth in municipality m during week-of-year t. $PM_{10,mt}$ denotes i) the PM_{10} concentration level expressed in mcg/m^3 averaged over the pregnancy or ii) the number of days with PM_{10} levels above the EU limit over the pregnancy; $\mathbf{X}_{mt}$ is a vector of mother- and child-specific characteristics (listed in Section 2.3.1) in the municipality-week-of-birth cell. We control for the average maximum and minimum temperatures during pregnancy denoted by $\mathbf{W}_{mt}$. I_{my} is the average per capita income at the municipality level in year y expressed in 2005 constant Euro. It serves as a proxy for maternal living conditions, which are likely to be correlated with both air quality and neonatal health. μ_m are municipality fixed-effects that control for time-invariant, unobserved determinants of birth outcomes for mothers living in a particular municipality m. θ_t are week-of-birth fixed-effects to account for any periodic co-movements between pollution and neonatal health such as improvements in healthcare. Finally, υ_{mt} is an idiosyncratic error component.

We cluster the standard errors at the municipality level, allowing for any spatial dependence of pollution exposure within the same municipality (Cameron and Miller, 2015), and use as weights the number of births in each municipality-year. Our two-way fixed effect model allows to compare babies born both in different weeks in the same municipality, and across municipalities in the same week, after controlling for climatic and temporal variability as well as predetermined mothers' and newborns' characteristics. The coefficient of interest is β, which captures the effect of i) one additional unit in the average PM_{10} level during pregnancy or ii) one additional day with high PM_{10} level during pregnancy on neonatal health outcomes for mothers living in a certain municipality m and giving birth in a given week-of-year t, holding constant all the other variables listed in equation (2.1). We additionally estimate a trimester-specific model to test whether the estimated effects are driven by a particular period of gestation, such as the first trimester when organs' formation takes place and the neonatal health may be extremely sensitive to environmental conditions, or the third trimester during which foetuses generally gain weight. Thus, we estimate the following model:

$$Y_{mt} = \sum_{k=1}^{3} \beta_k PM_{10,k,mt} + \mathbf{X}'_{mt}\delta + \sum_{k=1}^{3} \mathbf{W}'_{k,mt}\lambda_k + \gamma I_{my} + \mu_m + \theta_t + \upsilon_{mt} \tag{2.2}$$

where $PM_{10,k,mt}$ denotes: i) the average PM_{10} level during trimester of pregnancy $k = 1, 2, 3$ or ii) days during each trimester of pregnancy with high PM_{10} levels for mothers giving birth in municipality m during week-of-year t. $\mathbf{W}_{k,mt}$ measures the averaged maximum and minimum temperatures for each trimester k. Therefore β_k captures now the effect of interest for trimester $k = 1, 2, 3$.

Maternal exposure to PM_{10} during pregnancy is likely to be correlated with many observable and unobservable determinants of prenatal development and neonatal health. Including municipality fixed-effects in μ_m will absorb any time-invariant determinants of long-run characteristics unique to a specific municipality, while including week-of-birth fixed-effects θ_t will control for short- and long-run time trends-driven determinants of neonatal health common to all deliveries in a specific week of the year. For example, if relatively disadvantaged households live in more polluted areas and have poorer health for reason unrelated to air pollution, then the municipality fixed effects will control for this time-invariant unobserved heterogeneity.

In this baseline setup a causal interpretation of the effects would rely on the assumption that neonatal health outcomes are not correlated with any unobserved maternal and municipality characteristics. If local and transitory determinants of neonatal health covary also with air pollution, our fixed effect estimates of β will be biased. This is the case, for instance, if a mother suffers from poor air quality during pregnancy and moves towards areas with better air quality. Residential sorting arising from family wealth, heterogeneity in preferences for air quality, living conditions, access to medical care and other local amenities hints at endogeneity in maternal exposure to air pollution during pregnancy (Chay and Greenstone, 2005). Moreover, local economic activity may correlate with both air pollution and neonatal health as well as fertility decisions (Dehejia and Lleras-Muney, 2004, among others). In this case, an economic expansion is likely to increase pollution concentration but also to correlate with higher income levels and/or better healthcare facilities. As a result, there would be a positive correlation between air quality and the error term, which would bias the OLS estimates downward (Knittel et al., 2016). As a matter of fact, any unobserved transitory local shocks that covary with both air pollution concentrations and neonatal health will bias our baseline estimates of β (β_k).

In order to solve the problem of endogeneity in pollution exposure, we exploit the as-good-as-random variation in local weather conditions, which are able to amplify or mitigate air pollution concentrations. Indeed, stable weather conditions along with intense local economic activity can keep concentration levels above the limit for several days, while for instance on windy days air pollution can be dispersed far away from where it is locally produced.

Previous studies have successfully employed weather conditions, in most cases wind, to instrument for air pollution. Yang et al. (2017) uses wind-direction-adjusted SO_2 emissions from a coal-fired power plant located in Pennsylvania as an instrument for SO_2 concentrations in New Jersey. Similarly, Anderson (2015) uses quasi-random variation in ultra-fine particles, nitrogen oxides, and CO generated by wind patterns near major highways. A bunch of local weather conditions has been likewise employed to instrument for PM_{10} and CO in Knittel et al. (2016), while Arceo et al. (2016) exploit thermal inversions, which are likely to lead to a temporary accumulation of certain types of pollutants, to instrument for PM_{10}, CO, SO_2, and O_3. Finally, Schlenker and Walker (2015) account for the fact that wind speed and wind direction transport air pollutants in different ways, using interactions between taxi time, wind speed, and wind angle from airports in California to identify the specific effects of CO and NO_2.

Building on these studies and on the evidence presented in Figure 2.3 which links pollution and rainfall, we rely on quasi-experimental variation in PM_{10} exposure induced by rainfall to identify the causal effect of pollution on neonatal health. This instrument allows to solve the concerns mentioned above since it is unlikely that mothers sort according to the rainfall pattern.

The mechanism underlying the relationship between rainfall and local PM_{10} concentrations can be described as follows. The transportation of suspended particles from the earth's atmosphere to the ground occurs via dry and wet deposition processes. Wet deposition consists in removing particles from the atmosphere through precipitations such as rain, fog, and snow. As a raindrop falls through the atmosphere, it can attract numerous tiny aerosol particles to its surface before hitting the ground. The process by which droplets and aerosols attract particles is called coagulation, a natural phenomenon that can act to clear the air of particle pollutants such as PM_{10} (Ardon-Dryer et al., 2015). The effect of rainfall on pollution is broadly referred to as *wash-out* or *washing effect* (Guo et al., 2016).

We estimate equation (2.1) by 2SLS in a setup that includes the same set of socio-economic and demographic variables as reported in Appendix Table 2.A1 and fixed-effects, using the total precipitation ($Rain_{mt}$) during pregnancy for mothers living in municipality m and giving birth in week t as an instrument for pollution exposure.[10] When considering PM_{10} concentration level in each trimester as in equation (2.2), the instruments are the total precipitation levels ($Rain_{k,mt}$) during trimester $k = 1, 2, 3$, respectively.

Our key identifying assumption is that rainfall does not directly affect neonatal health through factors other than PM_{10} concentrations. This seems a plausible assumption once we control for municipality and week-of-birth fixed effects, temperature and predetermined mother's characteristics. While the identifying assumption is untestable, in Section 2.5.3 we address potential concerns that could threaten the validity of our instrument.

2.5 Results

2.5.1 OLS Estimates

Table 2.1 presents the OLS estimates of the effect of average PM_{10} exposure during pregnancy (Panel A) and in each trimester of pregnancy (Panel B) on neonatal health. The results in Panel A suggest that higher average PM_{10} values adversely affect prenatal development for most of the birth outcomes. In particular, an increase in PM_{10} concentration level significantly decreases the newborn's weight and the gestational length. Symmetrically, LBW, preterm birth, and low Apgar score (Low APGAR) significantly increase, while VLBW and IUGR are unaffected by pollution exposure during pregnancy.

The trimester-specific analysis provides evidence that the most harmful effects of pollution exposure are at the early gestational stage, the so-called embryonic period, and at the late gestational stage, also known as prenatal period. Birth weight and gestational age significantly decrease, while the incidence of LBW and preterm birth significantly increase, with the effects larger in the third trimester than in the first trimester of gestation. The effects for the second trimester are much smaller in size and not statistically significant. Additionally exposure to PM_{10} in the third trimester also leads to an increased probability of low APGAR score and IUGR, though the latter effect is only weakly significant.

[10] We have also considered the average precipitation during pregnancy as a possible instrument for the average PM_{10} concentration level with very similar results. Additionally, we have used the number of rainy days as an instrument for days with PM_{10} concentration levels above the limit but the instrument turned out to be weak.

Table 2.1: OLS Estimates of the Effect of Prenatal PM_{10} Exposure on Neonatal Health

	Birth Weight (1)	LBW (2)	VLBW (3)	IUGR (4)	Gestation Days (5)	Preterm Birth (6)	Low APGAR (7)
				A. Whole Pregnancy			
PM_{10}	-10.346***	0.003***	0.000	0.001	-0.187**	0.004**	0.006**
	(1.917)	(0.001)	(0.000)	(0.001)	(0.070)	(0.002)	(0.003)
				B. By Trimester			
PM_{10}, trim. I	-5.199***	0.002*	-0.000	0.000	-0.120**	0.002**	0.003
	(1.406)	(0.001)	(0.000)	(0.001)	(0.054)	(0.001)	(0.002)
PM_{10}, trim. II	1.431	-0.001	-0.001	0.000	0.079	-0.001	-0.005
	(2.419)	(0.001)	(0.000)	(0.000)	(0.066)	(0.001)	(0.003)
PM_{10}, trim. III	-8.037***	0.003**	0.001	0.001*	-0.183***	0.004***	0.009**
	(2.170)	(0.001)	(0.000)	(0.000)	(0.059)	(0.001)	(0.004)
Mean	3,272.12	0.05	0.01	0.02	273.27	0.05	0.31
S.d.	247.97	0.11	0.04	0.08	6.09	0.11	1.73

Notes: The table reports the OLS estimates of the effect of prenatal PM_{10} exposure during pregnancy (Panel A) and in each trimester (Panel B) on neonatal health. Pollution coefficients show the effect of an increase by 10 in the average PM_{10}. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. All regressions include municipality fixed effects and week-of-birth fixed effects and control for maternal and child characteristics: age, age squared, marital status (married=1), education level (less than high school (reference), high school, more than high school), labor market attachment (employed=1), professional position (housewife (reference), self-employed, dependent employee), child's sex (female=1), neonatal pediatrician at delivery (present (reference), absent, missing), type of hospital (public (reference), private, missing), citizenship (foreign=1), previous abortions including voluntary interruptions of pregnancy as well as miscarriages (yes=1), previous deliveries (yes=1). Controls also include yearly municipal income and average minimum and maximum temperatures during pregnancy or in each trimester. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 12,260 observations. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table 2.A3 presents the OLS estimates of the effects of high level of PM_{10} (n. of days above the hazard limit) over the pregnancy and in each trimester on neonatal heath. The estimates for the whole gestational period and separately by trimesters broadly confirm the results obtained in Table 2.1, with the early and the late gestational periods playing a major role for prenatal development.

As discussed in Section 2.4, the OLS regressions control for all time-invariant characteristics that may predict neonatal heath. However, fixed effects regressions cannot control for all time-varying forms of endogeneity. For example, including municipality fixed effects will ignore time-varying determinants of birth outcomes unique to a specific municipality, like economic conditions, improved hospital facilities or other local policies. Therefore, we use rainfall during pregnancy as a source of quasi-experimental variation to identify the causal effects of prenatal PM_{10} exposure on neonatal health.

2.5.2 IV Estimates

Consistent with previous studies that use weather conditions to instrument for pollution level (see Section 2.4), Table 2.2 shows a strong relationship between rainfall and PM_{10} concentrations.[11] Ten additional units in the total precipitation during pregnancy decreases the average PM_{10} concentration level by about 0.16 mcg/m^3, which corresponds to a 0.45% reduction (column 1 of Table 2.2). When considering the first stage estimates by trimester, we use three instruments (total rainfall in the I, II and III trimester) for the average PM_{10} concentration level in the I, II and III trimester, respectively (columns 2 to 4). The point estimates on the diagonal show that rainfall in a specific trimester is a strong predictor of particle pollution concentration in the same trimester. Interestingly, the coefficients on rainfall precipitations one trimester backwards are often statistically significant, though much smaller in size. This evidence is in line with studies suggesting that the relationship between rainfall and atmospheric particle concentrations might be non-linear with a lag effect (Barmpadimos et al., 2011), implying that the *wash-out* effect is long-standing but potentially decreasing over time due to new local PM_{10} emissions into the atmosphere. The F-statistics are in general above the threshold of 10 as indicated in Staiger and Stock (1997), confirming that rainfall is not a weak instrument.[12] In Appendix Table 2.A4 we report the first stage estimates for high level of PM_{10} during pregnancy. Column 1 shows that ten additional units in the total precipitation during pregnancy decreases the number of days with high level of pollution by about 0.55 days, which corresponds to a 1.01% reduction.

Table 2.2: First Stage Estimates of the Effect of Total Rainfall on PM_{10} Concentration

	PM_{10}			
	Whole Pregnancy (1)	Trimester I (2)	Trimester II (3)	Trimester III (4)
Rainfall during pregnancy	-0.156**			
	(0.049)			
Rainfall during trim. I		-0.288***	-0.108**	0.015
		(0.074)	(0.049)	(0.036)
Rainfall during trim. II		-0.068	-0.323***	-0.139**
		(0.057)	(0.076)	(0.046)

[11] We proxy rainfall during gestation by total rain because in the meteorological literature this measure is considered to carry more information on the precipitations dynamics (Ouyang et al., 2015). Using average rainfall yields similar results though.

[12] The only exception is the F-statistics for average pollution during whole pregnancy in column (1).

Rainfall during trim. III		0.026	-0.078	-0.263***
		(0.059)	(0.057)	(0.071)
F-statistics	9.82	26.99	38.15	16.27

Notes: The table reports the first stage estimates of the effect of total rain on average PM_{10} during pregnancy and in each trimester. The coefficients show the effect of an increase by 10 mm in total rain. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. All regressions include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy or in each trimester. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 12,260 observations. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.3 and Appendix Table 2.A5 report the IV estimates of the effects of PM_{10} exposure (in average level and in days of high pollution level, respectively) on neonatal health. Compared to the OLS estimates, the IV estimates are about two to four times as large and allow to identify the most susceptible period of prenatal exposure. Table 2.3 shows that health outcomes related to weight (birth weight, LBW, VLBW) and gestational duration (gestational age and preterm birth) are significantly affected by prenatal exposure to PM_{10}. In particular, ten additional units in the average PM_{10} concentration level would decrease birth weight by about 17.2 grams and gestational age by almost 0.6 days, a reduction of about 0.5% and 0.2%, on average, respectively. Moreover, the same increase in PM_{10} concentration level would increase the probability of LBW, VLBW, and preterm birth by about 0.009 (or 18%), 0.002 (or 20%), and 0.01 (or 20%), respectively. These results suggest that newborns at risk of low, very low birth weight and premature birth are most likely to be affected by particle pollution while *in utero*. Concerning the prevalence of IUGR and low APGAR, we do not find a statistically significant relationship.

Table 2.3: IV Estimates of the Effect of Prenatal PM_{10} Exposure on Neonatal Health

	Birth Weight (1)	LBW (2)	VLBW (3)	IUGR (4)	Gestation Days (5)	Preterm Birth (6)	Low APGAR (7)
				A. Whole Pregnancy			
PM_{10}	-17.210***	0.009***	0.002**	0.003	-0.559**	0.010***	0.015
	(5.089)	(0.003)	(0.001)	(0.002)	(0.224)	(0.004)	(0.009)
				B. By Trimester			
PM_{10}, trim. I	-8.799	0.004	0.000	0.002	-0.250*	0.001	-0.007
	(6.406)	(0.003)	(0.001)	(0.002)	(0.146)	(0.003)	(0.016)

PM_{10}, trim. II	11.949	-0.009**	-0.001	-0.001	0.292	-0.003	-0.014
	(10.706)	(0.005)	(0.002)	(0.001)	(0.298)	(0.006)	(0.017)
PM_{10}, trim. III	-26.601***	0.017***	0.003**	0.003	-0.818***	0.014**	0.038**
	(8.212)	(0.004)	(0.001)	(0.002)	(0.210)	(0.006)	(0.015)
Mean	3,272.12	0.05	0.01	0.02	273.27	0.05	0.31
S.d.	247.97	0.11	0.04	0.08	6.09	0.11	1.73

Notes: The table reports the IV estimates of the effect of average PM_{10} exposure during pregnancy (Panel A) and in each trimester of pregnancy (Panel B) on neonatal health using total rain over pregnancy and total rain in each trimester as an instrument, respectively. Pollution coefficients show the effect of an increase by 10 in the average PM_{10} expressed in mcg/m3. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. All regressions include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy or in each trimester. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 12,260 observations. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.3 also presents the IV estimates from the trimester-specific model, which suggests that the total effect observed for the entire pregnancy period is mainly driven by prenatal exposure during the last gestational period. Since foetuses gain about 200 grams in weight per week in the final month of pregnancy (Cunningham et al., 2010), a 0.82-days reduction in gestation would translate into a reduction of 23.4 grams in weight, which is very close to our estimate of the impact on birth weight of 26.6 grams. Therefore, the reduction in birth weight in the third trimester seems to be mainly due to shorter gestation, rather than to growth retardation. In support of this hypothesis, we do not find a statistically significant relationship between exposure to PM_{10} and IUGR.

Appendix Table 2.A5 reports the IV estimates for days with high levels of PM_{10} over the pregnancy and in each trimester. The results follow a similar pattern as in Table 2.3. In particular, results by trimester of gestation largely confirm that the largest adverse effects on neonatal health arise in the third trimester.

In explaining the pattern of our causal findings we rely on epidemiological studies on placental growth and function across gestational periods in response to prenatal exposure to PM_{10} (Janssen et al., 2012; van den Hooven et al., 2012, among others). These studies show that placental mitochondria, which is important for the proper formation and functioning of the placenta, are negatively associated with exposure to PM_{10} while *in utero*. This implies that the damaged placental content may be one of the underlying mechanisms explaining our findings. More recently, epidemiological literature has underscored the importance of the process of PM deposition on the mother's placenta (Bové et al., 2019). If the deposition process of PM particles is proportional to the size of the placenta increasing across trimesters, then PM_{10} effects might be more harmful in the late gestational phase when the foetus gains weight and size more rapidly.

Next, we investigate whether our results are in line with the literature linking maternal exposure to PM_{10} and birth outcomes. As mentioned in Section 2.2, the economic literature on neonatal health is rather scarce, focusing in most cases on infant mortality in response to exposure to other air pollutants, mainly SO_2 and CO. Currie et al. (2009) is one of the few exceptions because they investigate the effects of exposure to CO, O_3, and PM_{10}, both during pregnancy and after birth. In their analysis of health impacts of air pollution, the authors find consistent effects only for CO. Their estimates for prenatal exposure to CO in the third trimester of gestation suggest that a one unit increase in the mean level of CO would reduce birth weight by about 0.5%, increase LBW by almost 8%, and shorten gestation by about 0.2%. These estimates are roughly two to six times larger than ours, though they refer to CO effects and not to PM_{10}. Similarly, Lavaine and Neidell (2017) find that birth weight and gestational age of the newborns are particularly affected by exposure to SO_2 during the first and the third trimesters of pregnancy, with the estimates in the third trimester being much larger than ours for PM_{10}. Finally, in the study by Currie and Walker (2011) focusing on the reduction of air emissions caused by the introduction of electronic toll collection (E-ZPass), the associated reduced NO_2 levels substantially decreased prematurity and LBW among mothers within two kilometres of a toll plaza by 10.8% and 11.8%, respectively, relative to mothers living further away from a toll plaza. Although the estimates for CO, SO_2, and NO_2 found in these studies are larger than those we find for PM_{10}, Chay and Greenstone (2003) allow to reconcile all the results. Indeed, focusing on the effects of a decline in TSP (Total Suspended Particle), which consists of even larger particles than PM_{10}, on birth weight and infant mortality, they find that the effects on birth weight are much smaller for TSP than for CO, SO_2, and NO_2 obtained from other studies, which suggests that our PM_{10} effects are plausible.

Finally, we also contribute to the debate on environmental justice and investigate to what extent the burdens of prenatal exposure to air pollution are equally shared across various socio-economic groups in the population (Banzhaf et al., 2019).[13] The idea underlying the concept of environmental inequality is that more disadvantaged groups, for instance low-income groups or ethnic minorities, bear disproportionate environmental burdens, in the form of polluted air and water, unsafe jobs, and under-enforcement of environmental laws (Evans and Kantrowitz, 2002). A number of measures for socio-economic status (SES) has been adopted in the literature, including income, wealth, education, labor force status, and race/ethnicity to show that health effects of air pollution are larger for low SES groups (Neidell, 2004; Hsiang et al., 2019, among others). For instance, low SES groups may be more likely to live in areas with higher levels of air pollution, next to industrial districts for example, and at the same time less likely to move from one area to another to avoid pollution externalities. In this scenario, environmental inequality can reinforce the

[13] European policy makers have only recently included the notions of environmental justice and environmental equality in their goals (EEA, 2018), which have been part of the US policy objects for almost two decades (Laurent, 2011).

negative impact of early-life exposure to air pollution on children's health and their future educational and labor market outcomes, especially when exposure starts already *in utero* (Currie, 2011; Isen et al., 2017).

To investigate this issue, we analyze the effects of prenatal exposure to PM_{10} on neonatal health by mother's labor market status and education level. We define low SES mothers as unemployed (vs employed) or less-educated (vs more educated). Table 2.4 presents the IV estimates of the effect of average PM_{10} exposure during pregnancy on neonatal outcomes by maternal employment status. The results indicate that, though the effect is significant for both groups, babies born to unemployed mothers suffer more from particulate pollution exposure especially in terms of birth weight and, to a lesser extent, of APGAR score. We observe the same pattern when estimating the effects of PM_{10} for less and more educated mothers in Table 2.5, with babies born to less educated mothers being more exposed to negative pollution effects.[14] This result is consistent with Yang and Chou (2018) who have shown that less educated mothers benefited more from the shutdown of a power plant in Pennsylvania in terms of a greater reduction in preterm birth and LBW probabilities as well as greater increases in average birth weight and gestational age.

Table 2.4: IV Estimates of the Effect of Prenatal PM_{10} Exposure on Neonatal Health by Employment Status

	Birth Weight (1)	LBW (2)	VLBW (3)	IUGR (4)	Gestation Days (5)	Preterm Birth (6)	Low APGAR (7)
				A. Employed Mothers			
PM_{10}	-7.920*	0.006**	0.001	0.004	-0.479**	0.008**	-0.001
	(4.534)	(0.003)	(0.001)	(0.003)	(0.193)	(0.003)	(0.005)
Mean	3,272.61	0.04	0.006	0.02	273.31	0.05	0.02
S.d.	266.87	0.11	0.04	0.087	6.53	0.11	0.16
				B. Unemployed Mothers			
PM_{10}	-30.048**	0.008*	0.003	0.000	-0.633*	0.011	0.026*
	(13.761)	(0.004)	(0.002)	(0.003)	(0.352)	(0.007)	(0.015)
Mean	3,280.48	0.05	0.01	0.02	273.17	0.06	0.04
S.d.	319.19	0.14	0.05	0.10	7.65	0.14	0.20

[14] Appendix Tables 2.A6 and 2.A7 convey the same message when investigating the effects of days of high pollution exposure.

Notes: The table reports the IV estimates of the effect of average PM_{10} exposure during pregnancy on neonatal health for employed mothers (Panel A) and for unemployed mothers (Panel B) using total rain over pregnancy as an instrument. Pollution coefficients show the effect of an increase by 10 in the average PM_{10}. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. Both panels include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 11,676 observations for employed mothers and 10,100 observations for unemployed mothers. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.5: IV Estimates of the Effect of Prenatal PM_{10} Exposure on Neonatal Health by Education Level

	Birth Weight	LBW	VLBW	IUGR	Gestation Days	Preterm Birth	Low APGAR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A. More Educated Mothers			
PM_{10}	-7.433	0.005*	0.001	0.004*	-0.351*	0.009***	0.010
	(5.240)	(0.003)	(0.001)	(0.002)	(0.177)	(0.003)	(0.010)
Mean	3,282.53	0.05	0.01	0.02	273.49	0.05	0.03
S.d.	259.55	0.11	0.04	0.08	6.30	0.11	0.17
				B. Less Educated Mothers			
PM_{10}	-30.173***	0.012**	0.004*	0.001	-0.589*	0.010	0.007
	(9.785)	(0.005)	(0.002)	(0.003)	(0.336)	(0.006)	(0.010)
Mean	3,257.95	0.06	0.01	0.02	272.78	0.06	0.05
S.d.	333.25	0.15	0.05	0.10	8.04	0.16	0.21

Notes: The table reports the IV estimates of the effect of average PM_{10} exposure during pregnancy on neonatal health for more educated mothers (Panel A) and for less educated mothers (Panel B) using total rain over pregnancy as an instrument. Pollution coefficients show the effect of an increase by 10 in the average PM_{10}. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. Both panels include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 11,601 observations for more educated mothers and 9,963 observations for less educated mothers. * significant at 10%; ** significant at 5%; *** significant at 1%.

A plausible explanation of our findings could be that unemployed mothers spend more time outdoors being more exposed to air pollution, while employed mothers spend a large portion of time at work. Therefore, employed mothers may be more likely to enjoy better air quality in presence of air conditioning that filters air inhaled at work. More educated mothers may be also better informed about air quality than less educated mothers and undertake actions to compensate for the adverse environmental conditions.

2.5.3 Threats to Validity

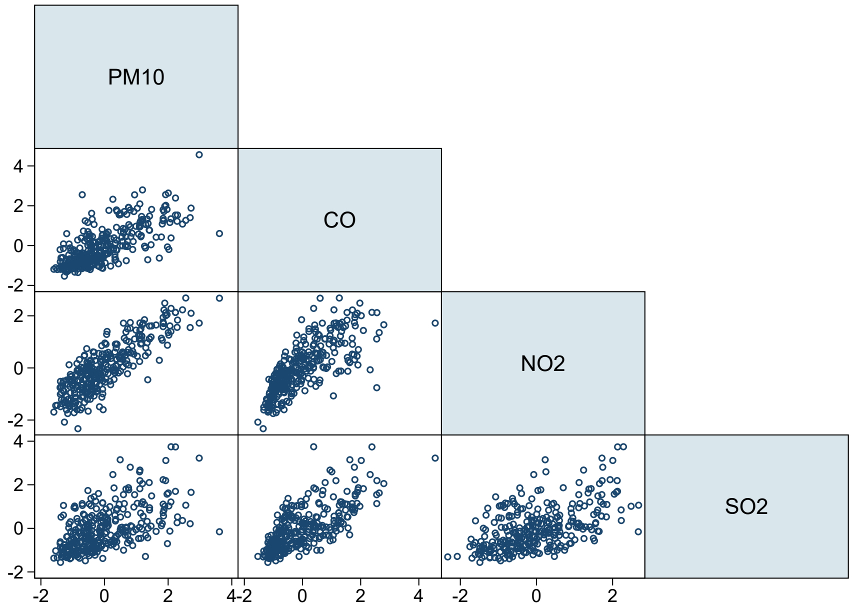
A possible concern on the validity of our estimates is that the sources of certain pollutants are similar and thus often vary jointly, which could make it difficult to establish which pollutant is responsible for the adverse health effects. If the observed PM_{10} concentrations are correlated with other pollutants not considered in this study, then our estimates are likely to be upward biased and the true effect of PM_{10} overestimated (Lavaine and Neidell, 2017). Figure 2.4 plots the correlations at weekly level between PM_{10}, CO, NO_2, and SO_2, obtained from the Airbase database. It shows that PM_{10} is highly correlated with other pollutants, coming from many of the same sources, posing the problem to determine which pollutant causes the observed changes in neonatal health. Recently, Schlenker and Walker (2015) have estimated the contemporaneous effect of multiple pollutants simultaneously (CO_2, NO_2 and O_3) using the interactions between taxi time and meteorological conditions to pin down the direct effect of each pollutant, while holding the others constant. However, they cannot isolate particulate matter effects due to the limitations in spatial and temporal coverage of monitors registering PM in California. Our research design provides a possible solution to this identification problem because the rainfall instrument has an impact only on PM. Indeed, when testing the correlation between rainfall and other pollutants such as CO, NO_2, or SO_2, our instrument does not show any statistical power.[15] This result suggests that the *wash-out* effect of rain applies exclusively to particle pollution and therefore we are able to isolate the health effects deriving only from variation in PM_{10} concentrations, induced by rainfall.

A possible threat to the validity of the estimates in Table 2.3 and Appendix Table 2.A5 comes from the fact that if rainfall, conditional on other covariates, directly affected neonatal health, our identifying assumption would be violated. This would be the case, for example, if the health status of pregnant women is to some extent affected by rainfall variation, with an indirect effect on prenatal health leading to worse birth outcomes. To exclude the existence of direct effects we run three tests.

First, suppose that hospital personnel or pregnant women have a preference for sunny days and systematically avoid rainy days for deliveries. In case of severe rainy forecast, this preference would lead to a reschedule of deliveries either to an earlier or a later date. This avoidance behavior would generate a problem of sample selection, acting through

[15] The results are available upon request.

Figure 2.4: Correlation between Pollutants

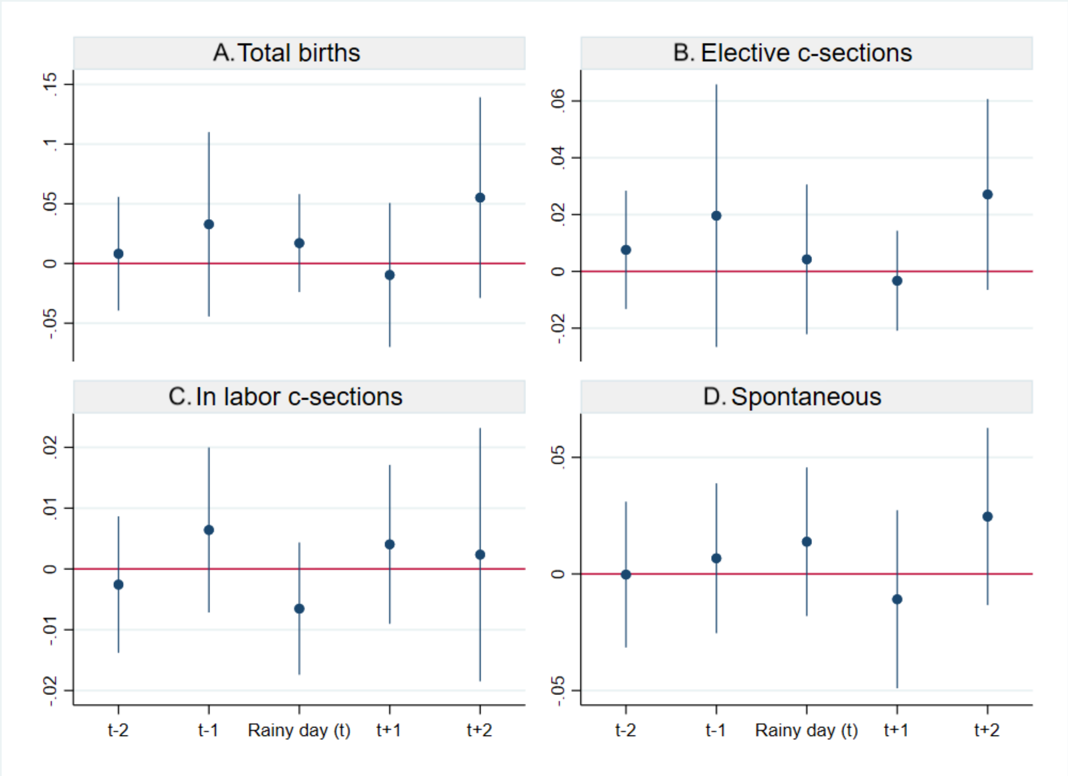


Notes : The figure shows the weekly correlations between PM_{10}, CO, NO_2 and SO_2, averaged over municipalities and expressed in standard deviations. Source: own elaboration based on the AirBase database.

anticipation or postponement of deliveries. In this case our instrument would not be as-good-as-random anymore. To test this hypothesis, after collapsing the dataset at municipality and delivery day level, we regress the total number of births on five rainy dummies (one indicating whether the delivery day was rainy or not, two daily lags to capture the anticipation effects and two daily leads to capture the postponement effects) controlling for municipality and day-of-week fixed effects. We then disentangle the effect by type of delivery to isolate the impact of rainy days on scheduled c-section births, which might be more subject to rescheduling. Figure 2.5 shows that rainfall does not significantly affect the number of births in a specific day. In particular scheduled c-section births are unlikely to be rescheduled in response to weather preferences. Second, in line with Angrist and Pischke (2009), we look at the reduced form relationship between rainfall and neonatal health. We regress the total rainfall during pregnancy on neonatal outcomes, separately for municipalities with above and below mean PM_{10} to test for possible direct effects of rainfall on prenatal development.[16] It is plausible to assume that in the absence of this potential direct effect, babies born in municipalities with better air quality, i.e. a relatively low PM_{10} concentration level, should not be affected by rainfall during pregnancy. On the contrary, weather conditions should have a significant positive impact on health outcomes of babies born in more polluted municipalities, i.e. with relatively high PM_{10} concentra-

[16] To reduce endogeneity, we separate the municipalities according to the mean of PM_{10} calculated for each municipality during 2002, i.e. one year before our period of analysis. Unfortunately, we cannot use years before 2002 due to the limited number of monitoring stations. Identical results are obtained when considering the PM_{10} mean based on the period 2003-2008 (full sample) or using the median PM_{10} concentration level. All results are available upon request.

Figure 2.5: Daily Rainfall and Birth Rates



Notes : The figure plots point estimates of the effect of daily rainfall on birth rates by types of delivery controlling for municipality and day of week fixed effects. Panel A shows the effects on total births, Panel B the effects on elective c-section deliveries, Panel C the effects on in labor c-section deliveries, while Panel D the effects on spontaneous deliveries.

tion level. Table 2.6 shows significant effects only in more polluted municipalities, which supports the idea that rainfall positively affects neonatal health exclusively through its impact on air pollution mitigation.

Table 2.6: The Effect of Rainfall on Neonatal Health by Level of Pollution

	Birth Weight (1)	LBW (2)	VLBW (3)	IUGR (4)	Gestation Days (5)	Preterm Birth (6)	Low APGAR (7)
	A. Municipalities with High Pollution						
Rainfall	0.0387***	-0.0001*	-0.0001***	-0.0001**	0.0005*	-0.0001*	-0.0003
	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	B. Municipalities with Low Pollution						
Rainfall	0.0253	-0.0000	0.0000	0.0000	0.0006	-0.0000	0.0003
	(0.016)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: The table reports the estimates of the effect of total rainfall during pregnancy on neonatal health in municipalities with above mean PM_{10} level (Panel A) and below mean PM_{10} level (Panel B). The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. All regressions include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 4,361 observations for municipalities with high pollution and 2,844 observations for municipalities with low pollution. * significant at 10%; ** significant at 5%; *** significant at 1%.

Third, we test whether there is evidence of an increased number of hospitalizations in the female population during rainy days. The underlying idea is to identify the direct effect, if any, of rainfall days on pregnant mother's health which in turn might impact the foetus. In particular, we analyze the effect of rainfall days in the day of hospitalization and up to two days before on hospitalizations of women related to a particular diagnosis. We use the Hospital Discharge Data provided by the Italian Ministry of Health, which include detailed information on daily hospitalizations in both public and private hospitals for the whole Italian population. We restrict the sample to women aged 15-45 with hospitalization episodes related to four main categories of diseases: pneumonia and influenza, acute pulmonary diseases, mental diseases, and nervous system disorders.[17] Our period of analysis is from 2004 to 2008 for a total of 14,395,843 municipality-day cells (hospitalization data are not provided in 2003). The results in Appendix Table 2.A8 show not statistically significant coefficients for any of the diagnoses considered. The only exception is the effect of rainfall on hospitalizations due to nervous system disorders, though the magnitude is negligible and the coefficient is only weakly significant.

2.6 Measurement Error in Pollution Exposure

One potential issue in our analysis is related to measurement error in pollution exposure. Indeed, given that we use concentration values directly reported by monitoring stations to measure air pollution, pregnant women not living next to monitoring stations might be exposed to pollution levels different from those actually registered by the monitors, potentially generating a mismatch between the detected pollution level and the assigned one. However, we argue that it is unlikely to be a concern in our context because the geographical unit of analysis, i.e. the municipality, is extremely fine. This implies that although the exact mother's address is not available, the measurement error is minimized when matching mothers with pollution data.

[17] In selecting women hospitalized with these diagnoses, we follow the ICD-9 classification. The complete set of selected ICD-9 codes is available upon request.

Nevertheless, we test the robustness of our results by extending our sample to include municipalities whose centroid falls within a radius of 15 km from the monitors' geographical coordinates as in Currie et al. (2009) and Currie and Neidell (2005). This procedure allows us to expand the sample coverage to 1,029 municipalities and 13,143 municipality × week-of-birth cells. Comparison between column 1 (estimation sample) and column 5 (extended sample) in Appendix Table 2.A2 shows that the extended sample is still comparable to the estimation sample in terms of observable characteristics. If the distance to a monitoring station matters for the accuracy of pollution measures, then we expect weaker results in the extended sample. Indeed, Appendix Table 2.A9 and Appendix Table 2.A10 confirm that the estimates are smaller and partially wrong-signed, but not statistically significant.

A second problem relates to individual temporary mobility. Lack of information on maternal location throughout pregnancy might introduce an exposure misclassification, leading to biased results towards the null. Since we do not have information on temporary mobility of mothers during pregnancy we assume that it is negligible. In our context this is a plausible assumption because pregnant women tend to experience low mobility rates and preference for short distances (Chen et al., 2010).[18] More recently, using detailed information on all residential addresses between the date of conception and date of delivery, Warren et al. (2017) have shown that ignorance of residential mobility during pregnancy does not lead to exposure misclassification. A further indication that residential mobility is not an issue in our context comes from the Italian census data, which points to generally high percentages of owned dwellings, ranging from 61.9% in the region of Campania to 78.8% in the region of Molise, as registered in 2001 (ISTAT, 2001). Hence, we can expect relatively low mobility among resident families. For all these reasons underestimation of the true effects of pollution on neonatal health due to residential misclassification does not seem highly relevant in our case.

A potential measurement error problem in the pollution assignment might also be due to the inclusion in our sample of mothers with region of hospital different from region of residence, which might generate attenuation bias in the estimates. From the initial total births population of approximately 3,400,000 mothers, only 162,244 of them report region of residence different from hospital of delivery's region (less than 5%).[19] To check to what extent our estimates are sensitive to the inclusion of mothers declaring region of residence different from region of hospital, we exclude these observations from the estimation sample. As expected, we find slightly larger estimates with the new sample. The results are presented in Appendix table 2.A11.

[18] Potential residential mobility during pregnancy is defined as any change of address between the estimated date of conception and pregnancy termination. A few studies report the frequency, distance, and timing of moves during pregnancy (Bell and Belanger, 2012, among others). The mobility rates range from 9% to 32%, with the highest mobility during the second trimester. Most moves occur once and within short distances, with a median distance of less than 10 km.

[19] We cannot reduce the mismatch at the provincial or municipal level because mothers might choose to deliver in a hospital located in a different province in the same region of residence or might be forced to move to the closest municipality with a hospital if their municipality of residence lacks one. Indeed, out of almost 8,100 municipalities in Italy, less than 800 have a hospital with a maternity ward.

A final concern relates to the magnitudes of the estimated effects since our analysis is based on population data belonging to a period in which the levels of particulate concentrations were slightly higher than nowadays. However, the health response to lower PM_{10} levels experienced today might be of similar order of magnitude if our estimates reflect lower bounds of the true effects. Indeed, this is likely to be the case since we do not control for selective mortality, implying that the population of surviving newborns is positively selected.

2.7 Conclusion

This paper identifies the causal effects of prenatal exposure to PM_{10} on neonatal health in the early 2000s in Italy. We combine newly available data with arguably exogenous changes in air pollution that originate from rainfall and are unforeseen by the population. An instrumental variable setting allows us to isolate the effect of PM_{10} on neonatal health, net of the influence of other air pollutants which are known to covary with PM_{10}.

We find that prenatal exposure to PM_{10} negatively affects newborn health, with most of the effect concentrating during the third trimester of gestation. Both average PM_{10} concentrations and days with high levels of PM_{10} yield similar results. Our analysis of the treatment effect heterogeneity also shows that babies born to socially disadvantaged mothers are more vulnerable, implying that the health impacts of air pollution on newborns are unequally distributed. This knowledge gain is of direct policy relevance. Indeed, if disadvantaged families are more likely to live in more polluted areas, exposure to air pollution may contribute to explaining the existing differences in educational attainment and labor market outcomes across different socio-economic groups, or more generally, explaining social and economic inequality (Isen et al., 2017; Hsiang et al., 2019, among others). This implies that better air quality may help improve environmental conditions in low-income families and thus align endowments at birth, giving a fair chance in life to every child. If economic and environmental inequality reinforce each other, then actions directed to improve air quality may serve not only as environmental health policies but also as effective social policies to reduce economic inequality. Finally, if air quality is viewed as a factor of production which, similar to technology, is able to impact how other production factors such as labor, capital, and land can be combined to generate output, a better air quality may also contribute to economic progress.

References

Almond, D., K. Y. Chay, and D. S. Lee (2005). The costs of low birth weight. *The Quarterly Journal of Economics 120*(3), 1031–1083.

Anderson, H. R. (2009). Air pollution and mortality: A history. *Atmospheric Environment 43*(1), 142–152.

Anderson, M. L. (2015). As the wind blows: The effects of long-term exposure to air pollution on mortality. Working Paper 21578, National Bureau of Economic Research.

Angrist, J. and J.-S. Pischke (2009). *Mostly harmless econometrics: An empiricist's guide.* Princeton: Princeton University Press.

Apgar, V. (1966). The newborn (Apgar) scoring system: Reflections and advice. *Pediatric Clinics of North America 13*(3), 645–650.

Arceo, E., R. Hanna, and P. Oliva (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City. *The Economic Journal 126*(591), 257–280.

Ardon-Dryer, K., Y.-W. Huang, and D. J. Cziczo (2015). Laboratory studies of collection efficiency of sub-micrometer aerosol particles by cloud droplets on a single-droplet basis. *Atmospheric Chemistry and Physics 15*(16), 9159–9171.

Banzhaf, S., L. Ma, and C. Timmins (2019). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives 33*(1), 185–208.

Barker, D. J., C. Osmond, P. Winter, B. Margetts, and S. J. Simmonds (1989). Weight in infancy and death from ischaemic heart disease. *The Lancet 334*(8663), 577–580.

Barmpadimos, I., C. Hueglin, J. Keller, S. Henne, and A. Prévôt (2011). Influence of meteorology on PM10 trends and variability in Switzerland from 1991 to 2008. *Atmospheric Chemistry and Physics 11*(4), 1813–1835.

Barrett, J. R. (2016). Prenatal air pollution and reduced birth weight: Decline in placental mitochondria as a potential mechanism. *Environmental Health Perspectives 124*(5), A98.

Bell, M. L. and K. Belanger (2012). Review of research on residential mobility during pregnancy: Consequences for assessment of prenatal environmental exposures. *Journal of Exposure Science and Environmental Epidemiology 22*(5), 429–438.

Bharadwaj, P., M. Gibson, J. G. Zivin, and C. Neilson (2017). Gray matters: Fetal pollution exposure and human capital formation. *Journal of the Association of Environmental and Resource Economists 4*(2), 505–542.

Bharadwaj, P., P. Lundborg, and D.-O. Rooth (2017). Birth weight in the long run. *Journal of Human Resources 53*(1), 189–231.

Bové, H., E. Bongaerts, E. Slenders, E. M. Bijnens, N. D. Saenen, W. Gyselaers, P. Van Eyken, M. Plusquin, M. B. J. Roeffaers, M. Ameloot, and T. S. Nawrot (2019). Ambient black carbon particles reach the fetal side of human placenta. *Nature Communications 10*(1), 3866.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Casey, B. M., D. D. McIntire, and K. J. Leveno (2001). The continuing value of the Apgar score for the assessment of newborn infants. *New England Journal of Medicine 344*(7), 467–471.

Chay, K. Y. and M. Greenstone (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *The Quarterly Journal of Economics 118*(3), 1121–1167.

Chay, K. Y. and M. Greenstone (2005). Does air quality matter? Evidence from the housing market. *Journal of Political Economy 113*(2), 376–424.

Chen, L., E. M. Bell, A. R. Caton, C. M. Druschel, and S. Lin (2010). Residential mobility during pregnancy and the potential for ambient air pollution exposure misclassification. *Environmental Research 110*(2), 162–168.

Cunningham, F. G., K. J. Leveno, S. L. Bloom, J. C. Hauth, D. J. Rouse, and C. Y. Spong (2010). *Williams Obstetrics. Fetal growth and development*, Volume 23. McGrawHill, NY, USA.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature 47*(1), 87–122.

Currie, J. (2011). Inequality at birth: Some causes and consequences. *American Economic Review 101*(3), 1–22.

Currie, J., L. Davis, M. Greenstone, R. Walker, et al. (2015). Environmental health risks and housing values: Evidence from 1,600 toxic plant openings and closings. *American Economic Review 105*(2), 678–709.

Currie, J. and E. Moretti (2007). Biology as destiny? Short-and long-run determinants of intergenerational transmission of birth weight. *Journal of Labor Economics 25*(2), 231–264.

Currie, J. and M. Neidell (2005). Air pollution and infant health: What can we learn from California's recent experience? *The Quarterly Journal of Economics 120*(3), 1003–1030.

Currie, J., M. Neidell, and J. F. Schmieder (2009). Air pollution and infant health: Lessons from New Jersey. *Journal of Health Economics 28*(3), 688 – 703.

Currie, J. and H. Schwandt (2016). The 9/11 dust cloud and pregnancy outcomes: A reconsideration. *Journal of Human Resources 51*(4), 805–831.

Currie, J. and R. Walker (2011). Traffic congestion and infant health: Evidence from E-Zpass. *American Economic Journal: Applied Economics 3*(1), 65–90.

Dehejia, R. and A. Lleras-Muney (2004). Booms, busts, and babies' health. *The Quarterly Journal of Economics 119*(3), 1091–1130.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review 109*(12), 4178–4219.

Deschênes, O. and M. Greenstone (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *American Economic Journal: Applied Economics 3*(4), 152–85.

Deschênes, O., M. Greenstone, and J. Guryan (2009). Climate change and birth weight. *American Economic Review 99*(2), 211–17.

Deschenes, O., M. Greenstone, and J. S. Shapiro (2017, October). Defensive investments and the demand for air quality: Evidence from the nox budget program. *American Economic Review 107*(10), 2958–89.

Ebenstein, A., V. Lavy, and S. Roth (2016). The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics 8*(4), 36–65.

EEA (2016). Air quality in Europe - 2016 report. Technical Report 28/2016, European Environment Agency, Luxembourg: Publications Office of the European Union.

EEA (2018). Unequal exposure and unequal impacts: Social vulnerability to air pollution, noise and extreme temperatures in Europe. Technical Report 22/2018, European Environment Agency, Luxembourg: Publications Office of the European Union.

EEA (2019). Air quality in europe - 2019 report. Technical Report 10/2015, European Environment Agency,, Luxembourg: Publications Office of the European Union.

Elder, T., D. N. Figlio, S. A. Imberman, and C. Persico (2019). The role of neonatal health in the incidence of childhood disability. Working Paper 25828, National Bureau of Economic Research.

Evans, G. W. and E. Kantrowitz (2002). Socioeconomic status and health: The potential role of environmental risk exposure. *Annual Review of Public Health 23*(1), 303–331.

Figlio, D., J. Guryan, K. Karbownik, and J. Roth (2014). The effects of poor neonatal health on children's cognitive development. *American Economic Review 104*(12), 3921–3955.

Guo, L.-C., Y. Zhang, H. Lin, W. Zeng, T. Liu, J. Xiao, S. Rutherford, J. You, and W. Ma (2016). The washout effects of rainfall on atmospheric particulate pollution in two Chinese cities. *Environmental Pollution 215*, 195–202.

Halliday, T. J., J. Lynham, and Á. de Paula (2018). Vog: Using volcanic eruptions to estimate the health costs of particulates. *The Economic Journal 129*(620), 1782–1816.

Hanna, R. and P. Oliva (2015). The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City. *Journal of Public Economics 122*(C), 68–79.

He, J., S. Gong, Y. Yu, L. Yu, L. Wu, H. Mao, C. Song, S. Zhao, H. Liu, X. Li, et al. (2017). Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major Chinese cities. *Environmental Pollution 223*, 484–496.

Hsiang, S., P. Oliva, and R. Walker (2019). The distribution of environmental damages. *Review of Environmental Economics and Policy 13*(1), 83–103.

Isen, A., M. Rossin-Slater, and W. R. Walker (2017). Every breath you take - every dollar you'll make: The long-term consequences of the Clean Air Act of 1970. *Journal of Political Economy 125*(3), 848–902.

ISTAT (2001). 14 Censimento della popolazione e delle abitazioni 2001. Technical report, Istituto Nazionale di Statistica.

Janssen, B. G., E. Munters, N. Pieters, K. Smeets, B. Cox, A. Cuypers, F. Fierens, J. Penders, J. Vangronsveld, W. Gyselaers, et al. (2012). Placental mitochondrial DNA content and particulate air pollution during in utero life. *Environmental health perspectives 120*(9), 1346–1352.

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*(2), 350–366.

Laurent, É. (2011). Issues in environmental justice within the European Union. *Ecological Economics 70*(11), 1846–1853.

Lavaine, E. and M. Neidell (2017). Energy production and health externalities: Evidence from oil refinery strikes in France. *Journal of the Association of Environmental and Resource Economists 4*(2), 447–477.

Li, F., T. Wu, X. Lei, H. Zhang, M. Mao, and J. Zhang (2013, 07). The Apgar score and infant mortality. *PLOS ONE 8*(7), 1–8.

Lichter, A., N. Pestel, and E. Sommer (2017). Productivity effects of air pollution: Evidence from professional soccer. *Labour Economics 48*, 54–66.

Neidell, M. J. (2004). Air pollution, health, and socio-economic status: The effect of outdoor air quality on childhood asthma. *Journal of Health Economics 23*(6), 1209–1236.

Ouyang, W., B. Guo, G. Cai, Q. Li, S. Han, B. Liu, and X. Liu (2015). The washing effect of precipitation on particulate matter and the pollution dynamics of rainwater in downtown Beijing. *Science of the Total Environment 505*, 306–314.

Sager, L. (2016). Estimating the effect of air pollution on road safety using atmospheric temperature inversions. *Grantham Research Institute on Climate Change and the Environment, Working Paper 251*, 2–30.

Schlenker, W. and W. R. Walker (2015). Airports, air pollution, and contemporaneous health. *The Review of Economic Studies 83*(2), 768–809.

Simeonova, E., J. Currie, P. Nilsson, and R. Walker (2018). Congestion pricing, air pollution and children's health. Working Paper 24410, National Bureau of Economic Research.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

van den Hooven, E. H., F. H. Pierik, Y. de Kluizenaar, A. Hofman, S. W. van Ratingen, P. Y. Zandveld, H. Russcher, J. Lindemans, H. M. Miedema, E. A. Steegers, et al. (2012). Air pollution exposure and markers of placental growth and function: The generation R study. *Environmental Health Perspectives 120*(12), 1753.

Wang, Y., X. Zhang, J. Sun, X. Zhang, H. Che, and Y. Li (2015). Spatial and temporal variations of the concentrations of PM10, PM2.5 and PM1 in China. *Atmospheric Chemistry and Physics 15*(23), 13585–13598.

Warren, J. L., J.-Y. Son, G. Pereira, B. P. Leaderer, and M. L. Bell (2017). Investigating the impact of maternal residential mobility on identifying critical windows of susceptibility to ambient air pollution during pregnancy. *American Journal of Epidemiology 187*(5), 992–1000.

WHO (2006). Multicentre growth reference study group. WHO child growth standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Technical report, World Health Organization. Geneva.

WHO (2013). Review of evidence on health aspects of air pollution - REVIHAAP project. Technical report, WHO. Copenhagen, Denmark.

Whyatt, R. M. and F. P. Perera (1995). Application of biologic markers to studies of environmental risks in children and the developing fetus. *Environmental Health Perspectives 103*(Suppl 6), 105.

Xu, M., Y. Guo, Y. Zhang, D. Westerdahl, Y. Mo, F. Liang, and X. Pan (2014). Spatiotemporal analysis of particulate air pollution and ischemic heart disease mortality in Beijing, China. *Environmental Health 13*(1), 109.

Yang, M., R. A. Bhatta, S.-Y. Chou, and C.-I. Hsieh (2017). The impact of prenatal exposure to power plant emissions on birth weight: Evidence from a Pennsylvania power plant located upwind of New Jersey. *Journal of Policy Analysis and Management 36*(3), 557–583.

Yang, M. and S.-Y. Chou (2018). The impact of environmental regulation on fetal health: Evidence from the shutdown of a coal-fired power plant located upwind of New Jersey. *Journal of Environmental Economics and Management 90*, 269–293.

Yoo, J.-M., Y.-R. Lee, D. Kim, M.-J. Jeong, W. R. Stockwell, P. K. Kundu, S.-M. Oh, D.-B. Shin, and S.-J. Lee (2014). New indices for wet scavenging of air pollutants (O3, CO, NO2, SO2, and PM10) by summertime rain. *Atmospheric Environment 82*, 226–237.

Appendix 2.A

Table 2.A1: Summary Statistics

	Mean	SD
A. Outcomes		
Birth Weight (grams)	3,272.12	247.97
Low Birth Weight (LBW)	0.05	0.11
Very Low Birth Weight (VLBW)	0.01	0.04
Intra-Uterine Growth Restriction (IUGR)	0.02	0.08
Gestation (days)	273.27	6.09
Preterm Birth	0.05	0.11
Low Apgar score (Low APGAR)	0.31	1.73
B. Pollution Measures during Pregnancy		
Mean PM_{10} exposure (mcg/m3)	34.73	10.22
# Days with PM_{10} above the EU limit	54.77	37.13
C. Control Variables		
Age of mother	32.28	2.54
Female child	0.49	0.25
Foreign mother	0.19	0.21
Mother Education: less than high school	0.30	0.25
Mother Education: high school	0.44	0.25
Mother Education: more than high school	0.26	0.22
Housewife	0.34	0.25
Mother Dependent employee	0.56	0.26
Mother Self-employed	0.10	0.15
Mother Employed	0.68	0.24
Mother Married	0.73	0.25
Previous births	0.45	0.25
Previous abortions	0.20	0.20
Type of hospital: public	0.91	0.19
Type of hospital: private	0.07	0.17
Type of hospital: missing	0.01	0.08
Pediatrician: absent	0.31	0.31
Pediatrician: present	0.58	0.34
Pediatrician: missing	0.11	0.28
Municipal income (ave. gross per capita)	23,731.49	3,156.90
D. Environmental Variables during Pregnancy		
Mean of daily minimum temperature	9.18	3.40
Mean of daily maximum temperature	18.80	3.67
Total rain (mm)	598.61	224.63

Notes: Baseline Sample: N=12,260. The unit of observation is a municipality-week-of-birth cell made of mothers who live in the same municipality and give birth in the same week of the year. Data on births come from the Standard Certificates of Live Births. Air pollution data are from the European Air Quality Database (Airbase). Weather information is from the Gridded Agro-Meteorological Data-CGMS.

Table 2.A2: Samples Comparison

	Estimation Sample		Unrestricted Sample		15km-radius Sample	
	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Birth Weight	3272.12	247.97	3270.51	409.06	3273.40	176.53
LBW	0.05	0.11	0.05	0.18	0.05	0.08
VLBW	0.01	0.04	0.01	0.06	0.01	0.03
IUGR	0.02	0.08	0.02	0.13	0.02	0.05
Gestation Days	273.27	6.09	273.04	9.93	273.35	4.38
Preterm Birth	0.05	0.11	0.06	0.19	0.06	0.08
Low APGAR	0.03	0.17	0.06	0.19	0.02	0.12
Age of Mother	32.28	2.54	31.49	4.28	32.00	1.90
Female Newborn	0.49	0.25	0.49	0.41	0.48	0.18
Mother Foreign	0.19	0.21	0.14	0.29	0.14	0.15
Mother Education (high school)	0.44	0.25	0.46	0.42	0.48	0.19
Mother Education (> high school)	0.26	0.22	0.15	0.29	0.19	0.15
Mother Dependent employee	0.56	0.26	0.50	0.42	0.58	0.21
Mother's Self-employed	0.10	0.15	0.09	0.24	0.09	0.10
Mother Employed	0.68	0.24	0.60	0.42	0.68	0.20
Mother Married	0.73	0.25	0.76	0.37	0.76	0.18
Previous Births	0.45	0.25	0.47	0.41	0.46	0.18
Previous Abortions	0.20	0.20	0.17	0.32	0.19	0.14
Type of Hospital (private)	0.07	0.17	0.06	0.21	0.07	0.15
Pediatrician (present)	0.58	0.34	0.58	0.44	0.58	0.28
Number of Observations	12,260		860,473		13,143	

Notes: Each cell is made of mothers living in the same municipality and giving birth in the same week of the year. Estimation Sample is our sample of analysis and consists of birth data matched with environmental data, after sample selection and with no missing values in the main variables. Unrestricted Sample refers to the the birth data before matching with monitoring stations data. 15km-radius Sample includes municipalities whose centroid falls within a radius of 15 km from the monitors' geographical coordinates.

Table 2.A3: OLS Estimates of the Effect of High PM_{10} Days on Neonatal Health

	Birth Weight	LBW	VLBW	IUGR	Gestation Days	Preterm Birth	Low APGAR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A. Whole Pregnancy			
High PM_{10} Days	-2.584***	0.001***	0.000	0.000*	-0.033	0.001***	0.001

	(0.501)	(0.000)	(0.000)	(0.000)	(0.024)	(0.000)	(0.001)
				B. By Trimester			
High PM_{10} Days	-4.479***	0.002**	0.000	0.001**	-0.086*	0.002***	0.003*
during trim. I	(1.031)	(0.001)	(0.000)	(0.000)	(0.051)	(0.001)	(0.002)
High PM_{10} Days	1.214	-0.001	-0.001**	-0.000	0.081*	-0.001**	-0.004*
during trim. II	(1.541)	(0.001)	(0.000)	(0.000)	(0.044)	(0.001)	(0.002)
High PM_{10} Days	-5.676***	0.003***	0.001**	0.001**	-0.122***	0.003***	0.006***
during trim. III	(1.351)	(0.001)	(0.000)	(0.000)	(0.044)	(0.001)	(0.002)
Mean	3,272.12	0.05	0.01	0.02	273.27	0.05	0.31
S.d.	247.97	0.11	0.04	0.08	6.09	0.11	1.73

Notes: The table reports the OLS estimates of the effect of the number of high PM_{10} days (above 50 mcg/m3) during pregnancy (Panel A) and in each trimester of pregnancy (Panel B) on neonatal health. Pollution coefficients show the effect of an increase by 10 in days with PM_{10} above EU limit. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. All regressions include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy or in each trimester. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 12,260 observations. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A4: First Stage Estimates of the Effect of Total Rainfall on High PM_{10} Days

	PM_{10}			
	Whole Pregnancy (1)	Trimester I (2)	Trimester II (3)	Trimester III (4)
Rainfall during pregnancy	-0.551***			
	(0.162)			
Rainfall during trim. I		-0.370***	-0.092	-0.013
		(0.098)	(0.067)	(0.005)
Rainfall during trim. II		-0.073	-0.377***	-0.128**
		(0.077)	(0.085)	(0.054)
Rainfall during trim. III		0.008	0.081	-0.303***
		(0.080)	(0.071)	(0.085)
F-statistics	11.6	26.90	26.10	14.45

Notes: The table reports the first stage estimates of the effect of total rainfall on the number of high PM_{10} days (above 50 mcg/m3) during pregnancy and in each trimester. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. All regressions include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy or in each trimester. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 12,260 observations. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A5: IV Estimates of the Effect of High PM_{10} Days on Neonatal Health

	Birth Weight	LBW	VLBW	IUGR	Gestation Days	Preterm Birth	Low APGAR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A. Whole Pregnancy			
High PM_{10} Days	-4.885***	0.002***	0.001**	0.001	-0.159**	0.003***	0.004*
	(1.370)	(0.001)	(0.000)	(0.001)	(0.060)	(0.001)	(0.002)
				B. By Trimester			
High PM_{10} Days	-3.142	0.001	-0.000	0.001	-0.087	-0.001	-0.010
during trim. I	(3.239)	(0.002)	(0.001)	(0.001)	(0.115)	(0.001)	(0.010)
High PM_{10} Days	6.595	-0.006	-0.001	-0.000	0.142	-0.001	-0.008
during trim. II	(8.200)	(0.004)	(0.001)	(0.001)	(0.234)	(0.005)	(0.013)
High PM_{10} Days	-21.084***	0.013***	0.002**	0.002	-0.653***	0.012**	0.032***
during trim. III	(6.857)	(0.003)	(0.001)	(0.002)	(0.169)	(0.005)	(0.011)
Mean	3,272.12	0.05	0.01	0.02	273.27	0.05	0.31
S.d.	247.97	0.11	0.04	0.08	6.09	0.11	1.73

Notes: The table reports the IV estimates of the effect of the number of high PM_{10} days (above 50 mcg/m3) during pregnancy (Panel A) and in each trimester of pregnancy (Panel B) on neonatal health using total rain over pregnancy and total rain in each trimester as an instrument, respectively. Pollution coefficients show the effect of an increase by 10 in days with PM_{10} above EU limit. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. All regressions include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income and average minimum and maximum temperatures during pregnancy or in each trimester. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 12,260 observations. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A6: IV Estimates of the Effect of High PM_{10} Days on Neonatal Health by Employment Status

	Birth Weight (1)	LBW (2)	VLBW (3)	IUGR (4)	Gestation Days (5)	Preterm Birth (6)	Low APGAR (7)
				A. Employed Mothers			
High PM_{10} Days	-2.278*	0.002**	0.000	0.001	-0.138**	0.002***	-0.000
	(1.290)	(0.001)	(0.000)	(0.001)	(0.055)	(0.001)	(0.002)
Mean	3,272.61	0.04	0.006	0.02	273.31	0.05	0.02
S.d.	266.87	0.11	0.04	0.087	6.53	0.11	0.16
				B. Unemployed Mothers			
High PM_{10} Days	-8.325**	0.002*	0.001	0.000	-0.175*	0.003*	0.007*
	(3.632)	(0.001)	(0.000)	(0.001)	(0.093)	(0.002)	(0.004)
Mean	3,280.48	0.05	0.01	0.02	273.17	0.06	0.04
S.d.	319.19	0.14	0.05	0.10	7.65	0.14	0.20

Notes: The table reports the IV estimates of the effect of the number of high PM_{10} days (above 50 mcg/m3) during pregnancy on neonatal health for employed mothers (Panel A) and for unemployed mothers (Panel B) using total rain over pregnancy as an instrument. Pollution coefficients show the effect of an increase by 10 in days with PM_{10} above the EU limit. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. Both panels include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 11,676 observations for employed mothers and 10,100 observations for unemployed mothers. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A7: IV Estimates of the Effect of High PM_{10} Days on Neonatal Health by Education Level

	Birth Weight (1)	LBW (2)	VLBW (3)	IUGR (4)	Gestation Days (5)	Preterm Birth (6)	Low APGAR (7)
				A. More Educated Mothers			
High PM_{10} Days	-2.112	0.001*	0.000	0.001*	-0.100**	0.002***	0.003
	(1.479)	(0.001)	(0.000)	(0.001)	(0.049)	(0.001)	(0.003)
Mean	3,282.53	0.05	0.01	0.02	273.49	0.05	0.03
S.d.	259.55	0.11	0.04	0.08	6.30	0.11	0.17
				B. Less Educated Mothers			
High PM_{10} Days	-8.553***	0.003**	0.001*	0.000	-0.167*	0.003*	0.002

	(2.592)	(0.001)	(0.001)	(0.001)	(0.091)	(0.002)	(0.003)
Mean	3,257.95	0.06	0.01	0.02	272.78	0.06	0.05
S.d.	333.25	0.15	0.05	0.10	8.04	0.16	0.21

Notes: This table reports the IV estimates of the effect of the number of high PM_{10} days (above 50 mcg/m3) during pregnancy on neonatal health for more educated mothers (Panel A) and for less educated mothers (Panel B) using total rain over pregnancy as an instrument. Pollution coefficients show the effect of an increase by 10 in days with PM_{10} above the EU limit. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. Both panels include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 11,601 observations for more educated mothers and 9,963 observations for less educated mothers. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A8: OLS Estimates of the Effect of Weather Conditions on Hospitalizations

	Pneumonia & Influenza (1)	Acute Pulmonary Disease (2)	Mental Disease (3)	Nervous System Disorder (4)
Rain (t)	-0.003	-0.002	0.019	0.053*
	(0.003)	(0.002)	(0.020)	(0.029)
Rain (t-1)	0.001	0.002	0.015	0.046
	(0.003)	(0.002)	(0.020)	(0.029)
Rain (t-2)	0.003	0.001	0.006	0.044
	(0.003)	(0.002)	(0.019)	(0.029)
Max Temp	-0.000	-0.000	0.004	0.003
	(0.000)	(0.000)	(0.004)	(0.005)
Max Temp (t-1)	0.000	0.000	0.000	0.004
	(0.001)	(0.000)	(0.004)	(0.006)
Max Temp (t-2)	-0.000	-0.000	0.003	0.003
	(0.000)	(0.000)	(0.004)	(0.005)
Min Temp	-0.000	-0.000	-0.005	-0.012*
	(0.001)	(0.000)	(0.005)	(0.007)
Min Temp (t-1)	-0.000	0.000	0.002	0.004
	(0.001)	(0.000)	(0.005)	(0.007)
Min Temp (t-2)	-0.000	0.000	-0.003	0.002
	(0.001)	(0.000)	(0.004)	(0.006)

Notes: The table reports the OLS estimates of the effects of rain and temperature in the day of hospitalization and up to 2 days before hospitalization on the number of hospitalizations per 1,000 residents. The estimates are obtained from 4 separate regressions, one per diagnosis, and include municipality fixed effects and day fixed effects. The unit of observation is the municipality-day cell. The estimates are weighted by the number of women in each municipality-year. The coefficients and the standard errors are multiplied by 1,000. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 14,395,843 observations. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A9: IV Estimates of the Effect of Prenatal PM_{10} Exposure on Neonatal Health - Extended Sample

	Birth Weight (1)	LBW (2)	VLBW (3)	IUGR (4)	Gestation Days (5)	Preterm Birth (6)	Low APGAR (7)
				A. Whole Pregnancy			
PM_{10}	-11.818	0.011	0.004	0.010	1.784	-0.001	0.009
	(34.623)	(0.016)	(0.007)	(0.014)	(1.911)	(0.011)	(0.029)
				B. By Trimester			
PM_{10}, trim. I	-13.442	0.006	0.000	0.007	1.084	-0.003	-0.012
	(27.134)	(0.013)	(0.005)	(0.011)	(1.456)	(0.009)	(0.014)
PM_{10}, trim. II	-1.531	-0.001	0.004	-0.003	0.521	-0.006	0.019
	(16.251)	(0.010)	(0.003)	(0.008)	(0.739)	(0.010)	(0.023)
PM_{10}, trim. III	2.180	0.011	-0.002	0.014	0.959	0.009	0.000
	(26.828)	(0.013)	(0.004)	(0.013)	(1.441)	(0.009)	(0.023)
Mean	3,273.37	0.05	0.01	0.02	273.28	0.05	0.01
S.d.	403.92	0.18	0.07	0.12	9.73	0.18	0.12

Notes: The table reports the IV estimates of the effect of average PM_{10} exposure during pregnancy (Panel A) and in each trimester of pregnancy (Panel B) on neonatal health using total rain over pregnancy and total rain in each trimester as an instrument, respectively. Pollution coefficients show the effect of an increase by 10 in the average PM_{10}. The unit of observation is the municipality-week-of-birth cell, where the definition of municipality is extended to inclusion of municipalities whose centroid falls within a radius of 15 km from the monitor's geographical coordinates. The estimates are weighted by the number of births in each municipality. All regressions include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy or in each trimester. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 13,143 observations. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A10: IV Estimates of the Effect of High PM_{10} Days on Neonatal Health - Extended Sample

	Birth Weight	LBW	VLBW	IUGR	Gestation Days	Preterm Birth	Low APGAR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A. Whole Pregnancy			
High PM_{10} Days	-2.727	0.002	0.001	0.002	0.412	-0.000	0.002
	(7.369)	(0.003)	(0.001)	(0.003)	(0.352)	(0.003)	(0.007)
				B. By Trimester			
High PM_{10} Days	-8.627	0.003	0.001	0.003	0.625	-0.003	-0.007
during trim. I	(14.811)	(0.007)	(0.003)	(0.006)	(0.667)	(0.005)	(0.008)
High PM_{10} Days	-0.772	-0.001	0.003	-0.001	0.430	-0.004	0.015
during trim. II	(12.772)	(0.008)	(0.003)	(0.006)	(0.450)	(0.009)	(0.018)
High PM_{10} Days	3.689	0.009	-0.003	0.012	0.635	0.009	-0.001
during trim. III	(19.764)	(0.010)	(0.003)	(0.009)	(0.784)	(0.009)	(0.020)
Mean	3,273.37	0.05	0.01	0.02	273.28	0.05	0.01
S.d.	403.92	0.18	0.07	0.12	9.73	0.18	0.12

Notes: The table reports the IV estimates of the effect of the number of high PM_{10} days (above 50 mcg/m3) during pregnancy (Panel A) and in each trimester of pregnancy (Panel B) on neonatal health using total rain over pregnancy and in each trimester as an instrument, respectively. Pollution coefficients show the effect of an increase by 10 in days with PM_{10} above EU limit. The unit of observation is the municipality-week-of-birth cell, where the definition of municipality is extended to inclusion of municipalities whose centroid falls within a radius of 15 km from the monitor's geographical coordinates. The estimates are weighted by the number of births in each municipality. All regressions include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy or in each trimester. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 13,143 observations. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A11: IV Estimates of the Effect of Prenatal PM_{10} Exposure on Neonatal Health - Mothers with Region of Residence Different from Region of Hospital

	Birth Weight	LBW	VLBW	IUGR	Gestation Days	Preterm Birth	Low APGAR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A. Whole Pregnancy			
PM_{10}	-25.004**	0.014**	0.005*	0.004*	-0.687**	0.011*	0.014
	(11.320)	(0.005)	(0.003)	(0.002)	(0.268)	(0.005)	(0.009)
				B. By Trimester			

PM_{10}, trim. I	-2.042	0.003	-0.002	0.002	-0.125	-0.002	-0.007
	(13.964)	(0.007)	(0.003)	(0.002)	(0.281)	(0.006)	(0.017)
PM_{10}, trim. II	12.094	-0.013**	0.001	-0.001	0.406*	-0.008	-0.010
	(9.894)	(0.005)	(0.002)	(0.001)	(0.242)	(0.006)	(0.013)
PM_{10}, trim. III	-31.253***	0.025***	0.003	0.003*	-0.941***	0.019**	0.030***
	(11.355)	(0.006)	(0.003)	(0.002)	(0.307)	(0.008)	(0.011)
Mean	3,277.93	0.04	0.005	0.02	273.49	0.05	0.03
S.d.	245.46	0.10	0.038	0.07	5.93	0.10	0.18

Notes: The table reports the IV estimates of the effect of average PM_{10} exposure during pregnancy (Panel A) and in each trimester of pregnancy (Panel B) on neonatal health using total rain over pregnancy and total rain in each trimester as an instrument, respectively. The estimation sample includes mothers declaring region of residence different from region of hospital. Pollution coefficients show the effect of an increase by 10 in the average PM_{10} expressed in mcg/m3. The unit of observation is the municipality-week-of-birth cell. The estimates are weighted by the number of births in each municipality. All regressions include municipality fixed effects and week-of-birth fixed effects. All controls for maternal and child characteristics are listed in Table 1. Controls also include yearly municipal income as well as average minimum and maximum temperatures during pregnancy or in each trimester. Robust standard errors, clustered at the municipality level, are shown in parentheses. Sample size is 11,886 observations. * significant at 10%; ** significant at 5%; *** significant at 1%.

Authors' Contributions

Prenatal Air Pollution Exposure and Neonatal Health

Alessandro Palma, Inna Petrunyk, Daniela Vuri

Authors' contribution:

Joint contribution:

(i) conception and design of the work
(ii) choice of empirical methods
(iii) drafting the work

Alessandro Palma's contribution (30%):

(i) environmental data preparation
(ii) econometric analysis
(iii) results preparation

Inna Petrunyk's contribution (40%):

(i) literature review
(ii) health data set preparation
(iii) interpretation of statistical results

Daniela Vuri's contribution (30%):

(i) birth data access and preparation
(ii) coordination of work
(iii) wrap up of the paper

February, 17 2020

Inna Petrunyk | Alessandro Palma | Daniela Vuri

Chapter 3

Life Satisfaction in Germany after Reunification: Additional Insights on the Pattern of Convergence

Joint with: Christian Pfeifer. Authors' Contributions.

Published in: Journal of Economics and Statistics, 2016; 236 (2): 217-239.
doi:10.15456/jbnst.2016119.084659.

Data replication statement:
A scientific use file of Socio-Economic Panel (SOEP), data for years 1984-2013, was made available by the German Institute for Economic Research. The used version is 30, SOEP, 2014, doi:10.5684/soep.v30. Unfortunately, the data are confidential and can not be provided as this would represent a violation of the confidentiality agreement. Statistical programs with all steps of data preparation and cleaning are available and, in case of request, support in accessing the data to anybody interested in getting the data from the German Institute for Economic Research will be provided.

3.1 Introduction

Twenty-five years after the German reunification in the year 1990, convergence between the Eastern (new) and Western (old) German federal states has not been achieved completely. For example, employment levels, wages, and productivity are still significantly lower in East than in West Germany (Burda, 2013; Smolny, 2009, 2012; Smolny and Kirbach, 2011; Uhlig, 2008, among others). In addition to such standard economic variables, researchers have also looked at the East-West gap in overall life satisfaction, happiness or well-being, respectively.[1] The reported persistent lower levels of life satisfaction for people living in East Germany give rise to the question if equal living conditions can be achieved. We add to this stream of the literature by analyzing data from the German Socio-Economic Panel (hereafter, SOEP) for the years 1992 to 2013 in order to update previous research and to get additional new insights. At first, we use pooled and individual specific fixed effects OLS (Ordinary Least Squares) models to estimate the East-West gap in overall life satisfaction. By using different specifications, we explore in how far the raw gap is reduced by the inclusion of important covariates that control for differences in household income, unemployment status, retirement, schooling, age, gender etc. In the next step, we add interaction terms between living in East Germany and the survey years, which allows us to analyze the conditional trends in life satisfaction and in the East-West gap. Additionally, we analyze the East-West gaps and their trends separately for men and women as well as for four birth cohorts.

The remainder of this paper is structured as followed. In the next Section 3.2, we summarize the relevant literature. The data set, variables, and econometric approach are described in Section 3.3. In Section 3.4, we present and discuss the results from our regression analyses. The paper concludes with a short summary and discussion in Section 3.5.

3.2 Literature Review

In the growing literature on subjective well-being, more commonly identified with life satisfaction or happiness, researchers have paid particular attention to its determinants, investigating factors able to enhance social welfare (Gerlach and Stephan, 1996; Winkelmann and Winkelmann, 1998). The collapse of socialism in East Germany and the concomitant German reunification in 1990 have encouraged economists for empirical studies in this field of literature (Frijters et al., 2004a,b; Van Hoorn and Maseland, 2010). Thereby, the variables of main interest refer to life satisfaction, used as a well-established proxy measure for utility and economic conditions in Germany following reunification (Easterlin and Plagnol, 2008). In particular, more recent studies have addressed the issue of both the determinants

[1] In our empirical analysis, we focus on life satisfaction. Whereas happiness relates more strongly to emotional states, life satisfaction aims to measure more strongly permanent states and is a better proxy for the economic utility concept.

of life satisfaction in East and West Germany and its convergence over the course of time, focusing on implications for policy.

Frijters et al. (2004a) using the SOEP consider the years 1985 to 1999 for West Germans and 1991 to 1999 for East Germans. Contrary to the authors' expectation of respondents living on the border between East and West German states being affected by reunification to the largest extent, the data point to no significant difference with respect to the residence place within the two previously separated parts of the country in terms of life satisfaction immediately after reunification. However, movers from the East to the West of Germany following reunification reveal a higher life satisfaction level with respect to the stayers in the East. Despite the fact, that the average life satisfaction for East Germans after 1990 is characterized by a steady increase, there is evidence of a significantly lower level of life satisfaction in the East compared to the West in the studied period. On the contrary, West Germans experienced a slight downward trend in life satisfaction in the years directly after reunification, but on average little change can be registered over the years (Easterlin and Plagnol, 2008; Noll and Weick, 2010). As a matter of fact, the most remarkable convergence in East-West levels occurred immediately after reunification, followed by a rather stable life satisfaction differential. According to Easterlin and Plagnol (2008), the convergence is concentrated in the years between 1991 and 1997. This evidence is confirmed in the analysis of Noll and Weick (2010), which updates the existing findings and is based on the SOEP from 1990 up to 2007 and the European Social Survey. The authors provide an additional insight into the reaction of East Germans in the crucial 1990, which can be identified with a decline in subjective well-being by 0.6 on a scale of zero to ten. This observation, confirmed also in Frijters et al. (2004a), captures inevitable adjustments and echoes initial difficulties connected with the transition, e.g., from the labor market perspective. In fact, a rapid increase in unemployment among East Germans seems to have largely contributed to this decline. However, this initial discomfort appears to have promptly dissolved. Indeed, life satisfaction among East Germans continuously improved in the subsequent years. Nevertheless, the average life satisfaction in East Germany was significantly below that of the West and has never reached the West German level in the considered period.

The gap in life satisfaction levels between East and West Germany has stimulated economists to investigate the factors this gap can be attributed to. In order to understand why convergence in life satisfaction scores might have occurred only in the first years after reunification, it is necessary to look into the determinants of life satisfaction. In fact, different developments of income and unemployment of people living in different regions could partly explain the evolution of the gaps in unconditional life satisfaction between regions. A recent study of Vatter (2012) has concentrated on the drivers of regional variation in subjective well-being within Germany. According to his analysis based on the SOEP data from 1995 to 2009, about half of the gap is due to discrepant macroeconomic conditions, such as the GDP per capita or unemployment rate, which reflect objective living conditions. The focus lies on the regional heterogeneity within Germany in terms of subjective well-being in nineteen regions, according to which smaller federal states are included in other regions, while states with the largest population are split up. In 2009, the spread in the overall life satisfaction scores is equal to approximately 1.0, whereas the highest score is registered in Hamburg (7.36) and the lowest one in Brandenburg (6.34). The estimation results of reported life satisfaction at the regional level suggest that the existing regional differences can be to a greater extent explained by unemployment, measured as ratio of unemployed persons to the total civil workforce, rather than by aggregated income in terms of real GDP per capita.

Controlling for individual characteristics life satisfaction is largely explained by income and unemployment also in the studies of Winkelmann and Winkelmann (1998) and Frijters et al. (2004b). According to the results of Noll and Weick (2010), higher levels of household income strongly increase life satisfaction, whereas being unemployed significantly reduces it. In the empirical literature consensus on the negative causal relationship between own unemployment and life satisfaction has been reached. In fact, the detrimental impact of unemployment on subjective well-being is confirmed across time and data sets and is often referred to as stylized fact (Clark et al., 2010; Oswald, 1997). Being unemployed, nevertheless, has a different impact with respect to gender and age (Gerlach and Stephan, 1996). Indeed, the cost of unemployment in terms of reduced life satisfaction is higher for men than women and largest for younger workers (Winkelmann and Winkelmann, 1998).

Empirical studies do not find similar clear-cut impact of income on life satisfaction. In fact, no unambiguous results have so far been found concerning income. (Frijters et al., 2004b) focus their attention on East Germany. Using the SOEP data for the period 1991 to 2001 the authors conclude, that significant improvements in life satisfaction can to a great extent be explained by higher real household monthly incomes in the East, whereby the largest effects can be observed in the years immediately after reunification. Moreover, better circumstances proxied by year controls significantly contribute to higher satisfaction levels of Eastern inhabitants directly after the transition. Till the late 1990s life satisfaction in East Germany improved, notwithstanding the unemployment rate increased. Easterlin and Plagnol (2008) attribute this correlation to a remarkable increase in household income in East Germany due to substantial public transfers from West Germany. According to Oswald (1997) additional income is a less important factor in explaining life satisfaction. However, more recent empirical studies in this field emphasize that higher income levels may still increase life satisfaction scores significantly, once numerous materialistic demands have been accomplished already (Clark et al., 2008). In the face of evidence supportive of the explanatory power of income, economists have no unanimous view on whether income is a good predictor of subjective well-being, and if so, which income measure should be preferred. According to Ferrer-i Carbonell (2005) household income rather than personal income better accounts for individuals' real access to economic resources. Oswald (1997) contribute to this debate from another perspective. They review the effect on life satisfaction of both relative and absolute income measures. In the context of the German reunification the authors examine annual means of life satisfaction and economic conditions in East and West Germany using the SOEP data up to 2004. They conclude that relative income, defined as the ratio of mean household income for a given group to the mean for Germany as a whole or as a response to a "satisfaction with income" question having addressed the issue of personality traits bias, explains life satisfaction better than absolute income, expressed in terms of adjusted mean household income. Moreover, it seems that income inequality within one country has a negative impact on happiness of lower-income households, attributable to higher unfairness perception and lack of trust among the individuals in the lowest 20 percent income group (Oishi et al., 2011).

Using the SOEP data for the period 1991 to 2007, Van Hoorn and Maseland (2010) analyze the heterogeneity in the structure of life satisfaction in East and West Germany. In fact, variables related to work, income and education have different impacts among East and West Germans. In particular, East Germans value being a blue collar worker, a white collar worker and self-employed relative to being unemployed more positively than the West Germans. This result may be indicative of a stronger preference for working among the East Germans, irrespective of the employment type. With respect to education, any vocational training is more appreciated than a university degree, which reflects a higher consideration of manual labor in the East than in the West. However, higher educated individuals both in the East and West benefited most from the transition (Frijters et al., 2004b). Finally, the authors find evidence of a higher level of materialism among the East than the West Germans. The positive effect of income on the overall life satisfaction is almost 60 percent higher in the East.

With respect to age, the results of Frijters et al. (2004b) suggest that younger males and females experienced larger gains in life satisfaction following reunification than the older ones. Easterlin (2009) finds evidence in favor of this result, affirming that in this context individuals aged fifty and over showed the most adverse attitude towards reunification. Apart from the standard covariates, Noll and Weick (2010) introduce additional socio-economic variables, which have a significant impact on subjective well-being. These refer to the respondents' trust in country's parliament and legal system, and their worries that income in old age will not be adequate to cover later years, which should capture people's confidence in the performance of the pension system. Interestingly, life satisfaction gap between West and East Germany reduces considerably if the above mentioned measures of confidence in welfare state institutions and trust in political and legal systems are included in the analysis. Thereby, age becomes no longer significant.

More recent studies have addressed the issue of whether the reunification process since 1990 has reached its terminal point. Our contribution to the existing literature in this field is twofold. We first analyze the most recent data available (1992-2013) to discover whether the convergence course has succeeded and East-West differentials can nowadays be declared negligible. Moreover, we propose a novel perspective of analyzing the German reunification process by considering birth cohort differences.

3.3 Data and Econometric Approach

We use the German Socio-Economic Panel (SOEP) for the post-reunification period 1992 to 2013 (Frick et al., 2007).[2] The SOEP is a representative longitudinal survey of private households and comprises household as well as individual information such as household composition, labor market participation, income, education, health, and satisfaction levels. We start our analysis with the year 1992, because the survey questionnaires in East and West Germany are uniform for the first time in 1992 and because we want to exclude short-term adjustments in 1990 and 1991 driven by potential "honeymoon effects" directly after reunification from our analysis. Our main focus is on the differences in overall life satisfaction between individuals living in East and West Germany and its trends over the two decades. As we apply regression analyses for 22 years of panel data, our estimation sample excludes all observations with missing values in the used variables that are described subsequently.[3] The final sample consists of n=370,244 observations of N=50,180 individuals with an average panel length of T=7.38 years in an unbalanced panel design. The sample is almost equally split between men and women; for women n=192,288 and N=25,815 and for men n=177,956 and N=24,365.

Our empirical analysis is divided in two parts. First, we analyze in how far the mean raw East-West gap in life satisfaction is reduced by the inclusion of household income, individual employment status, and socio-demographic characteristics that control for differences between individuals living in East and West Germany. Second, we estimate conditional life satisfaction trends for East and West in order to analyze the development of the East-West gap in life satisfaction and to what extent convergence has been achieved. Equation (3.1) illustrates our basic model, in which Greek letters indicate parameters to be estimated:

$$\begin{aligned} LS = {} & \alpha + \beta_1 EAST + \beta_2 YEAR + \beta_3 EAST \times YEAR \\ & + \gamma_1 logHHINCOME + \gamma_2 UNEMPLOYED + \gamma_3 RETIRED \\ & + \gamma_4 EDUCATION + \gamma_5 NONEMPLOYED + \delta X + \epsilon \end{aligned} \tag{3.1}$$

The dependent variable LS measures overall life satisfaction on a 11-point Likert scale ("How satisfied are you at present with your life as a whole?"; 0: completely dissatisfied, 10: completely satisfied). Mean life satisfaction in the estimation sample is 6.98 (SD=1.78). Our main explanatory variable of interest is $EAST$, which is binary and takes the value one if an individual lives in East Germany, which is the case for 25.6 percent of our observations. We further include dummy variables for the survey years ($YEAR$) and interaction terms between living in East Germany and the survey years ($EAST \times YEAR$) for the estimation of trends in East and West Germany. $logHHINCOME$ is the log of

[2] Socio-Economic Panel (SOEP), data for years 1984-2013, version 30, SOEP, 2014, doi:10.5684/soep.v30.

[3] Note that some determinants of life satisfaction have not been considered in our analysis, because the relevant variables are not available for all years.

monthly nominal net household income in Euros, which has a mean of 7.69 log points (SD=0.57). $UNEMPLOYED$ is a dummy variable that takes the value one if an individual is in registered unemployment, which is the case for 6.8 percent of the observations. Note that the unemployment rate in our sample is rather low, as we do not impose an age restriction on our sample, i.e., the sample contains also non employed individuals in retirement or in education. Therefore, we also include different non-employment status variables so that the reference group includes only employed individuals. About 17 percent of the observations are in old age retirement ($RETIRED$), about 3.4 percent are in education ($EDUCATION$), and about 21.8 percent are non employed for other reasons ($NONEMPLOYED$). Furthermore, a set of socio-demographic control variables (X) is included in the regressions: female, kids under 16 in household, number of persons in household, age and age squared, German citizenship, secondary schooling degrees, apprenticeship degree, university degree, and marital status. In order to account for a potential bias stemming from the survey context (Chadi, 2012, among others) the control variables include also a dummy variable for being personally interviewed in the SOEP (in contrast to written or telephone survey), the years since the last SOEP interview, and the years since the first SOEP interview. A list of the variables, their mean values and standard deviations are presented in Table 3.1.

Table 3.1: Summary Statistics

	Mean	SD
A. Main Variables		
LS: overall life satisfaction	6.9774	1.7774
(0: completely dissatisfied, 10: completely satisfied)		
EAST (dummy): living in former East Germany	0.2564	0.4366
logHHINCOME: log of monthly nominal net household income in Euros	7.6931	0.5694
UNEMPLOYED (dummy): registered unemployment status	0.0681	0.2520
RETIRED (dummy): old age retirement	0.1700	0.3756
EDUCATION (dummy): in education	0.0335	0.1800
NONEMPLOYED (dummy): other reasons for non-employment	0.2177	0.4126
B. Control Variables (X)		
Female (dummy)	0.5194	0.4996
Kids <16 years in household (dummy)	0.3091	0.4621
Number of persons in household	2.7605	1.2960
Age in years	47.4944	17.4113
Age squared	2558.87	1745.01
German citizenship (dummy)	0.9102	0.2860
Medium secondary schooling degree ("Realschule") (dummy)	0.2944	0.4558
High secondary schooling degree ("Abitur") (dummy)	0.1834	0.3870
Apprenticeship degree (dummy)	0.5958	0.4907
University degree (dummy)	0.2109	0.4079

Marital status (dummies)		
Married, living together	0.6039	0.4891
Married, separated	0.0165	0.1274
Single	0.2250	0.4176
Divorced	0.0684	0.2525
Widowed	0.0649	0.2463
Personnel SOEP interviewer (dummy)	0.6464	0.4781
Years since last SOEP interview	0.2836	13.417
Years since first SOEP interview	9.7506	7.0135
C. Birth Cohorts (dummies)		
<1945	0.2858	0.4518
1945-1974	0.5632	0.4960
1975-1984	0.1120	0.3153
>1984	0.0391	0.1938

Notes: SOEP 1992-2013. Number of observations is n=370,244. Number of individuals is N=50,180. Average panel length is T=7.38.

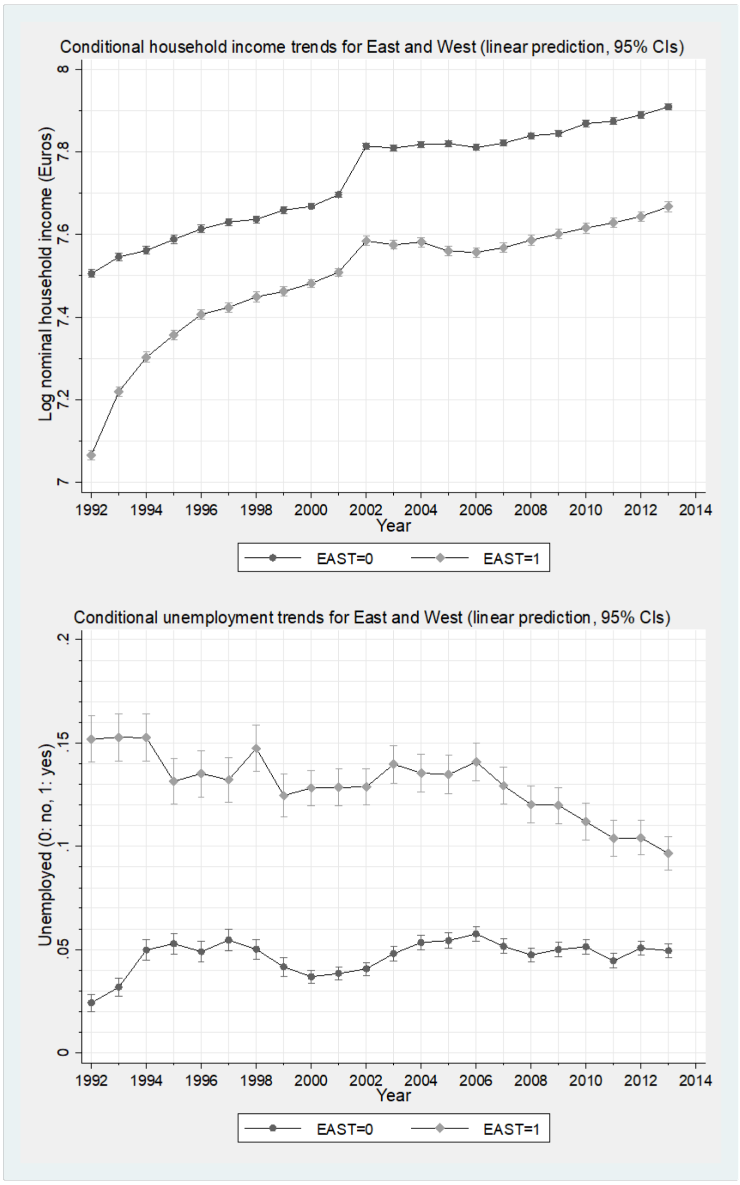
To estimate the life satisfaction equation, we apply OLS regressions. As the life satisfaction variable is measured on an ordinal scale, one might question if linear regressions are appropriate or if ordered Probit or Logit models might not be the better alternatives. In fact, it has been shown in many life satisfaction research papers that the cardinality or ordinality assumption is not very important (Ferrer-i Carbonell and Frijters, 2004). Of larger concern is unobserved heterogeneity that might bias the results. In our application we are interested in regional differences so that identification in an individual specific fixed effects model stems only from the within variation of moving between East and West. Although our sample is quite large, we can only observe 883 individuals who move from East to West or vice versa, which is however seldom the case. Additionally, location choice is not a random assignment, as a rational choice implies that individuals should only move if they gain in utility. Moreover, age and time trends are perfectly collinear in individual fixed effects models. To conclude, estimates of fixed effects models might be inefficient and inconsistent in our application so that we prefer pooled OLS regressions with clustered standard errors at the individual level. Nevertheless, we also present the results from individual specific fixed effects OLS regressions as a robustness check. Also note that we do not claim to establish a causal effect but rather want to describe regional differences between East and West Germany conditional on important individual characteristics such as income, unemployment, family, education, and age.

3.4 Estimation Results

3.4.1 Mean East-West Gap in Overall Life Satisfaction

Before we start our analysis of the East-West gap in life satisfaction and how it is affected by taking into account differences in household income, individual unemployment, and socio-demographic characteristics, let us have a look at the East-West differences in household income and registered unemployment in the estimation sample. For this purpose, we estimate pooled OLS regressions with interaction terms of East and years and further socio-demographic control variables. The estimation results are then used to predict trends of conditional log household income and conditional unemployment for an average individual in East and West in Figure 3.1.

Figure 3.1: Conditional Income and Unemployment Trends, Pooled OLS



It can be seen from the predicted trends in Figure 3.1 that monthly nominal net household income is significantly lower and unemployment significantly larger for individuals living in East than in West Germany. Mean household income in East Germany was about 0.4 log points lower in 1992, about 0.3 log points lower in 1993, about 0.25 log points lower in 1994, and between 0.2 and 0.25 log points lower from 1995 to 2013.[4] Thus, convergence in household income can only be observed in the early years following the reunification and the gap remained quite stable and large since the mid-nineties. The probability to be unemployed is more than twice as large for individuals in our sample who live in East Germany. But the East-West gap in unemployment decreases since 2006, which might be related to the labor market reforms in Germany.

Given the significant differences in household income and registered unemployment, the question arises in how far the mean raw East-West gap in life satisfaction is reduced by the inclusion of household income, individual unemployment, and socio-demographic characteristics that control for differences between individuals living in East and West Germany. Table 3.2 contains five different specifications of our life satisfaction equation, which all comprise $EAST$ and $YEAR$ but not their interaction terms. The first specification does not take into account any other covariates and the estimated coefficient of -0.6262 can be interpreted as the mean raw East-West gap in life satisfaction, which indicates that overall life satisfaction is about 0.63 points lower in East than in West Germany. The second specification includes $logHHINCOME$, which has the expected significant positive impact on life satisfaction and reduces the East-West gap to -0.4684. The third specification includes $UNEMPLOYED$, $RETIRED$, $EDUCATION$, and $NONEMPLOYED$ (reference group are employed individuals), which reduces the East-West gap to -0.4282. $UNEMPLOYED$ has the expected significant negative impact on life satisfaction, whereas $RETIRED$ and $EDUCATION$ seem to be positively correlated with life satisfaction. The fourth specification takes into account further control variables (X) as listed in Table 3.1. The East-West gap slightly increases to -0.4314. In the fifth specification, we add individual specific fixed effects, which reduces the East-West gap to -0.2235. But as discussed in the previous section, the fixed effects model serves only as a robustness check which supports the tendency of lower life satisfaction in East than in West Germany. We have also re-estimated the regressions separately for men and women, which did however not reveal noteworthy gender specific differences. To sum up, the regressions indicate that only part of the raw East-West gap in life satisfaction is due to differences in household income, unemployment status and other characteristics and that a significant conditional life satisfaction gap remains. The trends of this conditional life satisfaction gap are analyzed in more detail in the next section.

[4]Note that the jump in household income for East and West in the year 2002 is due to the inclusion of an additional sample with high income (>4500 Euros monthly net income) in the SOEP. The East-West gap in household income and also the subsequent results for life satisfaction remain however unaffected by the inclusion of this additional sample.

Table 3.2: East-West Gap in Life Satisfaction

	(1)	(2)	(3)	(4)	
EAST	-0.6262***	-0.4684***	-0.4282***	-0.4314***	-0.2235***
	[0.0168]	[0.0163]	[0.0160]	[0.0165]	[0.0456]
logHHINCOME		0.7064***	0.6270***	0.6381***	0.3505***
		[0.0113]	[0.0114]	[0.0130]	[0.0119]
UNEMPLOYED			-0.8232***	-0.7884***	-0.5762***
			[0.0226]	[0.0226]	[0.0175]
RETIRED			0.0586**	0.0403	0.0360
			[0.0183]	[0.0257]	[0.0200]
EDUCATION			0.3704***	0.1597***	0.0848***
			[0.0199]	[0.0224]	[0.0213]
NONEMPLOYED			-0.0869***	-0.0598***	-0.0151
			[0.0150]	[0.0158]	[0.0119]
YEAR	Yes	Yes	Yes	Yes	Yes
Controls X	No	No	No	Yes	Yes
Individual fixed effects	No	No	No	No	Yes
R squared	0.0268	0.0740	0.0919	0.1126	0.0353

Notes: SOEP 1992-2013. Number of observations is n=370,244. Number of individuals is N=50,180. Average panel length is T=7.38. OLS regressions. SE clustered at individual level in brackets. Significant at * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

3.4.2 Trends in Conditional Life Satisfaction and East-West Gap

At first, the unconditional life satisfaction trends in East and West are estimated in order to obtain a reference. For this purpose, the life satisfaction equation is estimated only with $EAST$, $YEAR$ and additional interaction terms of living in East Germany and the survey years ($EAST \times YEAR$). In the next step, all control variables are included in the regression in order to study the trends in conditional life satisfaction and in the East-West gap. The results for the East and West trends are presented in Table 3.3. The first and second specifications are estimated using pooled OLS and the third specification includes additionally individual fixed effects. Due to the collinearity problem between age and the time trend, the dummy variable for the year 2013 has been dropped from the third specification. Thus, the fourth specification includes individual fixed effects and excludes the age variables. When comparing the results across the specifications, it can easily be seen that the results for the year dummies, i.e., for the general time trend, differ strongly. But the results for the East-West gaps, i.e., for $EAST$ and $EAST \times YEAR$, are comparable in size. The comparison between the third and the fourth specification shows only small differences in the third digit of the coefficients so that the collinearity problem in individual fixed effects models does not seem to affect the estimated East-West gaps. The comparison with the pooled OLS results in the second specification and the individual fixed effects models indicates larger gaps of approximately 0.2 life satisfaction points for pooled OLS in every year, which is comparable in size with the findings in the previous section (see Table 3.2). More important is however that the trends in the gaps are comparable between the specifications so that the fixed effects results support the pooled OLS results. In the subsequent discussion we focus therefore on our preferred results from the specifications using pooled OLS.

Table 3.3: Trend in East-West Gap in Life Satisfaction

	(1)	(2)	(3)	(4)
EAST (dummy)	-1.1436***	-0.9592***	-0.7672***	-0.7712***
EAST $\times$ YEAR (dummies)				
1993	0.1893***	0.1233***	0.1664***	0.1663***
1994	0.3097***	0.1956***	0.2752***	0.2746***
1995	0.5094***	0.3681***	0.4296***	0.4287***
1996	0.5149***	0.3482***	0.4278***	0.4232***
1997	0.5853***	0.4129***	0.5216***	0.5172***
1998	0.5547***	0.4243***	0.5250***	0.5204***
1999	0.5915***	0.4562***	0.5603***	0.5556***
2000	0.5056***	0.4634***	0.5140***	0.5097***
2001	0.5027***	0.4590***	0.5165***	0.5121***
2002	0.4803***	0.4984***	0.5314***	0.5272***
2003	0.4762***	0.4917***	0.5197***	0.5152***
2004	0.4840***	0.4914***	0.5096***	0.5049***
2005	0.5020***	0.5213***	0.5367***	0.5322***
2006	0.4893***	0.5191***	0.5319***	0.5278***
2007	0.5117***	0.5315***	0.5541***	0.5497***
2008	0.6574***	0.6777***	0.7016***	0.6973***
2009	0.6253***	0.6540***	0.6856***	0.6817***
2010	0.7002***	0.7174***	0.7647***	0.7608***
2011	0.7091***	0.7239***	0.7692***	0.7656***
2012	0.6360***	0.6620***	0.7023***	0.6984***
2013	0.7030***	0.7300***	0.7481***	0.7437***
YEAR (dummies)				
1993	-0.0938***	-0.0946***	-0.0859***	-0.0300
1994	-0.1569***	-0.1397***	-0.1578***	-0.0458
1995	-0.1919***	-0.1856***	-0.1740***	-0.0056
1996	-0.1819***	-0.1797***	-0.1395***	0.0817
1997	-0.3089***	-0.2924***	-0.2647***	0.0125
1998	-0.1670***	-0.1732***	-0.1396***	0.1939
1999	-0.1394***	-0.1545***	-0.0960***	0.2933
2000	-0.0286	-0.1786***	-0.0847***	0.3614
2001	0.0057	-0.1445***	-0.0473**	0.4541

2002	-0.0572**	-0.3059***	-0.2038***	0.3533
2003	-0.1358***	-0.3449***	-0.2353***	0.3770
2004	-0.2969***	-0.4889***	-0.3698***	0.2980
2005	-0.1626***	-0.3309***	-0.2117***	0.5117
2006	-0.2047***	-0.3735***	-0.2521***	0.5267
2007	-0.1638***	-0.3304***	-0.1970***	0.6366
2008	-0.1581***	-0.3233***	-0.1765***	0.7118
2009	-0.1664***	-0.3488***	-0.2214***	0.7215
2010	-0.0486*	-0.2255***	-0.0787***	0.9184
2011	-0.1375***	-0.3234***	-0.1510***	0.9002
2012	-0.0292	-0.2543***	-0.0720***	10.332
2013	0.0263	-0.2115***		11.590
logHHINCOME, UNEMPLOYED RETIRED, EDUCATION NONEMPLOYED	No	Yes	Yes	Yes
Controls X incl. age and age squared	No	Yes	Yes	No
Controls X without age and age squared	No	No	No	Yes
Individual fixed effects	No	No	Yes	Yes
R squared	0.0282	0.1137	0.0370	0.0367

Notes: SOEP 1992-2013. Number of observations is n=370,244. Number of individuals is N=50,180. Average panel length is T=7.38. OLS regressions. SE clustered at individual level. Significant at * p<0.05, ** p<0.01, *** p<0.001.

In order to facilitate the interpretation of the regression results, we have used the results from the first and the second specification in Table 3.3 to predict unconditional and conditional life satisfaction trends for an average individual living in East and West Germany in the upper and lower graphs in Figure 3.2. The unconditional trends serve only as reference and it can easily be seen that the conditional East-West gap is substantially smaller than the unconditional East-West gap. The subsequent discussion focuses on the conditional trends. Mean life satisfaction was more than 0.9 points lower in East than in West Germany in the year 1992. Mainly due to a decrease in life satisfaction in West Germany until 1997, the life satisfaction gap decreased to about 0.5 points and remained quite stable until 2007. From 2008 to 2013 the life satisfaction gap is only about 0.3 points. We have also re-estimated the regressions and predicted conditional life satisfaction trends separately for men and women. The trends are depicted in Figure 3.3 and do neither reveal noteworthy gender specific differences in the levels nor in the trends. To sum up, the conditional life satisfaction trends for the complete sample and separately for men and women show significant lower life satisfaction in East than in West Germany. Moreover, convergence has only partly been achieved, namely directly in the years after reunification and again in the last years since 2008. One explanation for the observed lower East-West gap of about 0.3 points in the last years after a stable gap of about 0.5 points between 1997 and 2007 might be a change in the population and consequently in the composition of our sample, which is to some extent representative. As younger birth cohorts have entered the sample in more recent years, they might be responsible for the reduction in the East-West gap in life satisfaction. Therefore, we analyze cohort differences in the East-West gap and in the conditional life satisfaction trends in East and West in the next section.

Figure 3.2: Unconditional and Conditional Life Satisfaction Trends, Based on Specifications (1) and (2) in Table 3.3

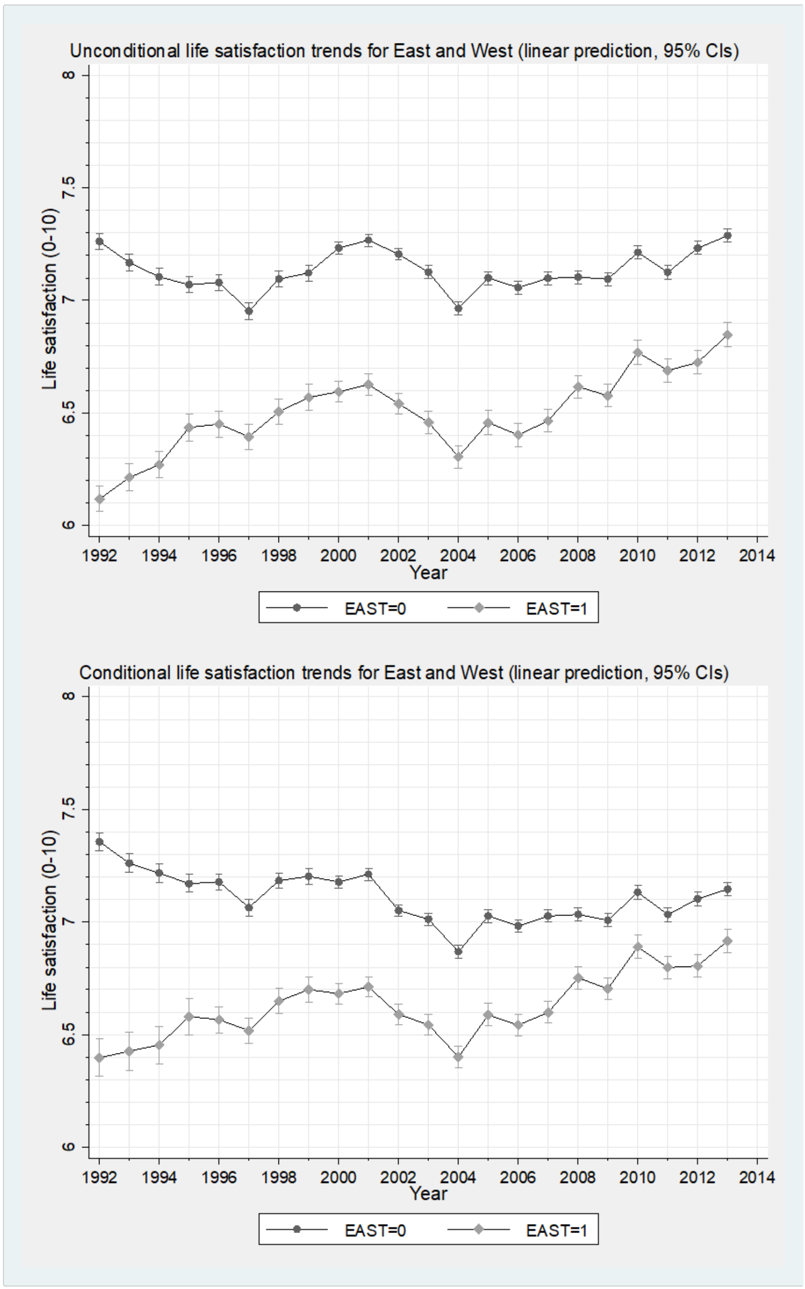
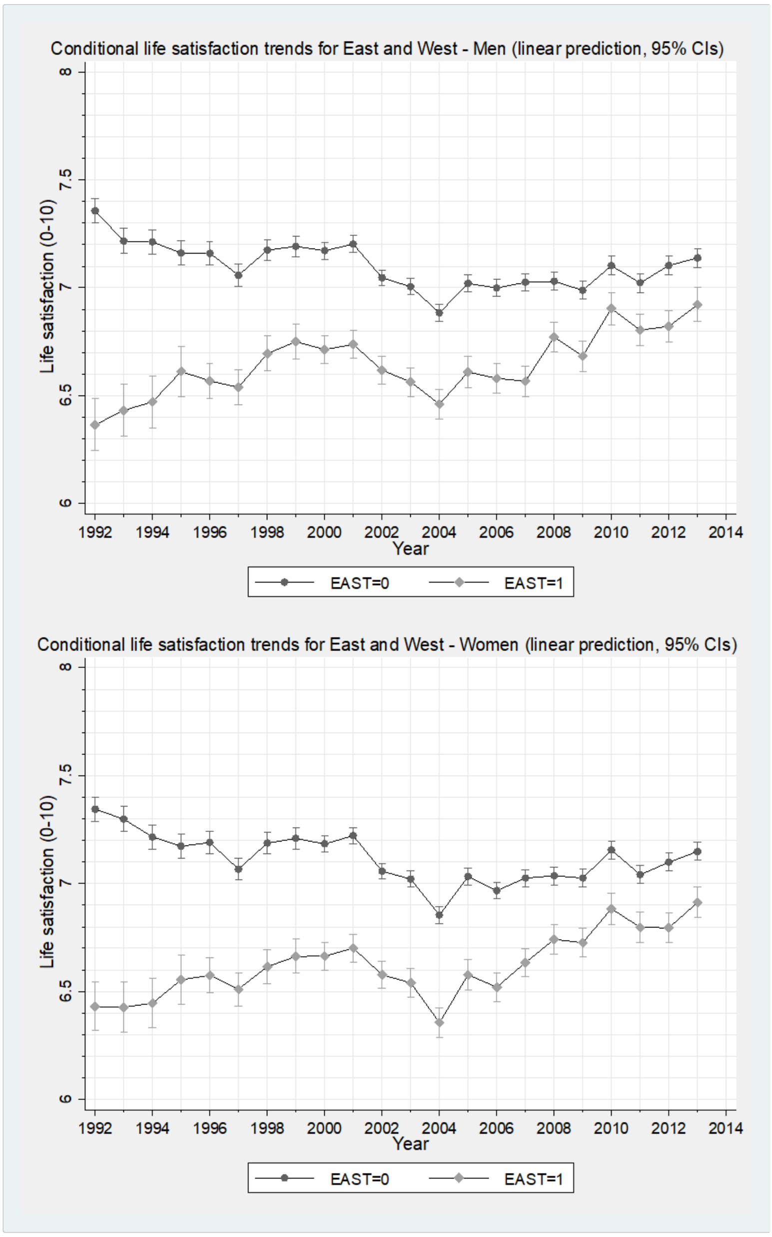


Figure 3.3: Conditional Life Satisfaction Trends for Men and Women



3.4.3 Cohort Differences

A tremendous system change such as the German reunification might affect birth cohorts differently. Even though the reunification was voluntary, at least for most people living in East Germany, many spheres of life might have been adversely affected and hopes might not have been fulfilled. For example, most people were probably surprised by the high level of persistent unemployment after reunification and the depreciation of many labor market skills. These adverse effects are likely to be larger for older individuals at the time of the reunification, who have been longer or even completely socialized in the old system and who have invested in human capital that had more value in the old than in the new system. Younger birth cohorts might however not been so deeply affected by such a system change, as they have not been socialized and did not invest in human capital in the old system.[5] In order to study cohort differences, we define four groups of birth cohorts. We have decided to use the following division in four birth cohorts based on the subsequent rationales:

Cohort 1 comprises individuals born before 1945. These individuals were born before the end of World War II, have been older than 6 years in 1950, when the Federal Republic of Germany (BRD) and the German Democratic Republic (DDR) have been independent separate states for the first time, and have been older than 45 years at time of the German reunification in 1990. Thus, cohort 1 has been socialized before and during World War II as well as in the separated German states. Number of observations in the sample is n=105,810 of N=13,173 individuals.

Cohort 2 comprises individuals born between 1945 and 1974. These individuals have been not older than 6 years in 1950 and older than 16 years in 1990. Thus, cohort 2 has been socialized mainly in the separated German states. More generally, cohort 2 belongs to the post-World War II and to the generation X. Number of observations in the sample is n=208,519 of N=25,770 individuals.

Cohort 3 comprises individuals born between 1975 and 1984. These individuals have been between 6 and 16 years in 1990. Thus, cohort 3 has been partly socialized in the separated German states and partly in reunified Germany. Cohort 3 also belongs to the generation X. Number of observations in the sample is n=41,451 of N=6,926 individuals.

Cohort 4 comprises individuals born after 1984. These individuals have been younger than 6 years in 1990 or even born after 1990. Thus, cohort 4 has been socialized in reunified Germany. Cohort 4 belongs to the generation Y. Number of observations in the sample is n=14,464 of N=4,311 individuals.

At first, we have re-estimated the pooled OLS regression with control variables for the complete sample. But instead of the interaction terms $EAST \times YEAR$, we include interaction terms of the four cohorts and living in East Germany ($EAST \times COHORT$). The results in Table 3.4 indicate a mean conditional East-West gap in life satisfaction of about 0.5 points for cohort 1 (<1945) and cohort 2 (1945-1974). The East-West gap is significantly lower for younger birth cohorts. Whereas the East-West gap for cohort 3 (1975-1984) is still 0.3 points, the gap is only 0.1 points for cohort 4 (>1984).[6]

[5] Van Hoorn and Maseland (2010) apply a robustness check in their analysis of the determinants of life satisfaction by using a subsample of individuals born between1946 and 1971. They argue that these individuals have been socialized in either the East or the West German system so that differences in values and preferences should be most pronounced in this birth cohort.

[6] Note that we do not aim to disentangle cohort from age effects, which is a typical problem when identifying cohort effects. We are interested in differences between the interaction terms of $EAST \times COHORT$ and in the different time trends in the East-West gap ($EAST \times YEAR$) for different cohorts, who have been at different age at the time of the reunification.

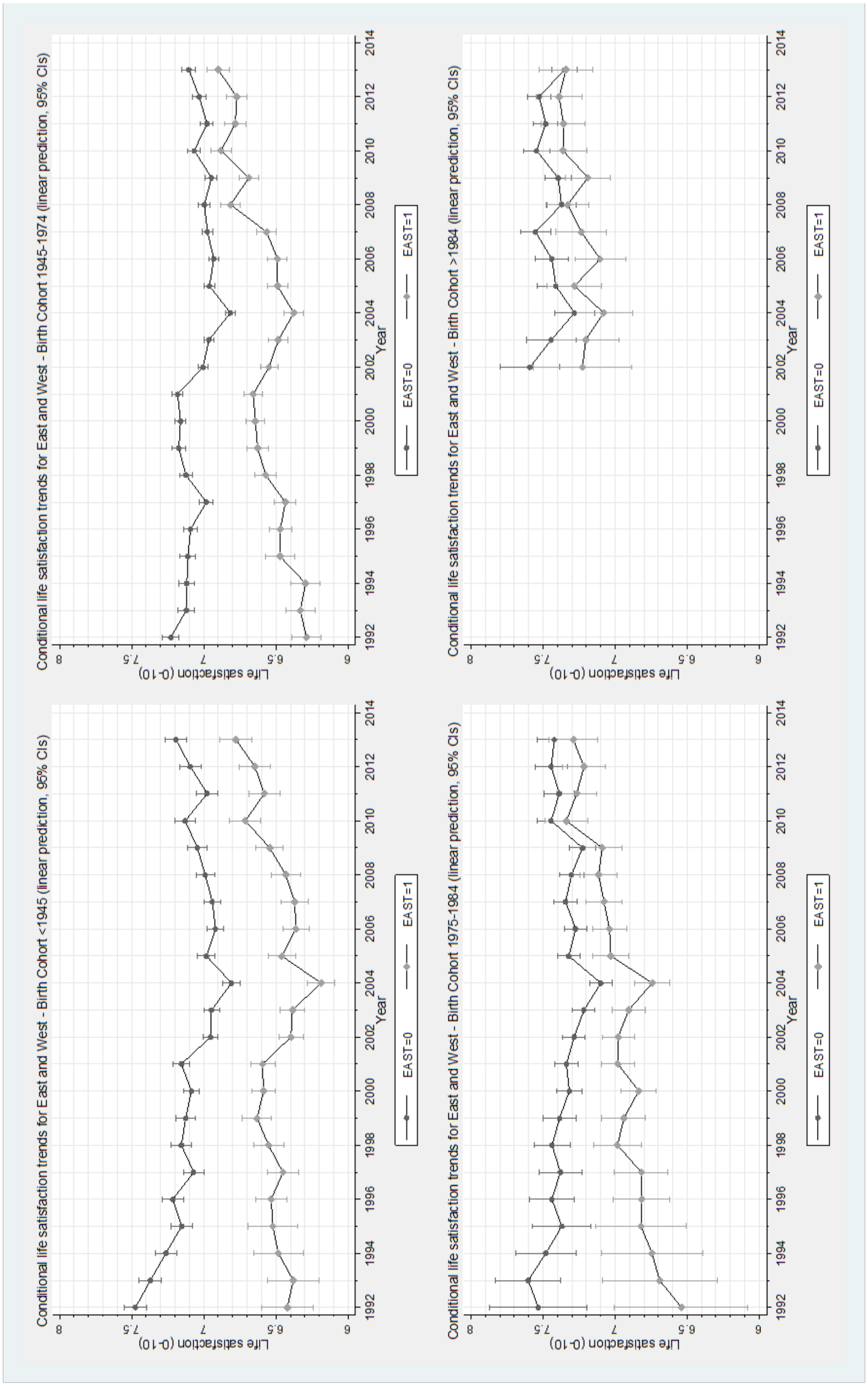
Table 3.4: Cohort Differences in the East-West Gap in Life Satisfaction

	(1)
EAST (dummy)	-0.5033***
	[0.0310]
EAST × COHORT (dummies)	
1945-1974	0.0326
	[0.0376]
1975-1984	0.2054***
	[0.0452]
>1984	0.4213***
	[0.0556]
COHORT (dummies)	
1945-1974	-0.3195***
	[0.0320]
1975-1984	-0.2193***
	[0.0480]
>1984	-0.2272**
	[0.0581]
logHHINCOME, UNEMPLOYED, RETIRED EDUCATION, NONEMPLOYED	Yes
YEAR	Yes
Controls X	Yes
Individual fixed effects	No
R squared	0.1155

Notes: SOEP 1992-2013. Number of observations is n=370,244. Number of individuals is N=50,180. Average panel length is T=7.38. OLS regressions. SE clustered at individual level in brackets. Significant at * p<0.05, ** p<0.01, *** p<0.001.

In the next step, we have re-estimated the pooled OLS regression with control variables and the interaction terms $EAST \times YEAR$ for each of the four groups of birth cohorts separately. Based on the regression results we predict the conditional life satisfaction trends for East and West separately for each cohort in Figure 3.4. It can easily be seen from the graphs that the conditional East-West gap in life satisfaction is smaller and that convergence is much stronger for younger birth cohorts. The East-West gap remains quite stable and sizeable at around 0.5 points since the mid-nineties for cohort 1 (<1945). For cohort 2 (1945-1974) the East-West gap decreases from more than 0.5 points in the mid-nineties to about 0.2 points in recent years. For cohort 3 (1975-1984) the East-West gap decreases even more strongly from more than 0.5 points in the mid-nineties to about 0.1 to 0.2 points in recent years. Information about the youngest cohort 4 (>1984) is only available in the data since 2002, as they have been too young to participate in the SOEP before that year. These young individuals living in East Germany are on average not even 0.2 points less satisfied with their lives than their counterparts in West Germany and the gap is virtually nonexistent in some of the recent years.

Figure 3.4: Conditional Life Satisfaction Trends for Birth Cohorts



3.4.4 Robustness Checks

We have performed several robustness checks to analyze the sensitivity of our results, which we summarize shortly in this section. In our regressions, we have used nominal net household income. Our results do not change noteworthy, if we use real household income or the adjusted household income variable provided by the SOEP. Unfortunately, the adjusted income variable is not available for 2012 and 2013. Moreover, the computation of real household income using the consumer price index provided by the SOEP is not unproblematic, because the CPI is separated for East and West until 2000 and not separated afterwards. Two further robustness checks are concerned with our estimation sample. In our data, Berlin is still separated in East and West. But our results do not change noteworthy, if we exclude Berlin from the sample. Moreover, our results do not change noteworthy, if we exclude individuals with non-German citizenships from the sample. A last robustness check is kind of a Placebo check for the cohort differences.[7] Instead of dividing the regions in East and West, we divide the Western German federal states without Berlin in the regions North (Schleswig-Holstein, Hamburg, Bremen, Niedersachsen, Nordrhein-Westfalen, Hessen) and South (Bayern, Baden-Württemberg, Rheinland-Pfalz, Saarland). The conditional life satisfaction levels and trends do neither differ significantly between North and South nor between the four birth cohorts. Thus, it seems unlikely that the results for the cohort differences in the East-West gap and its trends, which have been presented in the previous section, are only a statistical artifact.

3.5 Concluding Remarks

Our results indicate that part of the raw East-West gap in life satisfaction is due to differences in household income and individual unemployment status. The conditional gap remains however sizeable. As we control for important economic and socio-demographic variables at the individual and household level, the remaining conditional East-West gap might be driven by individual non-monetary and unobserved differences between people living in East and West Germany. These differences can include psychological factors related to the system change such as a loss in identity as well as values and preferences driven by socialization (Van Hoorn and Maseland, 2010). Noll and Weick (2010) report, for example, evidence from the European Social Survey that the East-West gap is significantly reduced after controlling for differences in worries about the retirement system and in trust in institutions such as the legal system and the parliament. Moreover, regional economic aspects such as the provision of public goods and external effects of higher unemployment rates might be responsible for the lower conditional life satisfaction in Eastern Germany.

The second part of our analysis has focused on trends in conditional life satisfaction and the evolution of the East-West gap. Our results indicate that the conditional East-West gap in life satisfaction decreased in the first years after the German reunification and remained

[7] We thank Joachim Wagner for suggesting this robustness check.

quite stable and sizeable since the mid-nineties for the complete sample. Thus, it seems as if the lower life satisfaction in East Germany is quite persistent even 25 years after reunification and that convergence has not been achieved yet and might be achieved only very slowly in the future. This rather pessimistic outlook is however attenuated by our analysis of cohort differences, which indicate that the conditional East-West gap in life satisfaction is smaller and that convergence is much stronger for younger birth cohorts. For the youngest cohorts the gap is not even significant in recent years. Consequently, we can be more optimistic that convergence is achievable for younger birth cohorts with socialization in the reunified Germany. For older birth cohorts with socialization in the old system, i.e., in the German Democratic Republic, the gap remains however sizeable and convergence is rather unlikely. From these findings, we can indirectly conclude that the East-West gap seems rather to be driven by socialization in the old system of the German Democratic Republic than by individual and regional economic factors, which are either controlled for in our regressions or the same for different birth cohorts living in the same region, though the cohorts might be affected differently.

References

Burda, M. (2013). The East German economy in the twenty-first century. S. 195-216 in: H. Berghoff, U. Balbier (eds.), The East German Economy, 1945-2010.

Chadi, A. (2012). I would really love to participate in your survey!: Bias problems in the measurement of well-being. *Economics Bulletin 32*(4), 3111–3119.

Clark, A., A. Knabe, and S. Rätzel (2010). Boon or bane? Others' unemployment, well-being and job insecurity. *Labour Economics 17*(1), 52–61.

Clark, A. E., P. Frijters, and M. A. Shields (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature 46*(1), 95–144.

Easterlin, R. A. (2009). Lost in transition: Life satisfaction on the road to capitalism. *Journal of Economic Behavior & Organization 71*(2), 130–145.

Easterlin, R. A. and A. C. Plagnol (2008). Life satisfaction and economic conditions in East and West Germany pre-and post-unification. *Journal of Economic Behavior & Organization 68*(3-4), 433–444.

Ferrer-i Carbonell, A. (2005). Income and well-being: An empirical analysis of the comparison income effect. *Journal of Public Economics 89*(5-6), 997–1019.

Ferrer-i Carbonell, A. and P. Frijters (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal 114*(497), 641–659.

Frick, J. R., J. Schupp, and G. G. Wagner (2007). Enhancing the power of the German Socio-Economic Panel Study (SOEP)—evolution, scope and enhancements. *Schmollers Jahrbuch 127*, 139–169.

Frijters, P., J. P. Haisken-DeNew, and M. A. Shields (2004a). Investigating the patterns and determinants of life satisfaction in Germany following reunification. *Journal of Human Resources 39*(3), 649–674.

Frijters, P., J. P. Haisken-DeNew, and M. A. Shields (2004b). Money does matter! Evidence from increasing real income and life satisfaction in East Germany following reunification. *American Economic Review 94*(3), 730–740.

Gerlach, K. and G. Stephan (1996). A paper on unhappiness and unemployment in Germany. *Economics Letters 52*(3), 325–330.

Noll, H.-H. and S. Weick (2010). Subjective well-being in Germany: Evolutions, determinants and policy implications. *Happiness and Social Policy in Europe 7088*, 70–90.

Oishi, S., S. Kesebir, and E. Diener (2011). Income inequality and happiness. *Psychological Science 22*(9), 1095–1100.

Oswald, A. J. (1997). Happiness and economic performance. *The Economic Journal 107*(445), 1815–1831.

Smolny, W. (2009). Wage adjustment, competitiveness and unemployment–East Germany after unification. *Jahrbücher für Nationalökonomie und Statistik 229*(2-3), 130–145.

Smolny, W. (2012). Cyclical adjustment, capital-labor substitution and total factor productivity convergence–East Germany after unification. *Jahrbücher für Nationalökonomie und Statistik 232*(4), 445–459.

Smolny, W. and M. Kirbach (2011). Wage differentials between East and West Germany: Are they related to the location or to the people? *Applied Economics Letters 18*(9), 873–879.

Uhlig, H. (2008). The slow decline of East Germany. *Journal of Comparative Economics 36*(4), 517–541.

Van Hoorn, A. and R. Maseland (2010). Cultural differences between East and West Germany after 1991: Communist values versus economic performance? *Journal of Economic Behavior & Organization 76*(3), 791–804.

Vatter, J. (2012). Well-being in Germany: GDP and unemployment still matter. *RatSWD Working Paper Series 196.*

Winkelmann, L. and R. Winkelmann (1998). Why are the unemployed so unhappy? Evidence from panel data. *Economica 65*(257), 1–15.

Authors' Contributions

Life Satisfaction in Germany after Reunification:
Additional Insights on the Pattern of Convergence

Inna Petrunyk and Christian Pfeifer

Authors' contribution:

Joint contribution:

(i) conception and design of the work
(ii) choice of empirical methods
(iii) interpretation of statistical results
(iv) drafting the work

Inna Petrunyk's contribution (50%):

(i) data set preparation
(ii) literature review

Christian Pfeifer's contribution (50%):

(i) econometric analysis
(ii) results preparation

February, 17 2020

Inna Petrunyk

Christian Pfeifer

Chapter 4

Unemployment Benefits Duration and Labor Market Outcomes: Evidence from a Natural Experiment in Germany

Joint with: Christian Pfeifer. Authors' Contributions.

Available as: IZA DP No.11300.

Presented at:

International Association for Applied Econometrics **(IAAE)**, Sapporo (JP), 2017.
Italian Association of Labour Economists **(AIEL)**, Cosenza (IT), 2017.
Italian Society of Public Economics **(SIEP)**, Catania (IT), 2017.
American Economic Association (poster session) **(AEA)**, Philadelphia (USA), 2018.
Italian Workshop of Econometrics: Panel Data Models and Applications, Milan (IT), 2018.
Annual Congress of European Economic Association **(EEA)**, Köln, 2018.
Annual Congress of European Association of Labour Economists **(EALE)**, Lyon (FR), 2018.

Data replication statement:

A scientific use file of the data on completed rehabilitation in the course of insurance 2002-2009 was made available by the Research Data Centre of the German Pension Insurance. Unfortunately, the data are confidential and can not be provided as this would represent a violation of the confidentiality agreement. Statistical programs with all steps of data preparation and cleaning are available and, in case of request, support in accessing the data to anybody interested in getting the data from the Research Data Centre of the German Statutory Pension Insurance will be provided.

4.1 Introduction

Since the Lisbon Strategy, launched in March 2000, promotion of employment has become a priority for European policy-makers. Since then, a bunch of reforms aimed at re-establishing incentives to work and delay withdrawals from the labor market have been implemented. Among other policy changes, a profound rearrangement of the unemployment insurance (hereafter, UI) systems was initiated almost contemporaneously in many European countries. Older workers were the target population because the rate of longer unemployment spells is generally higher for this age group, either due to poor employment outlooks or disincentives of reemployment. Economic theory hints at poor work incentives when unemployment benefits become available, indicating that the generosity of the UI system matters (Moffitt and Nicholson, 1982; Mortensen, 1970). Theoretical literature suggests that a less generous UI scheme is related to in-creased job search effort of unemployed workers and shorter unemployment duration. Moreover, around the date benefits are exhausted, the intensity of recipients' job searches rises, which is associated with a spike in the unemployment exit rate at this point (Card et al., 2007; Card and Levine, 2000). Finally, a static labor supply model predicts lower reservation wages and higher probability to choose employment over unemployment in response to lower unemployment benefits. There is a growing literature investigating the optimal design of the UI scheme (Cahuc and Lehmann, 2000; Hopenhayn and Nicolini, 1997; Hurd, 1980; Schmieder et al., 2012b; Shavell and Weiss, 1979). The core issues are potential decreases in UI benefits over the unemployment spell as well as increases in potential benefits durations (PBD) during recessions and for different demographic groups. This literature explores the trade-off between the insurance function of and disincentives derived from UI, whereby welfare changes of affected individuals play an important role. On the one hand, UI ensures consumption smoothing for the unemployed, while on the other one, benefits-induced disincentives shape their job search efforts and moral hazard behavior. Because unemployed individuals shrink from accepting job offers, delaying their reemployment, and because monitoring of job search behavior is limited, the optimal UI scheme provides incentives that discourage the unemployed from persistent unemployment.

In 2004, the German government announced a major reform of the UI system that was implemented in February 2006 as part of the Hartz-Reforms. It involved a substantial reduction in the potential duration of regular unemployment benefits (unemployment benefits 1, hereafter UB-1) to stimulate employment among older workers by alleviating the disincentive effect of long compensation. The reform affected older age groups, while younger workers were not subject to the policy change. The design of the reform provides a natural experiment setting, on which our difference-in-differences identification strategy relies to investigate the causal relationship between the potential duration of UB-1 and the labor market outcomes of the affected individuals. The contribution of our paper is threefold. To begin with, we use an alternative outcome measure. In particular, as opposed to unemployment duration, which is a well-established out-come in the UI literature, our outcome of interest is aggregated for the complete calendar year from spell data. Hence, unlike the previous research that focuses on the initial spell limiting the analysis of a policy change to its short-term outcome, our estimates present the combined effect from the incidence and duration of recurring spells within a well-defined time period, thus capturing the reform effects in the middle-run that go beyond the first unemployment spell. Indeed, short-term effects may either under- or overstate the total cost of the reform if these also impact the incidence and duration of future unemployment. The most closely related to our work is that of Schmieder et al. (2012a), in which the outcome is aggregated over the first five years after the start of the initial UI spell.

Next, our sample consists of individuals with some health impairment, who underwent medical rehabilitation treatments. The main objective of rehabilitation measures directed to the working age patients is to retain their working capacity, to facilitate their reintegration into the labor market, and to avoid early retirement (Deutsche Rentenversicherung, 2014a). The use of routine data on labor market performance of participants in medical rehabilitation allows us to study the reform effects on this population group, opening an interesting perspective in evaluating the policy. Intuitively, a potentially positive impact of the policy change for this population group would encourage us to expect even larger effects for healthy individuals with no need for medical rehabilitation because their labor market opportunities should be better. In fact, empirical evidence suggests that work-limited workers with physical or nervous conditions suffer large and persistent declines in annual earnings as well as hours worked following work-limitation onset (Charles, 2003; Mok et al., 2008). Furthermore, the post-onset annual hours contraction of individuals who are older at work-limitation onset chiefly stems from reduced probability of labor participation. Worse labor market prospects of participants in medical rehabilitation might also derive from demand-side factors such as discrimination in recruiting or performance evaluation as well as from the coworkers' side that might adversely affect the workplace integration of work-limited workers (Colella and Bruyère, 2011). Adverse treatment of work-limited workers hampers their return to work as well as job retention. Therefore, we expect individuals with some health impairment to be less responsive to the UI incentives, interpreting our estimates as lower and upper bounds for the treatment effects in the total population. Official statistics on medical rehabilitation treatments signals that this population group has gained more importance in recent years (Deutsche Rentenversicherung, 2014b). In fact, in 2012 the German Statutory Pension Insurance approved 1,097,538 applications for medical rehabilitation, which is 0.9% more in comparison to the previous year and 3.2% more with respect to 2010. Therefore, we believe that it is important to analyze the labor market performance of this group in response to policy changes and the findings from this study are intended to provide possible directions for future research in this area.

Finally, we analyze the reform in a wider framework of institutional interactions from the labor market perspective, neglecting the incidental fiscal effects of the policy change. We distinguish between the intended and unintended labor market effects of the reform and show that it had a structural impact on the distribution of unemployment and employment. Based on a difference-in-differences approach, our results provide causal evidence for a significant decrease in the number of days in UB-1 and increase in the number of days in employment subject to social insurance contributions. However, the findings also suggest a significant increase in the number of days in social assistance (unemployment benefits 2, hereafter UB-2) granted to unemployed job-seekers upon exhaustion of UB-1 to provide them a living at the subsistence level. This result is consistent with recent work on the effectiveness of more comprehensive reforms of labor market institutions as opposed to one policy reform at a time (Fremigacci et al., 2010; Pellizzari, 2006). From the labor market and social policy perspective, transitions to UB-2 represent an unintended consequence of the reform, limiting the success of a policy change that aims to increase labor supply via reductions in the generosity of the UI system. To the best of our knowledge, this is the first paper that explicitly investigates this important aspect of the major reform of UB-1 in Germany.

The remainder of the paper is organized as follows. Section 4.2 presents a literature review. Section 4.3 describes the institutional setting of the German UI system. Section 4.4 presents the data set and samples. Section 4.5 describes our econometric approach and summary statistics for the variables of interest. Section 4.6 reports the estimation results. Section 4.7 concludes with a short summary and discussion of the main findings.

4.2 Literature Review

This paper explores the impact of PBD on workers' labor market performance and thus con-tributes to the stream of literature analyzing the disincentives provided by the UI system. Empirical findings broadly support the predictions of job search models. Insurance protection offered by the UI system significantly affects exits from unemployment of benefits recipients. A large body of literature explores the impact of changes in UI parameters on the duration of the first nonemployment spell, limiting the analysis of a policy change to its short-term outcome, partially because the identification of precise longer-term effects is empirically challenging (Lalive, 2007; Meyer, 1990). However, in their recent contribution, Schmieder et al. (2012a) remedy this shortcoming. The authors use a regression discontinuity design in Germany and examine the long-term impact of an extension of UI duration captured by the total days receiving UI benefits and the total days in nonemployment in the first five years after the start of the initial UI spell. Thus, the positive effect of longer UI duration on the sum of days spent in nonemployment combines the effect from the initial nonemployment spell and the incidence and duration of additional spells. Their results further indicate that a large part of the effect of UI extension is captured by a longer initial nonemployment spell and allowing for multiple spells reduces the impact of longer PBD. In other words, the effect of PBD on total nonemployment is smaller than the effect on the duration of the initial nonemployment spell. This implies that the long-term effect of UI on overall nonemployment is smaller.

Hunt (1995) investigates the impact of large increases in PBD for older workers in former West Germany, distinguishing between escapes to employment and out of the labor force. Applying a difference-in-differences method, the author reports longer unemployment spells that lower the hazard rates to both exit destinations. A positive relationship between the generosity of the UI system and the duration of unemployment spells has become a stylized fact, although less is known about the underlying mechanisms. Higher UI benefits are associated with a strong negative effect on the probability of leaving unemployment that rises dramatically just prior to benefits exhaustion (Meyer, 1990). This evidence has been generally attributed to moral hazard caused by a substitution effect, according to which UI alters the relative price of leisure and consumption, lowering the marginal incentive to search for a new job. More recent studies focus on identification of the channels through which unemployment benefits affect search behavior. In a sample of U.S. workers, Chetty (2008) examines the importance of moral hazard (substitution effect) versus liquidity (income effect), referring to the principle according to which, as with any uncompensated labor supply elasticity, the total labor supply response to a change in UI benefits incorporates both effects. Indeed, the author finds that more than half of the labor supply elasticity is due to a liquidity effect. This indicates that a large fraction of benefits recipients is liquidity constrained and that when unemployed workers are unable to smooth consumption due to a transitory income shock, unemployment benefits affect job search behavior through a liquidity effect in addition to the moral hazard mechanism high-lighted in previous studies. However, when consumption can be smoothed perfectly, UI benefits raise unemployment durations essentially through moral hazard.

Lalive et al. (2006) show that replacement ratio (hereafter, RR) as well as PBD are both important policy tools that can alter behavior, although they prompt rather different behavioral responses. In the example of an increased RR and extended PBD in Austria in 1989, the authors observe an increase in unemployment duration, which is larger in case of a simultaneous increase in replacement rate and potential benefits duration compared to isolated increases in these UI parameters. Furthermore, they find a strong association between increases in PBD and exit rates from unemployment around the date of benefit expiration, while behavioral adjustments in response to an increase in RR follow a more uniform distribution over the unemployment spell. This pattern indicates that an increased RR has the largest impact on the behavior from the start of the unemployment spell, while extended PBD does so around the date of benefits expiration. Indeed, in many studies, a large spike is observed in the exit rate from registered unemployment at the point of benefits exhaustion, suggesting that as the remaining period of a benefits receipt declines, the value of remaining unemployed declines as well. The latter adversely affects the reservation wage and boosts job search intensity as workers approach benefits exhaustion. After UI benefits have expired, reservation wage and job search efforts remain constant. As a result, employment hazard increases up to the date of benefits exhaustion and does not change afterwards (Mortensen, 1977).

The spike in exit rates is mostly interpreted as a manifestation of non-stationary search behavior of benefits recipients who wait until their benefits expire to return to work, potentially accepting jobs with lower stability and lower wages compared to those unemployed who exit unemployment at an earlier stage (Caliendo et al., 2013). This strategic job search behavior discloses the distortionary effects of the UI system and social insurance programs in general, although the effective moral hazard effect is supposed to be significantly lower in recessions than in booms (Feldstein, 2005; Schmieder et al., 2012b). Empirical literature finds evidence of strategic job search behavior caused by more generous unemployment benefits durations. For instance, Lichter (2016) uses an exogenous variation in PBD originated from a policy change in Germany. The reform, implemented in 2008, involved an extension of potential benefits duration for workers of specific age groups. Applying a difference-in-differences technique, the estimates provide causal evidence of reduced search effort, measured by the number of job applications and the probability of applying for jobs in distant areas, in response to the extension of PBD. These findings are in line with the theoretical predictions of standard job search models. Furthermore, instrumental variables estimates show that the reduction in search effort induced by the reform caused a significant decrease in the short-run job finding rate. Evaluated at the mean, a 10 percent increase in the number of filed applications is associated with an increase in the short-run job-finding rate by about 1.3 percentage points. Additionally, individuals may make arrangements about the date of return to work. Empirical estimates support the hypothesis that at an early stage of an unemployment spell, firms and employees plan its ending date (Katz and Meyer, 1990; Meyer, 1990). Intuitively, if workers are bound to firms by issues such as implicit contracts, moving costs, or specific human capital, there is a strong incentive to tie recall decisions to the length of UI benefits.

A growing strand of literature focuses on transitions from employment into unemployment induced by UI incentives (Winter-Ebmer, 2003). In fact, unemployment entry and unemployment duration both explain the dynamics of the aggregate unemployment rate. Studying the unemployment incidence before and after the major reform of UB-1 in Germany in 2006, Dlugosz et al. (2014) apply a difference-in-differences method and find decreased unemployment inflows for individuals aged 52 and older. Moreover, the results indicate large anticipation effects of the reform in the three months before the policy change came into force, which greatly distorted the short-term effects of the reform. In particular, relative to younger workers, transition rates into unemployment of workers aged 52 and older substantially increased in the anticipation period, which suggests a change in the composition of the unemployed in response to the reform. In fact, for a limited period right after the introduction of the reform, unemployment inflow decreased in a more pronounced way than in the absence of anticipation. This observation could be explained by the anticipation of dismissals and resignations from the post-reform period to the pre-reform one. Nevertheless, the decrease in unemployment inflows following the reform far outweighed the anticipation effect. Based on this knowledge gain, Lo et al. (2017) focus on the age group with the smallest anticipation effect, for which no systematic decrease in unemployment inflows after the reform has been observed. Furthermore, the authors exclude periods with unemployment inflows during the potential anticipation period to remove anticipation effects. Based on a sample of male unemployed with full-time employment before unemployment, they distinguish between transitions to desired destinations such as non-low-wage full-time employment, transitions to less desired destinations such as subsidized self-employment, low-wage full-time employment and transitions to other states such as part-time employment, previous employer, secondary labor market or long training programs. A central contribution of their study is that it provides more detailed insights into the impact of the unemployment compensation system on the labor market performance of the affected individuals. In particular, the authors conclude that (non) low-wage workers tend to take up (non) low-wage employment. The probability of being recalled to the previous employer is higher for low-wage workers, while the probability to take up subsidized self-employment is higher for non-low-wage workers than for low-wage ones. Although the authors analyze the impact of the major German reform on transitions to important exit destinations, they disregard transitions to UB-2 that are perhaps no less important.

In the European context, the Finnish 1997 reform decreased PBD for older workers to enhance employment incentives and cut expenditures on unemployment indemnity. Kyyrä and Ollikainen (2008) apply a difference-in-differences approach to analyze transitions to employment, which represent a targeted exit destination, and less desired transitions out of the labor force and into active labor market programs that are more relevant in the Nordic countries. Based on the altered flows into unemployment in anticipation of the policy change, the authors exclude the involved groups from the analysis. Their findings do not reveal large increases in the employment hazard around the date of benefits exhaustion. Instead, the hazard rates for labor market programs and non-participation present substantial rises. These results point to an important interaction between labor market institutions in which the UI system and the early retirement scheme represent an attractive pathway to labor market withdrawal prior to the regular old-age pension. The French reform of the UI system in 2003 also involved substantial shortenings in PBD for older workers (Fremigacci et al., 2010). The findings point to in-creased transition rates out of unemployment in response to the policy change. A decomposition of the outflows from unemployment allows a closer insight into the exit destinations. It reveals a positive, although tiny, effect on exits to employment, but a substantial positive effect on transitions to unemployment assistance, which is granted to individuals who have exhausted the unemployment benefits or did not qualify to receive them. In fact, a common feature of the European institutions is the interaction of UI with other social security programs (Schmieder and Von Wachter, 2016). Thus, upon exhaustion of unemployment benefits, job seekers can shift to unemployment assistance or other basic income support programs. The author concludes that the major effect of the policy change in France was to shift job seekers from unemployment benefits to unemployment assistance. These non-negligible transitions represent an unintended consequence of the reform that might limit the success of a policy change aiming to increase labor supply via reductions in the generosity of the

UI system (Pellizzari, 2006).

The shift to unemployment assistance following a reduction in PBD largely explains the identified spikes in exits from unemployment around the date of benefits exhaustion and supports the relevant work of Card et al. (2007) on the true mechanisms behind these spikes. In their study based on a large sample of job losers in Austria, Card et al. (2007) show that the ob-served spikes may exaggerate the extent of moral hazard induced by UI. The authors underscore the importance of how unemployment spells is measured (time spent on the unemployment system vs. time to next job), which determines the magnitude of the spikes at benefits exhaustion. The results indicate that the hazard rate of reemployment accounts only for a small part of the exit rate from registered unemployment. This finding reveals that many unemployed workers leave the unemployment register around the date of benefits exhaustion without returning to work, which sheds light on the divergence between the two measures of unemployment spell.

Other significant contributions that rank as complementary research in this field analyze the impact of the UI reform in terms of job match quality. Contrary to the standard search models predictions, the evidence on the impact of UI benefits extension on post-unemployment job quality, as measured by earnings in the new job, is mixed. A large body of existing literature finds no effect. The study of Lalive (2007), based on a regression discontinuity approach, reveals that large extensions in UI benefits in Austria increased unemployment duration, reduced transitions to a regular job, and increased the duration until a new job, but did not affect average daily wages. In case of the reform in France, the results suggest that faster exits to jobs were not related to a decline in job stability, measured by transition rates from employment back to unemployment. Another study focuses on the French reform in 2000 that induced a large extension in PBD conditional on past employment duration over a reference period. Based on a regression discontinuity design, Le Barbanchon (2016) finds that this policy change had a significantly large and negative impact on unemployment exits to work, but no improvement of the match quality captured by hourly wage and employment duration. No effects on the quality of jobs that workers found after periods of unemployment have been identified in Slovenia, where a change in the UI system entailed substantial reductions in PBD for selected age groups of workers (Van Ours and Vodopivec, 2006, 2008). The results from this natural experiment indicate a positive effect on the exit rate from unemployment to new jobs without affecting the quality of post-unemployment job matches. In particular, no effect on wages, on the distribution between permanent and temporary jobs, or on the duration of the post-unemployment jobs has been detected. These findings, based on a difference-in-differences approach, allow the authors to conclude that the unemployment benefits reform in Slovenia diminished the moral hazard induced by the UI system.

In contrast to these studies, statistically significant negative and positive UI wage effects have also been identified. In their recent paper, Schmieder et al. (2016) adopt a regression discontinuity design and find negative effects of UI extensions on reemployment wages in Germany. Furthermore, they show that this effect results from the existence of several potentially offsetting components. The first may increase reemployment wages due to an increase in reservation wages or to stronger bargaining power. The second may reduce search effort, thus leading to longer nonemployment spells. If wage offers decrease over time due to stigma or human capital depreciation, workers will face a reduction in reemployment wages. Their results point to tiny reservation wage effects, leading the authors to conclude that reductions in reemployment wages over the unemployment spell cannot arise from changes in reservation wages. If longer unemployment spells do not help workers find a better job due to a negligible positive reservation wage effect and a large negative duration effect, the optimal UI length should be shortened. Contrary to this conclusion, Nekoei and Weber (2017), using a regression discontinuity design in Austria, show that extensions of relatively short UI benefits cause higher reemployment wages persistent over time that do not substitute other valuable job characteristics. This evidence supports the idea that UI subsidizes productive job search and not just unproductive leisure. Relevant studies in this research area are summarized in Table 4.1.

Table 4.1: Literature Review

Author (year)	**Country**	**Reform**	**Data**	**Research Design**	**Results**
			A. Unemployment Entry		
Dlugosz et al. (2013)	Germany	PBD shortening in 2006	Administrative data	DiD	Identification of bounds for the long-term reform effect due to strong anticipation effects. Probability of entering unemployment declined. The decrease in unemployment inflows after the reform far outweighed the anticipation effect.
Winter-Ebmer (2003)	Austria	PBD extension in 1988	Administrative data	DDD	Increase in unemployment entry, caused not by voluntary quits but by layoffs by employers.
			B. Unemployment Exit		
Hunt (1995)	West Germany	PBD extensions in 1985-1987; RR decrease in 1984	Survey data	DiD	Increase in unemployment duration, lower hazard rates to employment and to out of labor force. The impact of RR decrease for the childless is unclear.
Lichter (2016)	Germany	PBD extension in 2008	IZA Evaluation Dataset	DiD	Evidence for strategic job search behavior. Significant decrease in job search effort and short-run job-finding rate.
Schmieder et al. (2012a)	Germany	PBD extension in 1987-1999 (based on a discontinuity)	Administrative data	RDD	Increase in nonemployment duration until the 1st job. Much lower increase in days in nonemployment over 5 years. Allowing for added nonemployment spells lowers the effect of UI extensions on days of nonemployment.

Continued on next page

Schmieder et al. (2012b)	Germany	PBD extension in 1987-1999 (based on a discontinuity); PBD extension in 1999-2004 (based on a discontinuity)	Administrative data	RDD	Increase in nonemployment duration. Effective moral hazard effect of UI extensions is significantly lower in recessions than in booms.
Lalive et al. (2006)	Austria	PBD extension in 1989; RR increase in 1989	Administrative data	DiD	Increase in unemployment duration. Only PBD extension: a strong change in the exit rate around the date of benefits exhaustion. Only RR increase: the change is distributed more uniformly over the unemployment spell. Simultaneous RR increase & PBD extension: larger increase in unemployment duration compared to an isolated alteration in RR & PBD.
Kyyrä and Ollikainen (2008)	Finland	PBD shortening in 1997	Administrative data	DiD	No increase in employment hazard, but large increase in hazard rates to active labor market programs and to non-participation.
van Ours and Vodopivec (2006)	Slovakia	PBD shortening in 1998	Administrative data	DiD	Significant increase in the exit rate out of unemployment to employment and to active labor market programs.

Continued on next page

Card and Levine (2000)	USA	Short-term PBD extension in New Jersey in 1996. Politically motivated and unrelated to labor market conditions	State-level data and administrative data.	DiD	Modest impact on the fraction of recipients who exhausted regular benefits and on the average exit rate from regular UI. No impact on the average spell length.
Katz and Meyer (1990)	USA	Variation in UI parameters across states in 1978-1983	Survey data and administrative data	Semiparametric hazard model	Large increases in the escape rate from unemployment through recalls and finding of new jobs around the date of benefits exhaustion. A PBD extension has a larger disincentive effect on unemployment duration than an increase in UI benefits level with the same impact on the budget.
Meyer (1990)	USA	Variation in UI parameters across states in 1978-1983	Administrative data	Parametric and semiparametric hazard models	Higher UI benefits have a strong negative impact on the probability of leaving unemployment, which rises sharply at the benefits exhaustion date. When the length of benefits is extended, the probability of a spell termination is high in the week benefits were previously expected to lapse.
C. Post-Unemployment Jobs					
Schmieder et al. (2016)	Germany	PBD extension in 1987-1999 (based on a discontinuity)	Administrative data	RDD	Negative UI effect on reemployment wages, which combines two offsetting effects: the effect of UI on reservation wages and the effect of nonemployment durations on wage offers. The negative wage effect emerges chiefly from large nonemployment duration effects.

Continued on next page

Caliendo et al. (2013)	Germany	PBD extension in 2001-2003 (based on a discontinuity)	IZA Evaluation Dataset	RDD	Spike around the date of benefits exhaustion. Evidence for non-stationary search behavior (the unemployed who find a job at or after the date of benefits exhaustion experience less stable employment patterns and receive lower wages compared to their counterparts who receive extended benefits and exit unemployment in the same period).
Lo et al. (2015)	Germany	PBD shortening in 2006	Administrative data	DiD	Shorter unemployment duration. Effects are smallest for more desirable destination states such as recalls and non-low-paid full-time employment. (Non-) low-wage workers tend to take up (non-) low-wage employment. The probability to experience a recall to the previous employer is higher for low-wage workers, while the probability to take up subsidized self-employment is higher for non-low-wage workers than for the low-wage ones. Strong effects on subsidized self-employment at the benefits expiration date.
Lalive (2007)	Austria	PBD extension in 1989-1991 (based on a discontinuity)	Administrative data	RDD	Large benefit extensions increase unemployment duration, decrease transitions to a regular job, and increase the duration until a new job is taken, but do not have a distinct effect on post-unemployment job quality as measured by daily wages in the new job.
Nekoei and Weber (2017)	Austria	PBD extension in 1989-2010 (based on a discontinuity)	Administrative data	RDD	Positive UI effect on reemployment wages, which combines two offsetting effects (positive effect: workers are more selective in their job search and seek higher-wage jobs; negative effect: wages decrease due to longer unemployment spells induced by UI.
Le Barbanchon (2016)	France	PBD extension in 2000-2002 (based on a discontinuity)	Administrative data	RDD	No effect on wage and employment duration, but a significantly large and negative effect on unemployment exit to work.

Continued on next page

van Ours and Vodopivec (2008)	Slovakia	PBD shortening in 1998	Administrative data	DiD	No discernible effect on reemployment wages, the probability of finding a permanent rather than a temporary job, or the duration of post-unemployment job.
D. Interaction between Labor Market Institutions					
Card et al. (2007)	Austria	PBD extension in 1981-2001 (based on a discontinuity)	Administrative data	RDD	Spike around the date of benefits exhaustion, but reemployment hazards rise much less than unemployment exit hazards, because some individuals leave the unemployment system at the date of their benefits expiration and do not return to work. The reemployment rate depends on institutional factors among which the availability of post-exhaustion benefits.
Pellizzari (2006)	European countries	PBD extension or RR decrease in 1994-2001	Survey data	Proportional hazard model	Recipients of UI benefits eligible to complementary or ensuing welfare schemes are less responsive to changes in the level or duration of their benefits. Comprehensive reforms of labor market policies are more effective than reforming one program at a time.
Fremigacci (2010)	France	PBD shortening in 2003	Administrative data	RDD	Increase in transition rates out of unemployment. Substantial positive effect on transitions to unemployment assistance. Positive though modest effect on exits to employment. Faster exits to jobs not associated with a downturn in job stability.

End of table

4.3 Institutional Setting

4.3.1 The German Unemployment Compensation System

Similar to other European countries (Pellizzari, 2006, among others), the unemployment compensation system in Germany relies on two main pillars, UB-1 ("Arbeitslosengeld 1") and UB-2 ("Arbeitslosengeld 2"). UB-1 is funded by employee and employer contributions and is administered by the Federal Employment Agency. All employees subject to social security contributions are covered by this UI. However, entitlement to receive UB-1 is conditioned on contribution to the insurance scheme for at least 12 months within the last 24 months before a job loss, and its duration depends on the age and employment history of unemployed workers. PBD discontinuously increases with age to account for difficulties that older unemployed individuals might have in re-entering the job market. Workers who reached the statutory retirement age are excluded from the UI coverage. Monthly benefits replace 60% (67% for claimants with children) of the last net salary (capped at the social security ceiling). Payments are usually annulled for up to 12 weeks if employees take the initiative to terminate the employment relationship, therefore reducing the maximum benefits duration. Furthermore, recipients of UB-1 are required to actively search for a job and to prove their job searching activities (applications and responses by potential employers) upon request from the local employment office. Lack of compliance with these requirements may lead to benefit cuts. Upon exhaustion of UB-1 or in case of no entitlement to them, needy unemployed jobseekers receive tax-financed UB-2, which is unconstrained by previous earnings and is granted without temporal restrictions. UB-2 is means-tested against household income and aims at providing a living at the subsistence level. Non-compliance with the rules can result in benefit sanctions that reduce the compensation level.

4.3.2 The German Reform of Unemployment Benefits 1

Our study evaluates the major reform of UB-1 in Germany, originated from an institutional change called Hartz-Reforms. The reform was announced in 2004 and came into force in February 2006, affecting workers who lost their jobs after 31 January 2006. This major policy change implied a substantial reduction in the potential duration of UB-1 and largely annulled the extensions of the 1980s that were motivated by an increasing unemployment rate and long average spell duration among older workers in West Germany (Hunt, 1995). The core motivation of this reform was poor labor market performance of workers above 50 (Dietz and Walwei, 2011). Aiming at promoting reemployment among seniors, the introduced innovations were particularly penalizing for older workers with the maximum reduction in the potential benefits duration by 14 months. However, this reform lasted only till December 2007. In fact, as early as in January 2008, the German government enacted a new reform of the UI scheme, re-extending the PBD for older age groups. The main driving force were fairness considerations, according to which workers with contributions to the UI system for a longer period should be granted longer benefits durations. Nevertheless, this was a minor policy change that did not lead to the pre-reform state. In fact, the maximum extension in the potential benefits duration amounted to only six months, and only few age groups were affected by this adjustment. Table 4.2 illustrates the major and minor policy changes in the potential duration of UB-1 for each age category.

Table 4.2: Maximum Duration in Months of Unemployment Benefits in Germany

Age category	Before 2/2006	Reduction	2/2006-12/2007	Extension	Since 1/2008
<45	12	0	12	0	12
45-46	18	6	12	0	12
47-49	22	10	12	0	12
50-51	22	10	12	3	15
52-54	26	14	12	3	15
55-56	26	8	18	0	18
57	32	14	18	0	18
>57	32	14	18	6	24

The major reform in February 2006 also modified eligibility criteria and work history requirements for receipts of UB-1. Under the old regime, workers were eligible if they had worked at least 12 out of the 36 months preceding unemployment. After the reform, employment during at least 12 out of the last 24 months is required. Work history at the moment of the claim is crucial for qualification for the maximum benefits duration. Before the reform, individuals must have worked during the previous 84 months for a number of months equal to at least twice the potential benefits duration. Under the post-reform regime, they must have worked for a number of months equal to at least twice the potential benefits duration within the last 36 months prior to unemployment. The replacement rate for the level of benefits was not affected by the reform.

4.3.3 Other Relevant Policy Changes

UB-2 was introduced in January 2005. It largely replaced two previous components of the German unemployment compensation system, unemployment assistance ("Arbeitslosenhilfe"), granted to unemployed jobseekers upon exhaustion of unemployment benefits, and social assistance ("Sozialhilfe"), granted to all other needy individuals, in particular to those who have never been employed. This policy change aimed to reduce the dependence on benefits. To this end, the reform introduced strict rules aiming to motivate recipients to intensely cooperate with job centers and actively search for a job, on the one hand, and enhanced support to at least increase their employability by promoting education programs in skills valued on the labor market, on the other one. Practical enforcement of new rules is achieved through a complex of benefits sanctions.

Apart from the minor reform of UB-1 in January 2008 that partially re-extended PBD for older workers, the German old age pension system has also been redesigned. In particular, until 2003, workers unemployed for at least one year could take advantage of the early retirement scheme without pension shortenings at the age of 60. From 2004 on, entry into pension due to unemployment became possible at the age of 63 at the earliest, thus postponing early retirement. The analysis of the effects of UI reforms for workers approaching retirement age differs from those for other workers and is not the focus of this paper. The rationale behind this choice lies in the specificity of this population, which requires a simultaneous consideration of the unemployment compensation system and the retirement scheme. Just before retirement age, search intensity severely decreases, which weakens the incentive effect of cutting unemployment benefits. Indeed, in many European countries, extended benefits programs for long-term unemployed and early retirement schemes allow senior workers to leave the labor market before the legal retirement age (Kyyrä and Ollikainen, 2008). This suggests that retirement can not be modeled separately from other spells out of work, justifying the exclusion of workers approaching retirement age in this study. Finally, during the period under study, the state legal requirements for the approval of applications for medical rehabilitation have not been modified so that the pre- and post-reform participants in rehabilitation treatments do not systematically differ, leaving no room for a composition effect.

4.4 Data Set and Samples

For our analysis, we use the routine data collected by the German Statutory Pension Insurance.[1] The longitudinal data set includes a random sample of 20% of all individuals who completed medical rehabilitation treatments granted by this insurer. A characterizing feature of medical rehabilitation consists in treating, among others, health deficiencies such as renal failure, disorders involving the metabolic and endocrine systems (e.g., diabetes mellitus), nervous system (e.g., migraine and sleep disorder), circulatory system (e.g., heart failure), respiratory system (e.g., asthma), digestive system (e.g., liver disorder), musculoskeletal system (e.g., back pain), mental and behavioral disorders (e.g., depression and alcohol abuse), and skin diseases (e.g., dermatitis. The most recurrent health disorder is the low back pain, which in 2013 accounted for 31.5% of all medical and other rehabilitative services provided by the German Statutory Pension Fund (Deutsche Rentenversicherung, 2013). In fact, it is the largest finance provider of medical rehabilitation treatments for the

[1]While the administrative data from the Sample of Integrated Labour Market Biographies (SIAB) contain daily spell information on employment periods subject to social security contributions, job search periods, participation in active labor market programs, and claim periods of UB-1 as well as UB-2, the advantage of our data over the SIAB data consists in the availability of information on individuals' health limitations and other important control variables.

employed individuals in Germany, followed by the statutory health insurance, and aims essentially at preventing costs connected with early retirement following the principle of rehabilitation before pension. A scientific use file of the data on completed rehabilitation in the course of insurance 2002-2009 was made available by the Research Data Centre of the German Pension Insurance (Deutsche Rentenversicherung, 2012). The data set consists of three databases.

SUFRSDV09BYB: It is a pension insurance follow-up database that provides information on insurance relationship and amount of contribution payments. Information on the outcome variables of interest in this research field such as number of worked days, days in UB-1, and days in UB-2 are also collected in the database and employed in this study.

SUFRSDV09MCB: It includes all the cases with at least one completed medical rehabilitation, which in single cases may be supported by vocational rehabilitation and/or followed by granted pension benefits. The following variables contain detailed information on rehabilitation events during the reporting period 2002-2009: type of granted rehabilitation, implementation form on an inpatient or outpatient basis, begin/end of the treatment and its duration in days, rehabilitation region, and medical discharge diagnoses. Moreover, labor market variables at the moment of or shortly before the application for a rehabilitation treatment such as labor status, most recent activity, and occupational status are also available.

SUFRSDV09KOB: Standard socio-demographic characteristics such as birth/death year, nationality, residence region, gender, marital status, and education of the sample complete the data.

We restrict our sample to individuals aged between 38 and 62 in the outcome year who participated in only one medical rehabilitation in the observation period (approximately 75% of the whole sample), either before the UB-1 reform or thereafter. In this way, the data set takes the form of pooled cross-sections with information before and after rehabilitation. Our dependent variables measure days in employment subject to social insurance contributions and days in registered unemployment in the outcome year. Non-employment, such as retirement due to health reasons or other labor market exits, is not considered. Taking into account the timing of the reform of the old age pension system, we keep only individuals with completed rehabilitations between 2003 and 2008, for whom we can observe labor market outcomes in the years between 2004 and 2009. Observations with missing values are dropped. Based on this general sample composition, we consider three distinct samples.

Preferred Sample A: In our preferred sample A, we keep only years 2005 (pre-reform) and 2007 (post-reform) and focus on individuals employed before rehabilitation. This temporal restriction relies on the following considerations. First, UB-2 was introduced only in January 2005, while the potential duration of UB-1 was partially re-extended as early as in January 2008. The enacted policy changes prompt us to exclude the years prior to 2005 and after 2007. Second, the exclusion of the year 2006 is motivated by the potential transition period and anticipation effects of the reform. Indeed, most individuals with completed rehabilitation in 2005 and days in unemployment benefits measured in 2006 are more likely to have entered unemployment under the old regime. This is, however, not the case for rehabilitations finished in 2006, when the reform came into force. Although this restriction does not allow for controlling for pre-reform trends, it provides important insights into the impact of the policy change and results in an analysis free from any distortions that stem from anticipation effects of the reform. To check whether the common trend assumption holds in our data, we extend the considered time period to years between 2004 and 2009 in sample B. We further restrict our preferred sample A to those employed at least 12 months in the two calendar years before rehabilitation and the year of rehabilitation, i.e., during three calendar years before the outcome year. Although imperfect in its nature due to data construction, this restriction is supposed to approximate the sample's fulfillment of eligibility criteria both under the old and the new regime. As a result, sample A reduces the total number of observations by less than 10% and adds up to 94,990 observations, of which 46% are female and 52% are in the post-reform period (year 2007). Table 4.3 shows the data structure for sample A.

Table 4.3: Data Structure (Example for Sample A: 2005/2007, Employed before Rehabilitation)

Year Group	2002	2003	2004	2005 [pre-reform]	2006 [reform]	2007 [post-reform]
2005 (pre-reform)	sample restriction (full entitlement length): work. days 2002/03/04≥365		- employed -rehab. - rehab. exit	outcomes		
2007 (post-reform)			sample restriction (full entitlement length): work. days 2004/05/06≥365		- employed - rehab. - rehab. exit	outcomes

Extended Sample B: We extend our preferred sample A to outcome years 2004-2009, which enables us to compare the pre-reform and post-reform trends and thus verify the fulfillment of the common trend assumption. The final sample B consists of 306,230 observations.

Additional Sample C: As for Sample A, we keep only years 2005 and 2007, but we focus on individuals either unemployed or non-employed before rehabilitation. The subsample with the unemployed amounts to 15,857 observations, while that with the non-employed consists of 16,529 observations.

4.5 Difference-in-differences (DiD) Design and Variables

The major reform of UB-1 affected only individuals aged 45 or older and had a more pronounced impact for older age groups. The natural experiment setting allows us to apply a standard difference-in-differences approach, with assignment to the treatment and control groups according to age. The general estimation framework for our specification as described in equation (4.1) can be estimated by using linear regressions with ordinary least squares (OLS). A robustness check adopting count data models reveals virtually the same results.

$$Y = \alpha + \beta_1 AGE + \beta_2 YEAR + \beta_3 AGE \times YEAR + \delta X + \epsilon \tag{4.1}$$

β_1 is the parameter for the treatment group specific effect (age trend), β_2 is the parameter for the time trend common to the control and the treatment groups, β_3 is the parameter of interest that provides the DiD estimate of the average treatment effect on the treated (ATT), X is a vector of covariates defined below, α is a constant, and ϵ is the error term.

Outcome variables Y: Y denotes the outcome of interest measured in the calendar year after medical rehabilitation and indicates days in UB-1, days in UB-2, and days in employment subject to social insurance contributions (WORK). They are aggregated for the complete calendar year from spell data and range from 0 to 365 days, providing information on the combined effect from the incidence and duration of recurring spells within a well-defined time period, thus capturing the reform effects that go beyond the first unemployment spell. All three outcome variables are, of course, highly correlated with each other, because more working days, ceteris paribus, decrease the number of days in registered unemployment. Nevertheless, we think it is important to analyze all these outcome variables separately. In fact, the total number of days can not only be divided in employment and registered unemployment, but also in other sources of non-employment (e.g., family responsibility, early retirement). Apart from non-employment, we also exclude self-employment, minor employment, and civil service. For our preferred estimation sample A (2005/2007, employed before rehabilitation), the number of days in UB-1 is on average 39.6, the number of days in UB-2 is 6.2, and the number days in employment is 261.7. Because our outcome variables do not exhaust all possible labor market states, they do not sum up to 365 days.

Time period YEAR: The time period dummy $YEAR$ captures aggregate factors that would cause changes in Y even in the absence of a policy change. In our preferred specification (sample A), we include only years 2005 and 2007 so that $YEAR$ is a dummy for 2007, indicating the post-reform period. In the extended specification (sample B), we examine the pre-reform and post-reform trends for the years 2004 to 2009, testing for the common trend assumption. Here, $YEAR$ is a set of dummy variables, while the year 2004 serves as the reference group.

Specifications AGE: The dummy variable AGE captures possible differences between the treatment and control groups independent of the policy change. As a first step, the treatment group is defined by all individuals aged 45 or older and amounts to 77% of individuals in our preferred sample A. Because reductions in the PBD implied by the policy change varied with respect to age categories, the treatment group is further defined according to these categories. For a more detailed picture, we estimate a specification with dummies for each year of age. Age distribution across years reveals at least 1,000 observations in each year for each year of age (see Appendix Table 4.A1). To facilitate the interpretation, we predict and plot non-linear age profiles instead of interpreting the coefficients. These definitions, however, might suffer from some imprecision. In particular, we only consider age in years so that in 2007 (post-reform period), individuals within a specific age threshold (e.g., age 45) are not affected by the reform for the complete calendar year. This measurement error would bias the coefficients toward zero, so we expect a lower treatment effect for these age threshold groups.

Control variables X: We further account for sex, marital status, nationality, education, job position, occupation, federal state, and rehabilitation diagnosis in our model. These are treated as control variables, which results are not further discussed. Appendix Table 4.A2 offers a closer look at the descriptive statistics of these variables for our preferred estimation sample A (2005/2007, employed before rehabilitation). It is noteworthy to mention that our estimates of the treatment effect are largely unaffected by the inclusion of the control variables, which indicates that our estimates are likely to be unbiased.

Treatment intensity: In the next step, we revise our DiD strategy to examine the difference in the treatment intensity between age groups, implied by the reform design. The treatment intensity was null for all age groups in 2005 and for those under the age of 45 in 2007. But there was positive and heterogeneous treatment intensity for individuals aged 45 and older in 2007. Younger workers were affected to a lesser extent than older ones. In fact, the age group 45-46 suffered the smallest benefits cut by only six months, while age groups 52-54 and >56 suffered the largest reduction by 14 months. To address the fact that treatment intensity is correlated with age, we use different specifications of age as controls (age, age_2, age_3, age dummies). Furthermore, we run a placebo test for years 2004 and 2005 and act as if the UB-1 reduction would have occurred in 2005, although 2004 and 2005 are both pre-reform years. To this end, we adopt the following specification in equation (4.2):

$$Y = \alpha + \beta_1 AGE + \beta_2 YEAR + \beta_3 REDUCTION2007 + \delta X + \epsilon \tag{4.2}$$

4.6 DiD Regression Results

4.6.1 Main Results for 2005/2007

In our main analysis, we focus on sample A, i.e., on individuals who were employed before the rehabilitation and for whom we observe labor market outcomes in the pre-reform year 2005 and the post-reform year 2007. We use three different specifications of age (age treatment dummy for age≥45, age treatment categories according to different reductions implied by the reform, and age dummies for each year to estimate age profiles), which indicate the treatment assignment and are interacted with the post-reform year 2007 in our DiD design.

Table 4.4 shows the regression results for the age treatment dummy, i.e., individuals younger than 45 years are the control group, and individuals equal to or older than 45 years are the treatment group. The general age trends (age≥45) indicate that individuals equal to or older than 45 years have on average about 17.8 more days in UB-1, 6.5 fewer days in UB-2, and 25.3 fewer days in WORK than individuals who are younger than 45 years, which supports the view that older workers perform worse in the labor market. The general time trends (year2007) show that individuals in 2007 have on average about 6.7 fewer days in UB-1, 5 fewer days in UB-2, and 10.6 more days in WORK than individuals in 2005, which might be driven by the overall labor market reforms induced by AGENDA 2010. The treatment effects (age≥45×year2007) of the reduction of the potential duration of UB-1 indicate on average about 10.5 fewer days in UB-1, 4.7 more days in UB-2, and 13.6 more days in WORK. The treatment effects are statistically significant and sizeable. If we put the absolute treatment effects simply in relationship to the sample mean outcomes, days in UB-1 have decreased by about 25%, days in UB-2 have increased by about 75%, and days in WORK have increased by about 5%. It should be kept in mind that the estimates in this study may not provide the true treatment effect for the entire eligible population. We argue that our estimated treatment effects for UB-1 represent the lower bounds in absolute value, for UB-2 the upper bounds, and for WORK the lower bounds for the average treatment effects in the total population. Also note that the treatment effects on our three outcome variables do not sum up to zero, because our three outcome variables do not exhaust all possible labor market states. In particular, we focus only on employment subject to social insurance contributions, excluding self-employment, minor employment, and civil service. Moreover, we exclude non-employment such as retirement due to health reasons or other labor market exits.

Table 4.4: DiD Results for Age Treatment Dummy (Sample A: 2005/2007, Employed before Rehabilitation)

	UB-1 (1)	UB-2 (2)	WORK (3)
age≥45	17.80***	-6.51***	-25.29***
	[0.97]	[0.59]	[1.57]
year2007	-6.66***	-4.96***	10.56***
	[1.01]	[0.67]	[1.79]
age≥45×year2007(post-reform)	-10.50***	4.65***	13.57***
	[1.22]	[0.72]	[2.06]
Control variables	Yes	Yes	Yes
R^2	0.11	0.07	0.19
Mean dep. variable	39.58	6.15	261.68
N	94,990	94,990	94,990

Notes: Sample A (2005/2007, employed before rehabilitation). Outcome variables are days per calendar year. OLS regressions. Robust standard errors in brackets. * p<0.05; ** p<0.01; *** p<0.001.

In the next step, we replace the age treatment dummy with age treatment categories, i.e., we split the treatment group into age categories according to the different reductions of the potential duration of UB-1 induced by the reform. The results in Table 4.5 support the previous findings. Older workers have on average more days in UB-1, fewer days in UB-2, and fewer days in WORK. Days in UB-1 and UB-2 are lower, and days in WORK are larger in 2007 than in 2005. More importantly, the treatment effects have the same signs as before. The different age categories allow us to further analyze how far the treatment effects differ within the treatment group of older workers. The reference group is the control group consisting of individuals younger than 45 years. The treatment effects are four fewer days in UB-1 for the age group 45 to 46, which experienced a reduction of potential duration of UB-1 by six months. For the age group 47 to 51 (reduction by 10 months), the treatment effect is eight fewer days in UB-1. The age groups 52 to 54 (reduction by 14 months) and 55 to 56 (reduction by eight months) each have a treatment effect of about nine fewer days in UB-1. The largest treatment effect is estimated for the age group older than 56 (reduction by 14 months), which has about 14 fewer days in UB-1. The treatment effects on days in UB-2 do not differ that strongly and range be-tween three more days in UB-2 for the youngest treatment age group (45-46) and six more days in UB-2 for the oldest age treatment group (age>56). Days in WORK have increased for the age group 45 to 46 by eight days, for the age group 47 to 51 by six days, for the age group 52 to 54 by 13 days, for the age group 55 to 56 by 16 days, and for the age group older than 56 by 17 days after the reform. Overall, we can conclude that the treatment effects on days in UB-1, days in UB-2, and days in WORK are significant for all age treatment categories and that the absolute treatment effects are larger for older individuals.

Table 4.5: DiD Results for Age Treatment Categories (Sample A: 2005/2007, Employed before Rehabilitation)

	UB-1 (1)	UB-2 (2)	WORK (3)
age 45-46	4.00*	-3.08**	-5.32*
	[1.58]	[0.95]	[2.62]
age 47-51	9.74***	-5.55***	-4.58*
	[1.24]	[0.70]	[1.95]
age 52-54	13.74***	-6.32***	-15.94***
	[1.46]	[0.73]	[2.23]
age 55-56	18.08***	-7.08***	-32.55***
	[1.69]	[0.78]	[2.53]
age>56	35.45***	-8.90***	-60.08***
	[1.40]	[0.63]	[2.03]
year2007	-6.61***	-4.97***	10.46***
	[1.01]	[0.67]	[1.79]
age 45-46 (reduction -6 months)×year2007 (post-reform)	-4.44*	2.56*	7.72*
	[2.00]	[1.19]	[3.47]
age 47-51 (reduction -10 months)×year2007 (post-reform)	-7.87***	4.30***	5.88*
	[1.54]	[0.87]	[2.55]
age 52-54 (reduction -14 months)×year2007 (post-reform)	-9.28***	4.04***	12.68***
	[1.79]	[0.90]	[2.91]
age 55-56 (reduction -8 months)×year2007 (post-reform)	-9.04***	4.86***	15.85***
	[2.12]	[0.99]	[3.36]
age>56 (reduction -14 months)×year2007 (post-reform)	-14.22***	5.75***	17.16***
	[1.78]	[0.79]	[2.70]
Control variables	Yes	Yes	Yes
R^2	0.12	0.07	0.20
Mean dep. variable	39.58	6.15	261.68
N	94,99	94,99	94,99

Notes: Sample A (2005/2007, employed before rehabilitation). Outcome variables are days per calendar year. OLS regressions. Robust standard errors in brackets. * p<0.05; ** p<0.01; *** p<0.001.

Finally, we use age dummies for each year that are interacted with the post-reform year 2007. This specification allows us to estimate and predict completely non-linear age profiles. Figures 4.1 to 4.3 present the predicted outcomes, and Figure 4.4 summarizes the differences between 2007 and 2005 for each year of age. It can easily be seen for days in UB-1 and days in WORK that the differences between 2007 and 2005 are larger for the treatment groups than for the control groups, i.e., the treatment leads to a reduction of days in UB-1 and an increase of days in WORK. For days in UB-2, the differences between 2007 and 2005 are close to zero for the treatment groups and negative for the control groups, i.e., the treatment leads to an increase of days in UB-2. A further inspection of the non-linear age profiles for days in UB-1 and days in UB-2 reveals that we can not measure a significant treatment effect for the age threshold at 45 years. As mentioned, we have a measurement problem at the age threshold values because we only have age in years and not in months or even in days in our data. Thus, individuals turn 45 in the year 2007 and are not affected by the reform for the complete calendar year.

Figure 4.1: Age Profiles UB-1 (Sample A: 2005/2007, Employed before Rehabilitation) (SE and CI by delta method)

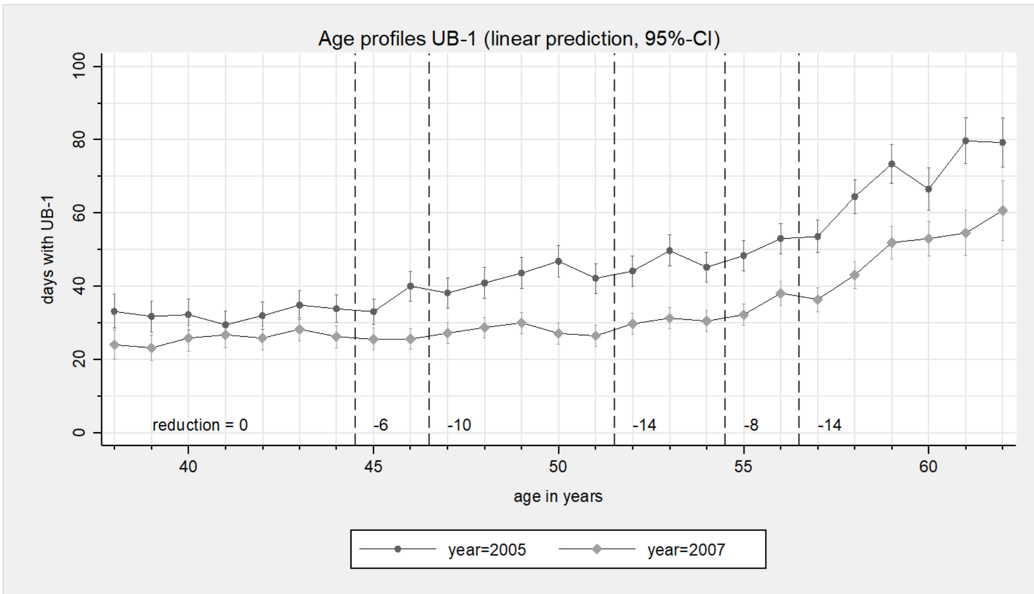


Figure 4.2: Age Profiles UB-2 (Sample A: 2005/2007, Employed before Rehabilitation) (SE and CI by delta method)

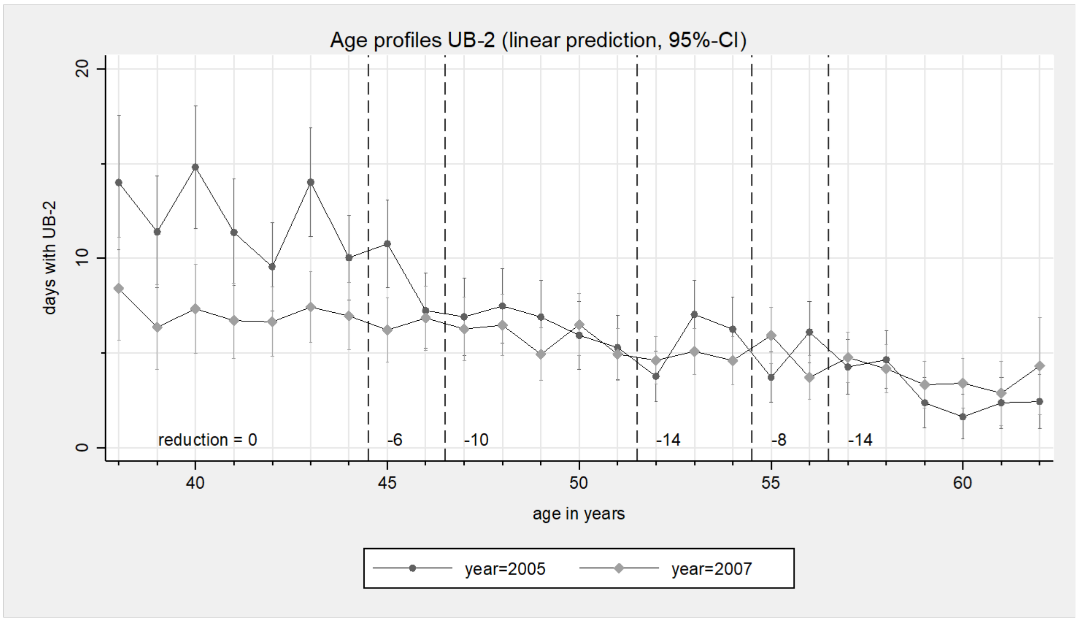
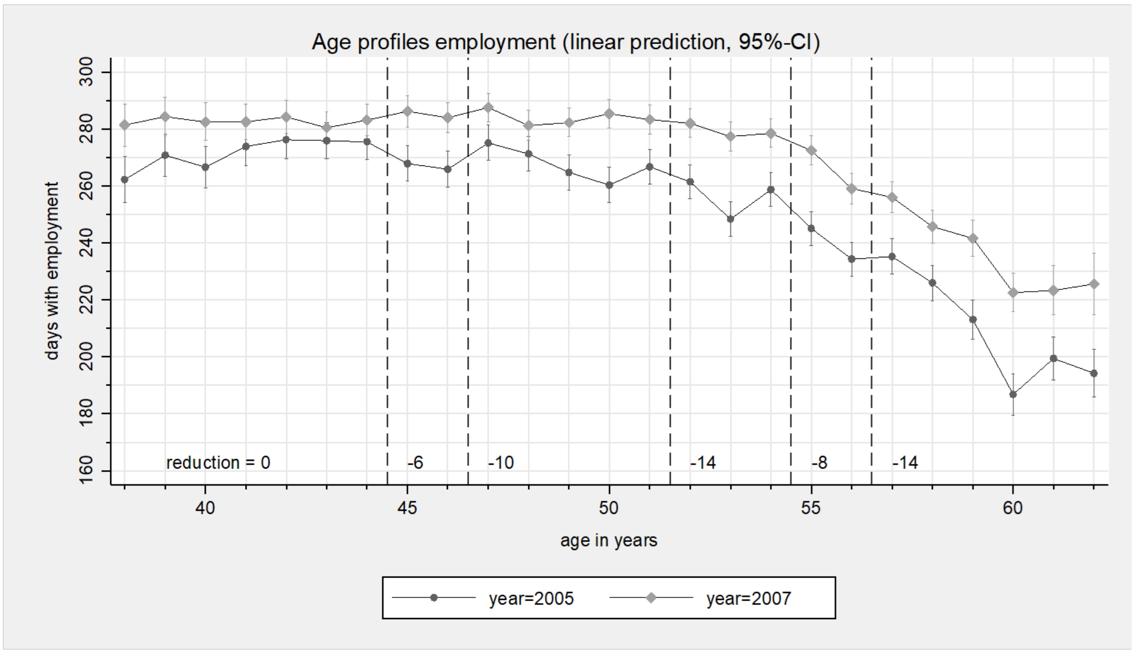


Figure 4.3: Age Profiles Employment (Sample A: 2005/2007, Employed before Rehabilitation) (SE and CI by delta method)



4.6.2 Pre-reform and Post-reform Trends for 2004-2009

Crucial for a causal interpretation of treatment effects in DiD designs is the parallel (common) trend assumption, i.e., the time trends must not differ between control groups (age<45) and treatment groups (age≥45) in the absence of the reform to estimate unbiased treatment effects, which are the coefficients of the interaction terms between treatment groups and post-reform years. To check the pre-reform and post-reform trends, we repeat the analysis for the years 2005 and 2007 (Sample A, employed before rehabilitation) from the previous section with the full set of years from 2004 to 2009 (Sample B, employed before rehabilitation). Table 4.6 shows that the treatment effects are only observed in the post-reform years and not in the pre-reform years, which supports the parallel trend assumption. More specifically, the coefficients of the interaction terms between age≥45 and the years 2005 and 2006 do not differ significantly from zero and from the reference year 2004, whereas the coefficients of the interaction terms between age≥45 and the post-reform years 2007, 2008, and 2009 differ significantly from zero and from the pre-reform years 2004, 2005, and 2006.[2] The more detailed age profiles in Figures 4.5 to 4.7 also support the parallel trend assumption. Because we do not find evidence for a violation of the parallel trend assumption and can identify a structural break between pre-reform years (2004-2006) and post-reform years (2007-2009), we are confident that our estimated

[2]Although the coefficient of the interaction term between age≥45 and year 2006 is positive and statistically significant at p<0.01, which might indicate some anticipation effect of the reform, it is small in magnitude (3.4 days) and does not affect our main results from the preferred specification (sample A, 2005/2007).

treatment effects are not a statistical artifact.

Figure 4.4: Differences 2007-2005 for Each Age Group (Sample A: 2005/2007, Employed before Rehabilitation) (SE and CI by delta method)

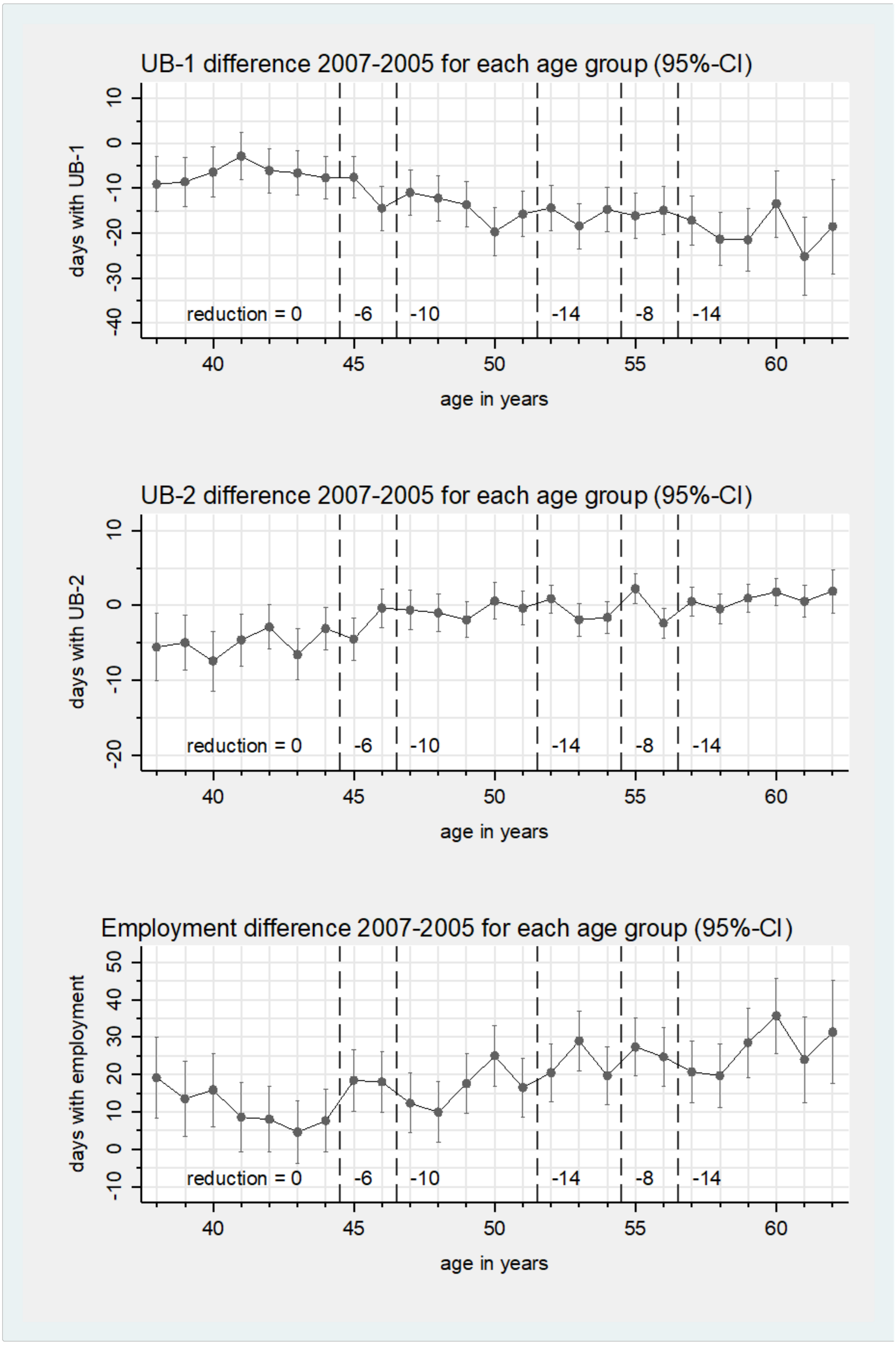


Table 4.6: DiD Results and Trends for Age Treatment Dummy (Sample B: 2004-2009, Employed before Rehabilitation)

	UB-1 (1)	UB-2 (2)	WORK (3)
age≥45	17.52***	-5.39***	-28.72***
	[0.94]	[0.43]	[1.48]
year2005	-3.73***	4.08***	4.84**
	[1.09]	[0.68]	[1.84]
year2006	-8.17***	5.27***	12.26***
	[1.04]	[0.69]	[1.79]
year2007	-10.34***	-0.98	15.33***
	[1.01]	[0.56]	[1.75]
year2008	-9.31***	-2.81***	15.48***
	[1.01]	[0.51]	[1.74]
year2009	-6.31***	-1.85***	9.06***
	[1.04]	[0.54]	[1.76]
age≥45×year2005	0.38	-1.27	3.61
	[1.33]	[0.72]	[2.13]
age≥45×year2006	3.35**	-1.05	4.61*
	[1.29]	[0.74]	[2.07]
age≥45×year2007 (post-reform)	-10.14***	3.43***	17.17***
	[1.21]	[0.61]	[2.01]
age≥45×year2008 (post-reform)	-10.99***	4.38***	19.74***
	[1.20]	[0.55]	[2.00]
age≥45×year2009 (post-reform)	-12.61***	3.61***	23.25***
	[1.22]	[0.58]	[2.00]
Control variables	Yes	Yes	Yes
R^2	0.11	0.06	0.18
Mean dep. variable	40.47	5.51	261.43
N	306,23	306,23	306,23

Notes: Sample B (2004-2009, employed before rehabilitation). Outcome variables are days per calendar year. OLS regressions. Robust standard errors in brackets. * p<0.05; ** p<0.01; *** p<0.001.

Figure 4.5: Age Profiles UB-1 (Sample B: 2004-2009, Employed before Rehabilitation) (SE and CI by delta method)

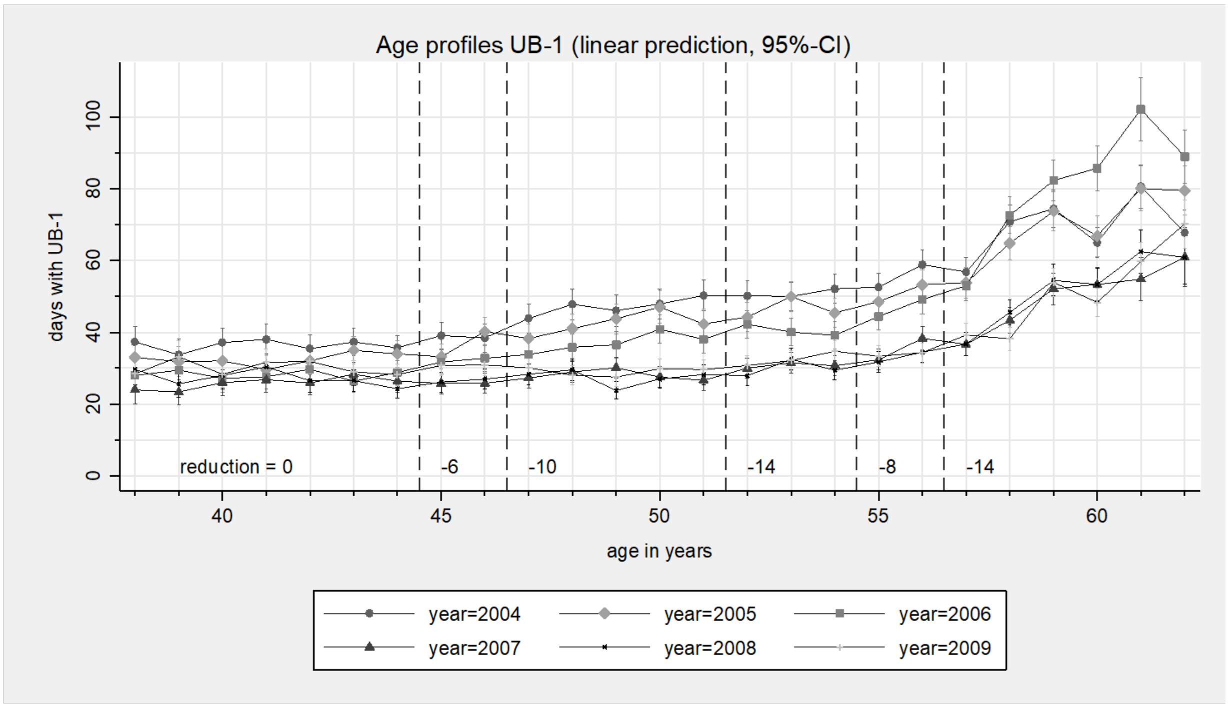


Figure 4.6: Age Profiles UB-2 (Sample B: 2004-2009, Employed before Rehabilitation) (SE and CI by delta method)

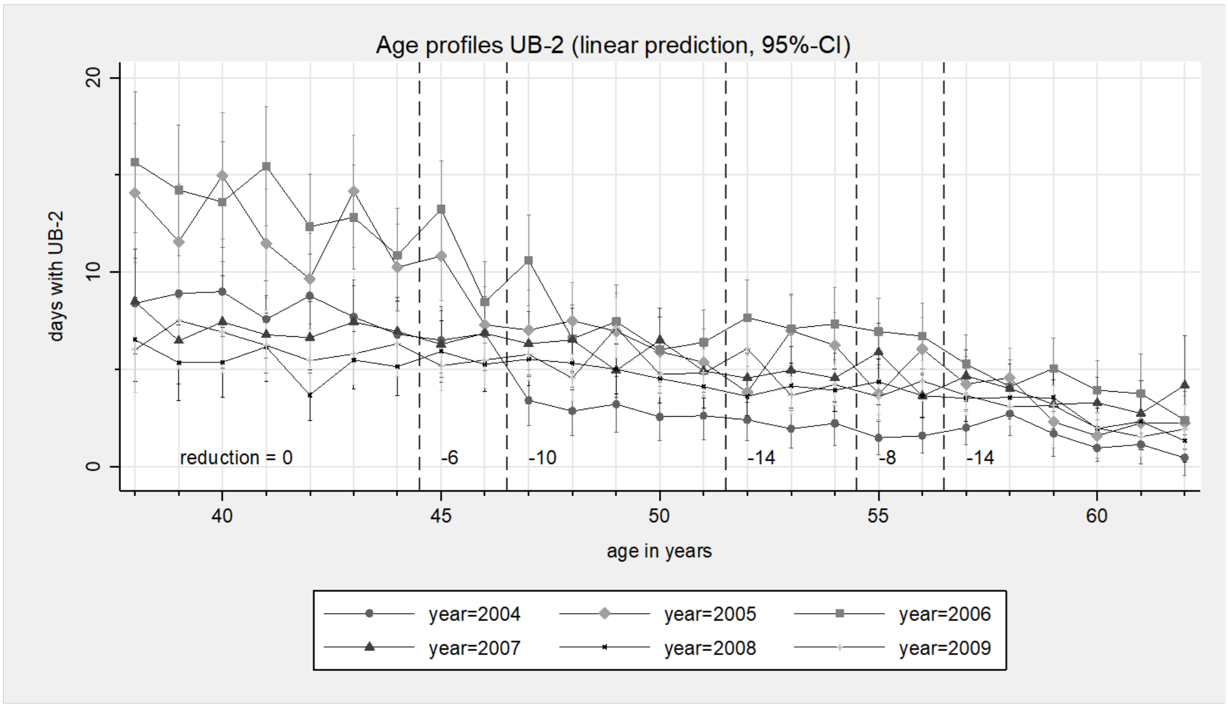


Figure 4.7: Age Profiles Employment (Sample B: 2004-2009, Employed before Rehabilitation) (SE and CI by delta method)

4.6.3 Gender and Regional Differences for 2005/2007

To check for potential gender and regional differences, we split our preferred sample A (2005/2007, employed before rehabilitation) between men and women (see Table 4.7) as well as between West (including Berlin) and East German Federal States (see Table 4.8). The results for the separate samples support our previous findings for the complete sample in Section 6.1. The small gender and regional differences indicate, if anything, that the reform affected women and people living in East German Federal States slightly more positively, i.e., days in UB-1 decreased and days in WORK increased even more, whereas days in UB-2 did not increase that much.[3]

[3]Estimates from the models with a triple interaction term (age$\geq$45$\times$year2007$\times$female and age$\geq$45$\times$year2007$\times$East, respectively) suggest that these differences are statistically insignificant.

Table 4.7: DiD Results Men vs. Women for Age Treatment Dummy (Sample A: 2005/2007, Employed before Rehabilitation)

	Men			Women		
	UB-1	UB-2	WORK	UB-1	UB-2	WORK
	(1)	(2)	(3)	(1)	(2)	(3)
age≥45	18.21***	-6.89***	-28.79***	17.35***	-6.15***	-20.65***
	[1.32]	[0.79]	[2.11]	[1.44]	[0.88]	[2.36]
year2007	-7.96***	-5.44***	13.85***	-4.91**	-4.50***	6.15*
	[1.36]	[0.90]	[2.37]	[1.53]	[1.01]	[2.70]
age≥45×year2007	-8.76***	5.70***	10.13***	-12.76***	3.48**	17.77***
	[1.64]	[0.97]	[2.76]	[1.83]	[1.09]	[3.10]
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.12	0.08	0.20	0.11	0.07	0.18
Mean dep. variable	39.66	6.14	260.02	39.49	6.16	263.59
N	50,903	50,903	50,903	44,087	44,087	44,087

Notes: Sample A (2005/2007, employed before rehabilitation). Outcome variables are days per calendar year. OLS regressions. Robust standard errors in brackets. * p<0.05; ** p<0.01; *** p<0.001.

Table 4.8: DiD Results West vs. East for Age Treatment Dummy (Sample A: 2005/2007, Employed before Rehabilitation)

	West			East		
	UB-1	UB-2	WORK	UB-1	UB-2	WORK
	(1)	(2)	(3)	(1)	(2)	(3)
age≥45	16.60***	-6.74***	-23.53***	23.44***	-5.32***	-33.36***
	[1.06]	[0.63]	[1.71]	[2.44]	[1.60]	[3.95]
year2007	-6.48***	-4.48***	10.29***	-7.72**	-7.58***	11.73**
	[1.11]	[0.73]	[1.95]	[2.52]	[1.73]	[4.52]
age≥45×year2007	-10.15***	4.85***	12.68***	-12.07***	3.78*	18.31***
	[1.33]	[0.78]	[2.25]	[3.06]	[1.87]	[5.18]
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.11	0.08	0.18	0.14	0.06	0.22
Mean dep. variable	38.73	5.92	262.93	43.83	7.31	255.43
N	79,098	79,098	79,098	15,892	15,892	15,892

Notes: Sample A (2005/2007, employed before rehabilitation). Outcome variables are days per calendar year. OLS regressions. Robust standard errors in brackets. * p<0.05; ** p<0.01; *** p<0.001.

4.6.4 Unemployed and Non-employed before Rehabilitation for 2005/2007

In the previous sections, we focused on individuals who were employed before medical rehabilitation. These individuals make up the majority of the complete sample. But there might al-so be a positive selection, and their decisions to enter unemployment are likely to be more important than their decisions to exit unemployment for our analyzed labor market outcome variables. To give a more complete picture of the labor market reform, we repeat our previous analyses for individuals who were unemployed and non-employed before the medical rehabilitation (sample C, 2005/2007). The overall effects are less positive for these unemployed and non-employed samples than for individuals who were employed before medical rehabilitation. Table 4.9 shows the average treatment effects for the unemployed sample of about six fewer days in UB-1, 10 more days in UB-2, and two fewer days in WORK, of which only the estimated treatment effect for days in UB-1 is statistically significant at $p<0.05$. The average treatment effects for the non-employed sample indicate about 10 fewer days in UB-1, 19 more days in UB-2, and 13 fewer days in WORK. Thus, the decreased days in UB-1 are largely due to a slip into UB-2 and the reform seems to even have a negative effect on WORK in the un-employed and non-employed samples.

Table 4.9: DiD Results Unemployed and Non-employed before Rehabilitation for Age Treatment Dummy (Sample C: 2005/2007)

	Unemployed			Non-employed		
	UB-1	UB-2	WORK	UB-1	UB-2	WORK
	(1)	(2)	(3)	(1)	(2)	(3)
age≥45	31.34***	-26.95***	-14.55***	20.83***	-23.86***	0.34
	[2.25]	[3.88]	[2.45]	[2.74]	[3.36]	[3.96]
year2007	-17.17***	14.30**	20.39***	-13.94***	-1.79	30.52***
	[2.20]	[4.84]	[3.41]	[2.67]	[3.78]	[4.46]
age≥45×year2007	-5.94*	9.62	-2.05	-9.93**	18.98***	-12.67*
	[2.93]	[5.53]	[3.77]	[3.27]	[4.21]	[5.07]
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.12	0.23	0.20	0.07	0.27	0.35
Mean dep. variable	55.93	159.31	42.86	47.47	61.23	146.31
N	15,857	15,857	15,857	16,529	16,529	16,529

Notes: Sample C (2005/2007, unemployed and non-employed before rehabilitation). Outcome variables are days per calendar year. OLS regressions. Robust standard errors in brackets. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

4.6.5 Treatment Intensity for 2005/2007

We extend the standard DiD strategy for our preferred sample A (2005/2007, employed before rehabilitation) by replacing the interaction terms between age and the post-reform year 2007 with a treatment intensity variable that measures the UB-1 reduction in months for the different age groups. The treatment intensity variable follows the DiD strategy because it is in principle an interaction between treatment intensity and the post-reform year. Table 4.10 shows that treatment intensity is zero for all observations in 2005 and for all observations younger than 45 years in 2007, whereas treatment intensity is positive for all observations equal to or older than 45 years in 2007. Because treatment intensity is correlated with age, we estimate different specifications of age as controls (age, age^2, age^3, age dummies). The results in Table 4.11 show that the size of the treatment intensity effect is indeed smaller if non-linearity of age is taken into account. Overall, the results in Table 4.11 support our previous findings, but allow a different quantitative interpretation. The reduction in potential duration of UB-1 by one month de-creases UB-1 on average by more than 0.6 days per year, increases UB-2 on average by about 0.3 days per year, and increases WORK on average by more than 0.8 days per year. To rule out that our estimated treatment intensity effects are a statistical artifact, we perform a placebo test for the years 2004 and 2005 and act as if the UB-1 reduction in months would have occurred in 2005, although 2004 and 2005 are both pre-reform years. Because the estimated coefficients in the placebo test are either not significantly different from zero or even have the opposite sign than the estimated treatment intensity effects for 2007, we are confident that the treatment intensity effects are not a statistical artifact.

Table 4.10: Number of Observations in Treatment Intensity Variable UB-1 Reduction in Months

UB-1 reduction in months	2005 (pre-reform)	2007 (post-reform)	Total
0	45,703	11,128	56,831
6	0	4,238	4,238
8	0	4,972	4,972
10	0	11,174	11,174
14	0	17,775	17,775
Total	45,703	49,287	94,99

Table 4.11: DiD Results for Treatment Intensity (Sample A: 2005/2007, Employed before Rehabilitation) and Placebo Tests (2004/2005)

	Different specifications of age			
	age	age, age^2	age, age^2, age^3	age dummies
	(1)	(2)	(3)	(4)
Outcome: days UB-1				
UB-1 reduction in months	-0.96***	-0.65***	-0.60***	-0.81***
	[0.09]	[0.09]	[0.09]	[0.10]
R^2	0.12	0.12	0.12	0.12
Mean dep. variable	39.58	39.58	39.58	39.58
N	94,990	94,990	94,990	94,990
Coefficient placebo 2004/2005	-0.03	0.22*	0.23*	0.12
Outcome: days UB-2				
UB-1 reduction in months	0.24***	0.27***	0.26***	0.36***
	[0.04]	[0.05]	[0.05]	[0.05]
R^2	0.07	0.07	0.07	0.07
Mean dep. variable	6.15	6.15	6.15	6.15
N	94,990	94,990	94,990	94,990
Coefficient placebo 2004/2005	-0.18***	-0.14**	-0.14**	-0.12*
Outcome: days WORK				
UB-1 reduction in months	1.74***	0.86***	0.81***	0.95***
	[0.14]	[0.15]	[0.15]	[0.16]
R^2	0.20	0.20	0.20	0.20
Mean dep. variable	261.68	261.68	261.68	261.68
N	94,990	94,990	94,990	94,990
Coefficient placebo 2004/2005	1.08***	0.29	0.28	0.15

Notes: Sample A (2005/2007, employed before rehabilitation). Placebo tests for years 2004/2005 (N= 97,513) as if UB-1 reduction in months would have occurred in 2005. All control variables included. Outcome variables are days per calendar year. OLS regressions. Robust standard errors in brackets. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

4.7 Discussion and Concluding Remarks

This study offers a comprehensive evaluation of the major German reform of regular unemployment benefits (unemployment benefits 1, UB-1), enacted in February 2006 to re-establish incentives to work and delay withdrawals from the labor market. The policy change induced a substantial reduction in the potential benefits duration for older workers, thus alleviating the disincentive effect of long compensation provided by the UI system. Our estimation results, based on a difference-in-differences approach, reveal partially positive effects of the reform and, in line with the analyses of similar reforms in the European context, suggest the need for more complete evaluations of policy changes in general (Fremigacci et al., 2010; Pellizzari, 2006).

We find evidence that individual labor market attachment matters, so does the value of the UI recipients' outside options (employment, non-employment, unemployment assistance or other social programs). In particular, for our preferred sample of individuals who were employed be-fore medical rehabilitation, we find that days in UB-1 decrease by 10.5 and days in employment increase by 13.6, which hints at a positive treatment effect. However, a deeper analysis indicates that days in unemployment assistance (unemployment benefits 2, UB-2), which is granted to unemployed jobseekers without temporal restrictions upon exhaustion of UB-1, in-crease by 4.7. Supplementary analysis of individuals who are less attached to the labor market (individuals who were unemployed or non-employed before medical rehabilitation) provides a more complete picture of the distribution of days in UB-1, days in UB-2, and days in employment. Although a positive impact on days in UB-1 is observed also for this population group, they perform worse on the labor market. In particular, days in UB-2 largely increase, while days in employment even decrease in response to the reform. These findings are consistent with the recent work on effectiveness of more comprehensive reforms of labor market institutions as opposed to one policy reform at a time. Indeed, from the labor market and social policy perspective, transitions to UB-2 upon exhaustion of UB-1 denote an unintended consequence of the reform, limiting the success of a policy change that aims to increase labor supply via reductions in the generosity of the UI system. It is noteworthy to mention that, contrary to the middle-run framework adopted in our work, a long-run perspective of this re-form evaluation might lead to more reassuring conclusions. In fact, a potential increase in the number of days in WORK registered in later years following the reform would encourage us to conclude that the worse outcome estimated in the middle-run, as indicated by an increase in days in UB-2, might be an expected intermediate outcome, which serves as an additional incentive for reintegration into the labor market. In other words, the negative effect on days in UB-2 and days in WORK, observed in the middle-run for those less attached to the labor market, might be smaller in the long-run. Unfortunately, our data do not allow us to investigate the dynamics of UB-1, UB-2, and WORK in the long-run to test this hypothesis.

Due to data construction, our study presents some limitations with respect to measurement ac-curacy of eligibility conditions to qualify for the potential benefits duration. In practice, the duration of benefits that an unemployed worker applying for unemployment compensation is entitled to is calculated from his or her work history over a reference period just prior to job separation. In our work, however, due to lack of information on age in months and work history of the unemployed at the date of unemployment entry, we measure their labor market attachment within entire calendar years. One might argue that this imprecision could compromise our results. Nevertheless, we are able to restrict our samples to approximate the fulfillment of eligibility criteria both under the old and the new regime. In line with Hunt (1995), who finds slightly larger coefficients when the treatment group is defined more accurately, our estimates should be interpreted as lower and upper bounds for the treatment effects in the total population. Furthermore, our sample consists of individuals with some health deficiency who participated in a medical rehabilitation program. We believe this aspect opens an interesting perspective in the reform evaluation, and positive effects of the reform for this population group that has gained more importance in recent years encourage us to predict even larger effects for healthy workers with no need for medical rehabilitation. Our expectation relies on several studies that focus on employment outcomes of work-limited workers and identify the underlying mechanisms of their adverse labor market prospects (Charles, 2003; Colella and Bruyère, 2011; Mok et al., 2008). Based on this literature, we expect individuals with no health impairment to be more responsive to the UI incentives, interpreting our estimates as lower and upper bounds for the treatment effects in the total population.

Our results hint at the importance to design labor market reforms in a wider framework of institutional interactions. In fact, a common feature of the European institutions is the interaction of unemployment insurance with other social security programs. Furthermore, institutional similarity in the European context may encourage policy-makers to learn from their neighbors' experience. In fact, only three years apart, the French government also enacted a major reform of the UI system that significantly shortened the potential benefits duration for older workers. Fremigacci et al. (2010) claims that the major effect of this policy change was to shift job seekers from regular unemployment benefits to unemployment assistance. When the objective of policy-makers is to discourage moral hazard behavior via shortening the duration of unemployment benefits, a broader consideration of labor market performance upon exhaustion of regular unemployment benefits seems appropriate. Thus, the unemployment benefits reform in Slovenia achieved the intended decline in moral hazard induced by the UI system because in-creased transitions from unemployment to new jobs did not occur at the expense of the quality of post-unemployment job matches (Van Ours and Vodopivec, 2006, 2008). On the contrary, the claim of abated moral hazard behavior of benefits' recipients might be unfounded if increased employment comes with a loss of job match quality in response to the reform and longer unemployment spells would have facilitated productive job searches. Furthermore, non-negligible exits from unemployment to non-employment as well as shifts from unemployment benefits to unemployment assistance or other welfare programs under-mine potentially positive effects on duration of regular unemployment benefits, calling into doubt the success of proposed mitigation of the disincentive effect of long compensation provided by the UI system.

References

Cahuc, P. and E. Lehmann (2000). Should unemployment benefits decrease with the unemployment spell? *Journal of Public Economics 77*(1), 135–153.

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*(4), 604–627.

Card, D., R. Chetty, and A. Weber (2007). The spike at benefit exhaustion: Leaving the unemployment system or starting a new job? *American Economic Review 97*(2), 113–118.

Card, D. and P. B. Levine (2000). Extended benefits and the duration of UI spells: Evidence from the New Jersey extended benefit program. *Journal of Public Economics 78*(1-2), 107–138.

Charles, K. K. (2003). The longitudinal structure of earnings losses among work-limited disabled workers. *Journal of Human Resources 38*(3), 618–646.

Chetty, R. (2008). Moral hazard versus liquidity and optimal unemployment insurance. *Journal of Political Economy 116*(2), 173–234.

Colella, A. J. and S. M. Bruyère (2011). Disability and employment: New directions for industrial and organizational psychology. In *APA handbook of industrial and organizational psychology, Vol 1: Building and developing the organization.*, pp. 473–503. American Psychological Association.

Deutsche Rentenversicherung (2012). Abgeschlossene Rehabilitation im Versicherungsverlauf 2002-2009. http://www.fdz-rv.de/FdzPortalWeb/dispcontent.do?id=main_fdz_forschung_laengsb&chmenu=ispvwNavEntriesByHierarchy34, last accessed on February 18, 2020.

Deutsche Rentenversicherung (2013). Completed medical rehabilitation in 2013. http://forschung.deutsche-rentenversicherung.de/ForschPortalWeb/contentAction.do?stataktID=B8E152EEF2EF07CBC1257CFA00283390&chstatakt_RehabilitationReha-Leistungen=WebPagesIIOP1430&open&viewName=statakt_RehabilitationReha-Leistungen#WebPagesIIOP1430, last accessed on February 18, 2020.

Deutsche Rentenversicherung (2014a). http://www.deutsche-rentenversicherung.de/Allgemein/de/Navigation/2_Rente_Reha/02_Rehabilitation/02_leistungen/01_medizinisch/medizinische_reha_node.html, last accessed on February 18, 2020.

Deutsche Rentenversicherung (2014b). Approved applications for medical rehabilitation. http://forschung.deutsche-rentenversicherung.de/ForschPortalWeb/contentAction.do?statzrID=DC13BBF15050174CC1256F2A00307C6A&chstatzr_Rehabilitation=WebPagesIIOP62&open&viewName=statzr_Rehabilitation#WebPagesIIOP62, last accessed on February 18, 2020.

Dietz, M. and U. Walwei (2011). Germany—No country for old workers? *Zeitschrift für Arbeitsmarktforschung 44*(4), 363–376.

Dlugosz, S., G. Stephan, and R. A. Wilke (2014). Fixing the leak: Unemployment incidence before and after a major reform of unemployment benefits in Germany. *German Economic Review 15*(3), 329–352.

Feldstein, M. (2005). Rethinking social insurance. *American Economic Review 95*(1), 1–24.

Fremigacci, F. et al. (2010). Maximum benefits duration and older workers' transitions out of unemployment: A regression discontinuity approach. *Document de recherche EPEE, Centre d'études des politiques économiques de l'université d'Évry* (10-12), 56.

Hopenhayn, H. A. and J. P. Nicolini (1997). Optimal unemployment insurance. *Journal of Political Economy 105*(2), 412–438.

Hunt, J. (1995). The effect of unemployment compensation on unemployment duration in Germany. *Journal of Labor Economics 13*(1), 88–120.

Hurd, M. (1980). A compensation measure of the cost of unemployment to the unemployed. *The Quarterly Journal of Economics 95*(2), 225–243.

Katz, L. and B. Meyer (1990). The impact of the potential duration of unemployment benefits on the duration of unemployment. *Journal of Public Economics 41*(1), 45–72.

Kyyrä, T. and V. Ollikainen (2008). To search or not to search? The effects of UI benefit extension for the older unemployed. *Journal of Public Economics 92*(10-11), 2048–2070.

Lalive, R. (2007). Unemployment benefits, unemployment duration, and post-unemployment jobs: A regression discontinuity approach. *American Economic Review 97*(2), 108–112.

Lalive, R., J. Van Ours, and J. Zweimüller (2006). How changes in financial incentives affect the duration of unemployment. *The Review of Economic Studies 73*(4), 1009–1038.

Le Barbanchon, T. (2016). The effect of the potential duration of unemployment benefits on unemployment exits to work and match quality in France. *Labour Economics 42*, 16–29.

Lichter, A. (2016). Benefit duration and job search effort: Evidence from a natural experiment. *IZA Discussion Paper 10264, Institute of Labor Economics*.

Lo, S. M., G. Stephan, and R. A. Wilke (2017). Competing risks copula models for unemployment duration: An application to a German Hartz reform. *Journal of Econometric Methods 6*(1). doi:https://doi.org/10.1515/jem-2015-0005.

Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica 58*, 757–782.

Moffitt, R. and W. Nicholson (1982). The effect of unemployment insurance on unemployment: The case of federal supplemental benefits. *The Review of Economics and Statistics 64*(1), 1–11.

Mok, W. K., B. D. Meyer, K. K. Charles, and A. C. Achen (2008). A note on "The longitudinal structure of earnings losses among work-limited disabled workers". *Journal of Human Resources 43*(3), 721–728.

Mortensen, D. T. (1970). Job search, the duration of unemployment, and the Phillips curve. *The American Economic Review 60*(5), 847–862.

Mortensen, D. T. (1977). Unemployment insurance and job search decisions. *ILR Review 30*(4), 505–517.

Nekoei, A. and A. Weber (2017). Does extending unemployment benefits improve job quality? *American Economic Review 107*(2), 527–61.

Pellizzari, M. (2006). Unemployment duration and the interactions between unemployment insurance and social assistance. *Labour Economics 13*(6), 773–798.

Schmieder, J. F. and T. Von Wachter (2016). The effects of unemployment insurance benefits: New evidence and interpretation. *Annual Review of Economics 8*, 547–581.

Schmieder, J. F., T. Von Wachter, and S. Bender (2012a). The long-term effects of UI extensions on employment. *American Economic Review 102*(3), 514–19.

Schmieder, J. F., T. Von Wachter, and S. Bender (2012b). The effects of extended unemployment insurance over the business cycle: Evidence from regression discontinuity estimates over 20 years. *The Quarterly Journal of Economics 127*(2), 701–752.

Schmieder, J. F., T. von Wachter, and S. Bender (2016). The effect of unemployment benefits and nonemployment durations on wages. *American Economic Review 106*(3), 739–77.

Shavell, S. and L. Weiss (1979). The optimal payment of unemployment insurance benefits over time. *Journal of Political Economy 87*(6), 1347–1362.

Van Ours, J. C. and M. Vodopivec (2006). How shortening the potential duration of unemployment benefits affects the duration of unemployment: Evidence from a natural experiment. *Journal of Labor Economics 24*(2), 351–378.

Van Ours, J. C. and M. Vodopivec (2008). Does reducing unemployment insurance generosity reduce job match quality? *Journal of Public Economics 92*(3-4), 684–695.

Winter-Ebmer, R. (2003). Benefit duration and unemployment entry: A quasi-experiment in Austria. *European Economic Review 47*(2), 259–273.

Appendix 4.A

Table 4.A1: Number of Observations in Year-Age Cells (Employed before Rehabilitation)

age	2004	2005	2006	2007	2008	2009	Total
38	1,375	1,140	1,206	1,150	1,125	1,064	7,060
39	1,399	1,272	1,282	1,352	1,28	1,293	7,878
40	1,575	1,393	1,456	1,385	1,435	1,524	8,768
41	1,712	1,441	1,614	1,645	1,679	1,677	9,768
42	1,724	1,609	1,679	1,711	1,847	1,796	10,366
43	1,883	1,683	1,795	1,922	1,974	2,111	11,368
44	1,930	1,765	1,870	1,963	2,194	2,177	11,899
45	1,981	1,812	2,014	2,013	2,164	2,369	12,353
46	1,949	1,888	1,958	2,225	2,275	2,506	12,801
47	2,070	1,834	2,084	2,163	2,389	2,593	13,133
48	2,065	1,825	2,106	2,250	2,439	2,752	13,437
49	1,928	1,927	2,166	2,273	2,633	2,763	13,690
50	2,121	1,889	2,138	2,221	2,419	2,880	13,668
51	2,164	1,985	2,164	2,267	2,583	2,748	13,911
52	2,346	2,117	2,271	2,495	2,620	2,864	14,713
53	2,447	2,178	2,296	2,508	2,609	2,995	15,033
54	2,505	2,202	2,328	2,384	2,674	2,811	14,904
55	2,677	2,330	2,407	2,503	2,597	2,900	15,414
56	2,641	2,465	2,473	2,469	2,683	2,904	15,635
57	2,568	2,181	2,448	2,483	2,611	2,805	15,096
58	2,483	2,244	2,315	2,283	2,506	2,587	14,418
59	2,178	1,969	2,148	1,983	2,233	2,595	13,106
60	2,802	1,661	1,825	1,817	1,949	2,289	12,343
61	1,880	1,570	1,084	1,116	1,338	1,619	8,607
62	1,407	1,323	1,307	706	938	1,180	6,861
Total	51,810	45,703	48,434	49,287	53,194	57,802	306,230

Table 4.A2: Summary Statistics (Sample A: 2005/2007, Employed before Rehabilitation)

	Mean	Std. dev.	Min	Max
OUTCOME VARIABLES (Y)				
days UB-1 in calendar year (UB-1)	39.5784	93.0501	0	365
days UB-2 in calendar year (UB-2)	6.1490	40.7390	0	365
days employed in calendar year (WORK)	261.6766	151.7951	0	365
DiD VARIABLES (AGE, YEARS)				
age in years	50.4924	6.5965	38	62

age≥45 (affected by reform)	0.7744	0.4180	0	1
year2005 (pre-reform)	0.4811	0.4996	0	1
year2007 (post-reform)	0.5189	0.4996	0	1
CONTROL VARIABLES (X)				
female (dummy)	0.4641	0.4987	0	1
MARITALSTATUS (dummies)				
Single (reference group)	0.1254	0.3312	0	1
Married	0.7147	0.4516	0	1
Divorced	0.1312	0.3376	0	1
Widowed	0.0287	0.1671	0	1
NATIONALITY (dummies)				
Germany (reference group)	0.9429	0.2321	0	1
Italy, Spain, Greece, Portugal	0.0108	0.1036	0	1
Former Yugoslavia	0.0122	0.1096	0	1
Turkey	0.0145	0.1196	0	1
Other EU and non-EU country	0.0167	0.1281	0	1
Stateless, unknown	0.0029	0.0541	0	1
EDUCATION (dummies)				
Unknown, not applicable (reference group)	0.1747	0.3797	0	1
Low/ medium secondary schooling degree without apprenticeship	0.1288	0.3350	0	1
Low/ medium secondary schooling degree with apprenticeship	0.6062	0.4886	0	1
High secondary schooling degree without apprenticeship	0.0039	0.0622	0	1
High secondary schooling degree with apprenticeship	0.0261	0.1594	0	1
University of Applied Science degree	0.0292	0.1684	0	1
University degree	0.0311	0.1735	0	1
JOBPOSITION (dummies)				
Unknown, not applicable (reference group)	0.0034	0.0579	0	1
Apprentice	0.0006	0.0249	0	1
Unskilled blue-collar worker	0.1076	0.3099	0	1
Low skilled blue-collar worker	0.1020	0.3027	0	1
Skilled blue-collar worker	0.2724	0.4452	0	1
Master craftsman, foreman	0.0132	0.1143	0	1
White-collar worker	0.4955	0.5000	0	1
Civil servant	0.0004	0.0195	0	1
Self-employed	0.0047	0.0687	0	1
OCCUPATION (dummies)				
Unknown, not applicable (reference group)	0.0621	0.2413	0	1

Agriculture, forestry and fishing	0.0124	0.1106	0	1
Mining and quarrying	0.0036	0.0595	0	1
Manufacturing	0.0332	0.1791	0	1
Metal-making and metal-working	0.1236	0.3292	0	1
Textile-making and textile-processing	0.0056	0.0745	0	1
Accommodation and food service activities	0.0255	0.1576	0	1
Construction	0.0885	0.2841	0	1
Professional, scientific and technical activities	0.0555	0.2290	0	1
Trade and transportation	0.1846	0.3880	0	1
Administrative and support service activities	0.2059	0.4043	0	1
Health care	0.1144	0.3183	0	1
Teaching and training	0.0244	0.1544	0	1
Other	0.0607	0.2387	0	1
FEDERAL STATE (dummies)				
Berlin (reference group)	0.0428	0.2024	0	1
Schleswig Holstein	0.0287	0.1669	0	1
Hamburg	0.0159	0.1252	0	1
Lower Saxony	0.0998	0.2997	0	1
Bremen	0.0060	0.0775	0	1
Northrhine-Westphalia	0.2032	0.4024	0	1
Hesse	0.0740	0.2617	0	1
Rhineland Palatinate	0.0471	0.2118	0	1
Baden-Wurttemberg	0.1419	0.3489	0	1
Bavaria	0.1596	0.3663	0	1
Saarland	0.0137	0.1162	0	1
Brandenburg	0.0319	0.1758	0	1
Mecklenburg-West Pomerania	0.0222	0.1473	0	1
Saxony	0.0537	0.2254	0	1
Saxony-Anhalt	0.0260	0.1590	0	1
Thuringia	0.0336	0.1801	0	1
REHABILITATION DIAGNOSIS (dummies)				
166 medical diagnoses				

Notes: Sample A (2005/2007, employed before rehabilitation). Number of observations N=94,990.

Authors' Contributions

Unemployment Benefits Duration and Labor Market Outcomes: Evidence from a Natural Experiment in Germany

Inna Petrunyk and Christian Pfeifer

Authors' contribution:

Joint contribution:

(i) conception and design of the work
(ii) choice of empirical methods
(iii) interpretation of statistical results
(iv) drafting the work

Inna Petrunyk's contribution (50%):

(i) data set preparation
(ii) literature review

Christian Pfeifer's contribution (50%):

(i) econometric analysis
(ii) results preparation

February, 17 2020

Inna Petrunyk

Christian Pfeifer

Chapter 5

How Older Workers Respond to Raised Early Retirement Age: Evidence from a Kink Design in Germany

Data replication statement:
The administrative data BASiD5109 employed in this study are weakly anonymous data obtained under a strict confidentiality agreement with the Research Institute of the Federal Employment Agency. Access to weakly anonymous data set is only possible via on-site use or through remote data execution. Statistical programs with all steps of data preparation and cleaning are available. Unfortunately, the data can not be provided as this would represent a violation of the confidentiality agreement. Nevertheless, in case of request, full assistance to anybody interested in getting the individual data from the the Research Institute of the Federal Employment Agency will be provided.

5.1 Introduction

Population aging across OECD countries substantially increases the financial burden on the public pension systems (OECD, 2017). To cope with the demographic change, many countries implemented reforms aimed at delaying withdrawals from the labor market, thus prolonging the contribution period and lowering the number of benefits recipients at the same time. Previous studies on the relationship between social security benefits and retirement find a sharp increase in labor market exits at the age of first eligibility for retirement benefits (Gruber and Wise, 2007). In line with this evidence, increase in the early retirement age is likely to be a promising policy instrument to boost employment of older people and ultimately alleviate the burden on the public pension system.

Official statistics signals that early retirement in Germany is non-negligible (Deutsche Rentenversicherung, 2019c). In this respect, in 2006 about 40% of all retirement benefits due to old-age pensions were paid at a reduced rate, with benefits claimed earlier by almost 38 months, on average. The monthly average pension deduction amounted to €114, lowering the average pension benefits level to €812 per month. Early retirement among claimants of disability pensions is even more pronounced. In fact, almost 95% of claimants received the benefits by about 35 months earlier, on average, with a monthly average pension deduction of €80 and an average benefits level of €630. To restore the financial equilibrium in Germany, characterized by poor labor market performance among the elderly, early retirement age for the old-age pension due to unemployment or partial retirement has been raised. The reform introduced in 2004 scheduled a gradual increase in the early retirement age from age 60 to 63. An increase in the early retirement age is likely to be effective in postponing retirement. However, the positive employment response might be limited if there are other attractive social insurance programs that may be used as exit routes from employment (Staubli and Zweimüller, 2013). Therefore, it is important to explore how the increased early retirement age, or, in other words, an increased distance to retirement age, affects the decision of older workers to enroll in other government programs, e.g. unemployment or disability insurance. Given that theoretical predictions about the impact of raising the early retirement age depend on model assumptions (Burtless, 1986; French, 2005; Gustman and Steinmeier, 1985, 2005; Mitchell and Phillips, 2000; Rust and Phelan, 1997), it is necessary to lean on empirical evidence based on exogenous variation.

In this paper, I study how postponement of entry into retirement through an increase in the early retirement age affects labor supply and retirement decisions as well as claims of alternative benefits in the German context. The German old-age social security system shares many institutional features, among which the option to retire at the early retirement age or through unemployment and disability insurance, with other countries. Insights about the implications of the pension reform in Germany can be useful in other contexts and are therefore of general interest.

Similar to previous research, this study exploits a cohort-specific variation in incentives to retire. At the same time, it differs from the existing literature in several important aspects. Most studies for Germany present evidence on the effects of a variation in pension incentives solely in the female population. For example, based on administrative data from the German Public Pension Insurance, Engels et al. (2017) examine the impact of a cohort-specific introduction of deductions for early retirement for women born after 1939. More recently, using the same data Geyer and Welteke (2019) analyze in a regression discontinuity design framework the effects of a large increase in the early retirement age resulting from an abolished early retirement program for women born after 1951. To the best of my knowledge, only Giesecke and Kind (2013) evaluate the pension reform studied in this paper. However, the authors restrict their attention exclusively to reemployment probability and early retirement behavior of older unemployed men. Moreover, they use survey data from the German Socio-Economic Panel and rely on a difference-in-differences design. This approach presents several disadvantages. Similar to Geyer and Welteke (2019) and Engels et al. (2017), it does not allow to examine the differential behavioral responses of men and women. Moreover, it is likely to suffer from lack of data accuracy typical of survey responses. This paper improves along these dimensions and provides novel empirical evidence on how older workers respond to a raised early retirement age.

Unlike the extensive research on factors affecting the timing of retirement (Manoli and Weber, 2016; Seibold, 2017; Ye, 2018), there is little work on how changes in the retirement age can affect labor force participation. I make four major contributions to the existing literature in this research area. First, the pension reform implemented a gradual increase in the early retirement age over multiple birth cohorts, thus providing a clean quasi-experimental setting. I exploit the kink in the early retirement age induced by the policy rule and apply a regression kink design to identify the causal effects of the reform. More precisely, the research design exploits the kinked schedule by which the reform increases the early retirement age and relates the increase in the early retirement age to changes in average labor market outcomes. In this way, the paper contributes to the strand of literature on labor force participation at older ages (Geyer and Welteke, 2019, among others) and the growing empirical evidence from regression kink designs (Card et al., 2015; Landais, 2015, among others).

Second, I adopt high-quality administrative data from the Research Institute of the Federal Employment Agency. The data provide rich information stored in episode format on employment biographies, benefits receipt history, retirement decisions as well as socio-demographic characteristics. These data present an advantage over the data from the Public Pension Insurance used in Engels et al. (2017) and Geyer and Welteke (2019) with information on several variables potentially not representative of entire employment histories (e.g. information only about the last occupation at the time of data collection). Among variables capturing individual characteristics, the data set adopted in this study provides information on birth month which determines the age of first eligibility for retirement benefits. The reported birth month allows to precisely define each individual's early retirement age, thus minimizing the measurement error in comparing labor market behavior of older people born close to the cutoff date.

Third, I contribute to the scarce research evidence on the labor market impact of an increase in the early retirement age, which differs from studying an increase in the normal retirement age (Duggan et al., 2007; Mastrobuoni, 2009). While an increase in the normal retirement age is comparable to a reduction in benefits, an increase in the early retirement age implies a reduction in the choice set of older people, encouraging them to enter retirement later or claim benefits from other government programs. Although there is a large literature on program substitution effects in response to changes in pension policy (Borghans et al., 2014; Karlström et al., 2008, among others), the existing evidence on program substitution in the context of an increase in the early retirement age is mixed (Atalay and Barrett, 2015; Oguzoglu et al., 2016). For example, Staubli and Zweimüller (2013) find evidence for large spillover effects on the unemployment insurance program and negligible effects on disability insurance claims in Austria. Studying the same reform, Manoli and Weber (2016) conclude instead that workers adjust to the increased early retirement age by keeping their jobs longer rather than substituting early retirement with unemployment benefits or disability pensions. This paper is intended to shed light on this controversial issue, by quantifying the causal effects of the reform on claims of unemployment and disability pension benefits.

Fourth, I explore the distributional and persistence effects of the reform which are of direct policy relevance. To this end, I perform a treatment heterogeneity analysis to assess whether behavioral responses vary across population groups. I first consider economic sector and past earnings as potential determinants of retirement decisions (Bound et al., 2010; Staubli, 2011). Then, I analyze whether behavioral responses of men and women differ. Finally, I investigate whether the reform mainly affects labor market behavior prior to reaching the early retirement age or its effects tend to persist also beyond the early retirement age.

Using a regression kink design, I estimate that a one-year increase in the early retirement age increases employment by 11.6 percentage points and reduces retirement by 15 percentage points. This indicates that a rise in the early retirement age is an effective policy tool to delay employment exits. I further do not detect any spillover effects on the unemployment or disability insurance program, which implies that workers with relatively long employment careers in my sample adjust to the increased early retirement age by delaying withdrawals from the labor market rather than seeking benefits from other sources. The treatment heterogeneity analysis suggests that, compared to individuals employed in service sector, workers in manufacturing sector are more likely to claim disability pension benefits, thus relying on the alternative social support program as an exit route from employment. Furthermore, the employment response among high-wage workers is more pronounced compared to that of low-wage workers, which is in line with the incentives. Women are barely affected by the pension reform, potentially relying on other available retirement options for women. Finally, I find weak evidence for positive employment effects persisting even beyond the early retirement age.

The reminder of the paper is as follows. Section 5.2 outlines the institutional background and the pension reform design. Section 5.3 describes the theoretical framework. Section

5.4 presents the data and some descriptive evidence, while Section 5.5 illustrates the identification strategy and the estimation method. In Section 5.6, I show the graphical evidence and report the estimation results. Section 5.7 discusses potential threats to identification. Section 5.8 investigates the treatment effect heterogeneity. Section 5.9 concludes.

5.2 Institutional Background

5.2.1 German Public Pension System

The German public pension system was introduced in the 1880s as part of the Bismarckian insurance scheme and is one of the oldest pension systems in the world (Börsch-Supan and Schnabel, 1999; Börsch-Supan and Wilke, 2004). The public pension system is based on lifetime earnings-related pension points converted into pension benefits, with additional points granted e.g. during parental leave and care-giving periods. In 2017, the net replacement rate, i.e. individual net pension entitlement divided by net pre-retirement earnings, amounted to 51% and was thus below the OECD average of 63% (OECD, 2017). Public pension benefits represent the main income source for most retirees and are financed with contribution payments on a pay-as-you-go basis. Contribution payments are borne by employers and employees in equal shares for a total of 18.6% up to a yearly social security contribution ceiling of €82,800 in West Germany and €77,400 in East Germany in 2020 (Deutsche Rentenversicherung, 2019a).

Participation in the statutory pension insurance is mandatory, except for civil servants and self-employed workers. In this respect, the statutory public pension is the first and the most important pillar in the German three-pillar pension system. Voluntary occupational pension plans constitute the second pillar, which is arranged in a funding system organized at the company level or through sectoral collective agreements. The third pillar is given by voluntary private provision schemes with a funding system including e.g. insurance policies, real estate, or bank products. Among all pension benefits, about 80% of payments are provided by the state pension plan, while occupational pension plans and private provision plans are almost equally relevant covering about 10% of payments each.

The German pension system comprises multiple pension schemes generally classified in three major groups: i) *old-age pensions* including regular old-age pension, old-age pension for people with a long insurance record, old-age pension for people with an exceptionally long insurance record, old-age pension due to unemployment or partial retirement, old-age pension for women, old-age pension for severely disabled people, and old-age pension for miners; ii) *surviving dependants' pensions* including widow's or widower's pension, orphan's pension, and child-raising pension; and iii) *disability pensions* including pension for miners and disability pension due to fully or partially reduced earnings capacity.[1] In 2018, pension benefits were paid to about 25,700,000 claimants on total. Among all the beneficiaries, the

[1] All available pension schemes are listed in the German Social Security Legislation *(Sozialgesetzbuch, SGB)*, § 33 SGB VI.

majority received old-age pensions (about 18,250,000 in total) with an average monthly benefits level of €905, followed by about 5,620,000 surviving dependants' pensions with an average monthly benefits level of €590, and about 1,820,000 disability pensions with an average monthly benefits level of €795 (Deutsche Rentenversicherung, 2019b).

Benefits receipt does not basically restrain claimants from earning additional income. However, supplementary income limits can apply. Below the age threshold for the regular old-age pension, also called statutory retirement age, recipients of old-age and disability pensions are allowed to earn up to €6,300 per year without deductions, which arise only if supplementary income is above this threshold. Supplementary income includes income from employment, self-employment, and comparable sources. In the extreme case, pension deductions and pension benefits can completely cancel each other out.[2] Yet upon reaching the statutory retirement age, earned income becomes irrelevant for benefits level. Differently from old-age and disability pensions, supplementary income limits for surviving dependants' pensions are linked to the actual pension benefits level and are thus dynamic, yet they are not applicable to the orphan's pensions at all.

As in many other countries, the earliest claim of retirement benefits in Germany is possible at two statutory retirement ages, the normal retirement age (hereafter, NRA) and the early retirement age (hereafter, ERA). While retirement at the ERA implies permanent pension deductions of 0.3% per month for each month before the statutory NRA, retirement after the NRA provides pension supplements of 0.5% per month amounting to a supplement of 6% per year. Deductions for early retirement and supplements for late retirement were introduced with the pension reform 1992 to postpone retirement entry, thus alleviating the financial burden on the public pension system. As a result, the average effective age of retirement, defined as the average age of exit from the labor force during a 5-year period, increased between 2000 and 2018 from 61 to 64 years for men and from 60.3 to 63.6 years for women (OECD, 2018).

In Appendix 5.A, I provide an overview of the different pathways to retirement. Among all old-age pensions, old-age pension due to unemployment or partial retirement is the most common pension claimed after the regular old-age pension and the old-age pension for women. In 2018, about 11% of almost 18,250,000 old-age pension claimants were retirees due to unemployment or partial retirement with an average benefits level of about €1,300 per month (Deutsche Rentenversicherung, 2019b). Rules for retirement via this pension scheme are explained in more detail in the next Section 5.2.2.

5.2.2 Old-Age Pension due to Unemployment or Partial Retirement

The old-age pension due to unemployment or partial retirement allows unemployed individuals and workers under a progressive retirement plan to retire at age 60 at the earliest.[3]

[2] Rules for calculation of pension deductions due to high additional income are described in § 34 SGB VI.

[3] A progressive retirement plan allows workers aged 55 and older to work part-time. The program was introduced in 1996 and terminated in 2009. For details, see the German Partial Retirement Act

This scheme was de facto phased out by 2016 given that only insured individuals born before January 1952 were eligible for this pension.[4] Benefits receipt is subject to fulfillment of the following criteria: i) at least 15 insurance years (the so-called waiting period or qualifying period); ii) 8 years with paid compulsory contributions for insured employment accumulated in the last 10 years prior to retirement; iii) unemployment at retirement entry with at least 52 weeks of unemployment after reaching age 58.5 or at least 2 years of partial retirement at retirement entry after reaching age 55.

Retirement via this old-age pension can occur at the NRA or the ERA. Both age thresholds have been raised over multiple birth cohorts across time. In Appendix Table 5.B1 and Table 5.B2, I provide an overview of the implemented changes. To begin with, the NRA increased from 60 years by one month with every subsequent birth month starting from January 1937. In this way, the new NRA reached age 65 starting from the cohort born in December 1941 and remained fixed till this program was phased out. Further, the ERA increased from 60 to 63 years for younger cohorts starting from January 1946. The latter policy change is the object of this study and is described in the next Section 5.2.3. More recently, starting from the cohort born in 1947 the statutory retirement age for the regular old-age pension has been increased from 65 to 67 years.[5] This pension reform is gradually implemented between 2012 and 2029 and has therefore no impact on the old-age pension due to unemployment or partial retirement.

5.2.3 Pension Reform: Increase in the Early Retirement Age

The reform of the old-age pension due to unemployment or partial retirement passed into law in July 2004.[6] The pension reform, gradually implemented between January 2006 and December 2011, scheduled an increase in the ERA by 36 months from age 60 to 63. This implies that the first affected cohort were individuals who would have become eligible for early retirement under the old rule in January 2006. Starting with individuals born in January 1946, the ERA increased by one month per each birth month. In this way, when the reform was fully phased-in, individuals born in December 1948 or later could enter early retirement at the minimum age of 63. Birth month is thus the key determinant for the age of first eligibility. A graphical illustration of the policy rule is provided in Figure 5.1.

Relying on the principle of "protection of legitimate expectations", the reform design also entailed an exemption limited to a small group of people. In this respect, among the affected birth cohorts, individuals could be exempted from the increase in the ERA in case of i)

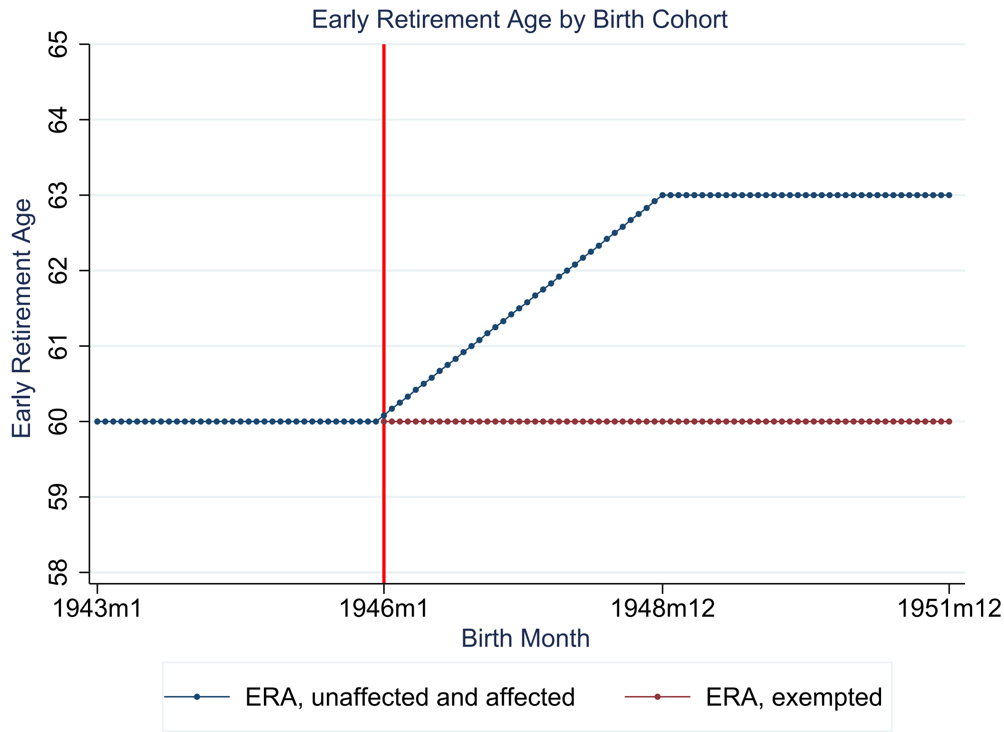
(Altersteilzeitgesetz, AltTZG).

[4]The relevant clause is outlined in § 237 SGB VI in combination with Annex 19 SGB VI.

[5]Gesetz zur Anpassung der Regelaltersgrenze an die demografische Entwicklung und zur Stärkung der Finanzierungsgrundlagen der gesetzlichen Rentenversicherung, 04/2007 *(RV-Altersgrenzenanpassungsgesetz).*

[6]Gesetz zur Sicherung der nachhaltigen Finanzierungsgrundlagen der gesetzlichen Rentenversicherung, 07/2004 *(RV-Nachhaltigkeitsgesetz).*

Figure 5.1: Pension Reform Design



Notes: The vertical line marks the beginning of the pension reform.

unemployment on January 1, 2004; or ii) termination of an employment relationship before January 1, 2004 and non-employment status on January 1, 2004; or iii) termination of an employment relationship after December 31, 2003 due to its fixed-term nature (including fixed-term training measures) or notice given before January 1, 2004; or iv) receipt of transition money for the miners.

5.2.4 Other Relevant Policy Changes

In addition to postponing access to early retirement, poor labor market performance of older workers motivated policy-makers to reduce the attractiveness of unemployment benefits as well. As in other European countries (Pellizzari, 2006, among others), the unemployment insurance (UI) system in Germany relies on two main pillars: Unemployment Benefits I (hereafter, UB I) and Unemployment Benefits II (hereafter, UB II). Originated from a large-scale institutional change called Hartz-Reforms, both unemployment schemes were rearranged in the 2000s. This implies that the implementation time frame for this major reform overlaps with that of the policy change in the public pension system under study. In this respect, altering the relative attractiveness of the UI program may cause variation in the individual labor market behavior which is not attributable to the raised ERA induced by the pension reform. To better understand how the reform of the UI system can interfere with the pension reform under study, I provide an overview of the implemented changes to the UI system below. How this interaction is taken into account in the empirical model is explained in Section 5.4.2.

Reform of Unemployment Benefits I

To begin with, the reform of UB I substantially reduced the maximum duration of unemployment compensation. UB I is funded by employee and employer contributions and is administered by the Federal Employment Agency. All employees subject to social security contributions are covered by this unemployment program. Entitlement to UB I is conditioned on paid contributions, while compensation duration depends on age as well as employment history of the claimant. Workers who reach the statutory retirement age are excluded from the UI coverage. Monthly benefits replace 60% (67% for claimants with children) of the last net salary capped at the social security ceiling. UB I recipients are required to actively search for a job and prove their job searching activities (e.g. applications and responses by potential employers) upon request from the local employment office. Lack of compliance with these requirements can lead to benefits cuts. Moreover, workers who voluntarily quit the job also experience a temporal benefits cut implying a reduction in the maximum duration of unemployment compensation.

The UB I reform was announced in 2004 and came into force in February 2006 affecting workers who lost their jobs after 31 January 2006. This major policy change implied a substantial reduction in the maximum duration of UB I and largely annulled the extensions of the 1980s implemented in response to an increasing unemployment rate and long average spell duration among older workers in West Germany (Hunt, 1995). The introduced innovations were particularly penalizing for workers over age 56 with an applied reduction in the maximum benefits duration by 14 months. Evidence suggests that the policy measure gave rise to strong anticipation effects, with additional unemployment entries from November 2005 till January 2006 (Dlugosz et al., 2014). This anticipation behavior was, however, offset by reduced unemployment in the post-reform period till December 2007.

As early as in January 2008, a new reform of the UI scheme was unexpectedly enacted, re-extending the maximum duration of UB I for older age groups. Nevertheless, this was a minor policy change that did not lead to the pre-reform state. In fact, for workers over age 57 the duration extension amounted to only six months. On top of that, the UB I reform also modified the eligibility criteria and work history requirements. To begin with, under the new (old) regime, claimants were eligible if they worked at least 12 out of the 24 (36) months preceding unemployment. Next, work history requirements for qualification for the maximum benefits duration became more stringent. After (before) the reform, individuals must have worked during the previous 84 months for a number of months equal to at least twice the maximum benefits duration within the last 36 (84) months prior to unemployment. Taken these major changes together, while the replacement rate for calculation of benefits level remained unaffected, rules for benefits eligibility and qualification for the maximum benefits duration became more demanding.

Reform of Unemployment Benefits II

Upon exhaustion of UB I or in case of no entitlement to them, needy unemployed jobseekers receive tax-financed help which is unconstrained by previous earnings and is granted without temporal restrictions. It is means-tested against the household income and aims at providing a living at the subsistence level. A second major reform enacted in January 2005 introduced UB II replacing two previous components of this financial help, namely unemployment assistance *(Arbeitslosenhilfe)* granted to unemployed jobseekers upon exhaustion of unemployment benefits and social assistance *(Sozialhilfe)* granted to all other needy individuals, in particular to those who have never been employed. This policy change aimed to reduce the dependence on benefits. To this end, the reform introduced strict rules with the purpose to motivate recipients to actively search for a job and to intensely cooperate with job centers. Moreover, the reform announced enhanced support to increase the claimants' employability by promoting training programs conveying skills valued on the labor market. Practical enforcement of new rules was achieved through a complex of benefits sanctions reducing the compensation level.

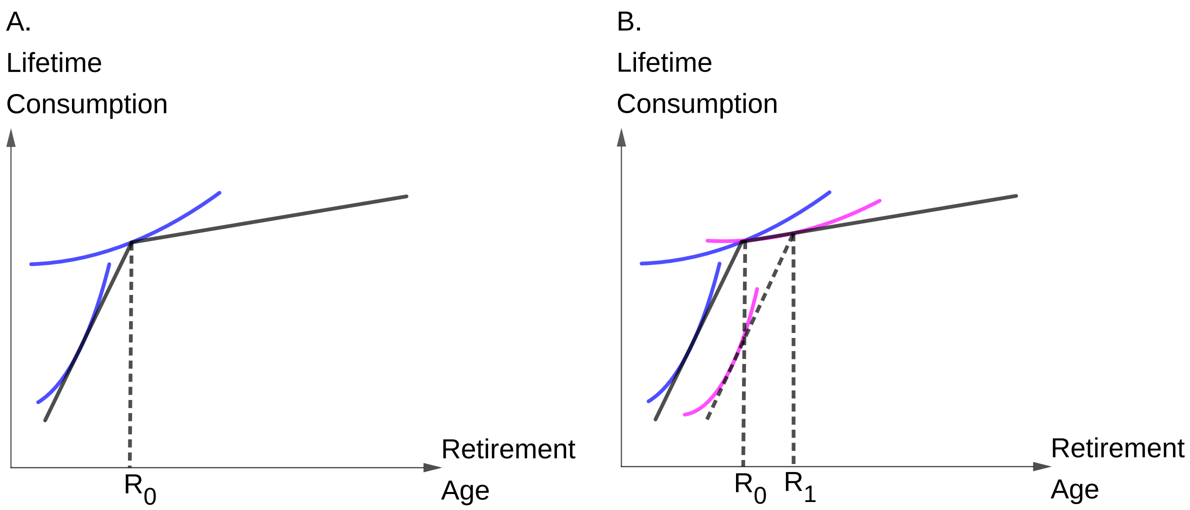
5.3 Theoretical Framework

In this section, I present a static model of retirement decisions in the presence of a binding early retirement age, which is directly related to the German pension reform. The model is common to both the public finance (Kleven, 2016; Kleven and Waseem, 2013; Saez, 2010, among others) and the retirement literature (Brown, 2013). Importantly, in this model the only financial incentive from the pension system is due to the availability of pension benefits at the ERA. The model provides a simple framework to interpret the patterns observed in the data and to motivate the empirical analysis based on the regression kink design.

I first describe the model of retirement decisions with a fixed ERA and then consider responses to a reform that increases the ERA. To keep the model simple, I abstract from uncertainty and time discounting and assume that individuals live for T periods with complete certainty. I further assume that each individual decides on labor supply over the life-cycle by choosing his or her retirement age R. The individual's utility function is defined over life-time consumption C and the retirement age R as $U(C, R) = u(C) - \phi(\theta, R)$, where $\phi(\theta, R)$ denotes the disutility from working R years and the parameter θ reflects heterogeneity in the tastes for work in the population. In this way, individuals with different levels of θ retire at different ages.

Consumption is based on total life-time income from wages w while working and pension benefits while retired. The pension system is defined around the ERA denoted by R_0. I assume that pension benefits are a function of age so that for ages prior to the ERA benefits are 0, and for ages at the ERA and higher, pension benefits are positive. Benefits are taxed at 100% if an individual stays employed after the ERA, and for simplicity, I assume that benefits are set at a constant level b after the ERA. Given these assumptions, the individual's budget constraint is given by $C = wR + \int_R^T [b \times 1(t \geq R_0)]dt$.

Figure 5.2: Retirement Decisions



The budget constraint, illustrated in Figure 5.2 Panel A, shows a kink at the ERA. At ages below the ERA, the slope of the budget constraint equals w as each additional year of work increases income by w. Individuals who remain employed beyond the ERA forgo a year of pension benefits, thus the slope in the budget set above the ERA is reduced to $w - b$. The implicit tax rate on working can be seen as the tax rate τ that solves $w(1 - \tau) = w - b$, or $\tau = \frac{b}{w}$. A key prediction of this model is that there will be bunching of retirements at R_0. The German pension reforms increased the ERA across birth cohorts. To capture this in the model, I consider a policy change that increases the ERA from R_0 to R_1. The increase in the ERA lowers life-time income, as the pension benefits can only be consumed for the shorter period from R_1 to T. Panel B in Figure 5.2 illustrates the change in the budget constraint, which moves from the solid line to the dashed line. It follows that the increase in the ERA shifts the kink point from R_0 to R_1, and above R_1 the budget set remains unchanged. Relative to the pre-reform scenario with the ERA at R_0, the model predicts that bunching shifts from the original ERA, R_0, to the new ERA, R_1. In other words, individuals who brought their retirements forward in response to the pre-reform kink will now shift their retirements to the new kink at R_1. In absence of adjustment costs or frictions, the model thus predicts a shift in bunching from R_0 to R_1.

Apart from the wealth effects of the increase in the ERA, more recent research suggests a potential effect of eligibility age on social norms. In this respect, Seibold (2017) show that statutory retirement age, such as early retirement age as defined by the pension rule, may serve as a reference point for workers' retirement decisions. The simple model can be extended to define a reference-point dependent utility that induces retirement at the ERA by introducing an additional cost a on consumption applied to individuals working after the ERA. In this case, the size of the kink would increase given that the implicit tax rate on working would change to $\tau = \frac{b+a}{w}$.

Finally, the negative wealth effect in response to the German pension reform may induce the affected individuals to adjust behavior along other dimensions. To be more precise, the increase in the ERA may provide incentives to claim benefits from other government programs used as alternative pathways to retirement, thus lowering the impact on expected wealth and labor force participation. In the next Section 5.4, I introduce the data used to quantify the effects of the ERA increase on labor force participation at older ages.

5.4 Data and Sample

5.4.1 Biographical Data of Social Insurance Agencies in Germany

The empirical analysis is based on the Biographical Data of Social Insurance Agencies in Germany (hereafter, BASiD) made available by the Research Institute of the Federal Employment Agency. For an extensive description of the dataset, see Hochfellner et al. (2012). The dataset provides detailed information on the life course of 568,468 individuals and is arranged in episode format. It combines administrative data from two social security agencies, the German Federal Employment Agency and the German Pension Insurance. The data sources include the Integrated Employment Biographies, the Establishment History Panel as well as the Sample of Insured Persons and their Insurance Accounts. The latter is the basis for data fusion. It consists of a 1% disproportionately stratified random sample by agency, sex, nationality, and year of birth of individuals aged between 15 and 67 with an account at the German Pension Insurance on December 31, 2007 as well as paid contributions in the respective year. The final dataset BASiD covers employment biographies in the period from 1951 till 2009 with information on retirement behavior until the end of 2007 from the pension insurance accounts. The observation period begins with entry into the education system and ends with entry into retirement. In this respect, BASiD is a unique dataset containing a full range of variables with longitudinal information on education, employment and benefit history, job seeking and training measures, periods of illness and motherhood, retirement decisions, and socio-demographic characteristics, among which month of birth is crucial for the analysis.

5.4.2 Sample and Variables Definition

The analysis sample is defined to reflect the target population group of the pension reform under study. To this end, I exclude from BASiD self-employed people, civil servants, people whose last labor market activity ends at or before age 60 as well as people whose last employment relationship ends due to death. People at least once employed in East Germany are similarly excluded given that employment biographies of workers in East

Germany have been recorded only since January 1, 1991. This restriction is necessary to calculate individual entitlements to various pension types over a longer time horizon.

Among the retirees in the data set, people receiving invalidity pension benefits or old-age pension benefits for severely disable people, old-age pension benefits for women, and old-age pension benefits for miners are also excluded from the sample due to special pension rules. A synopsis of the pension schemes relevant for the analysis sample is presented in Table 5.1. To resemble the fulfillment of the eligibility criteria for the old-age pension due to unemployment or partial retirement, I restrict the sample to individuals with at least 15 insurance years at age 58. This constraint is applied to approximate the criteria fulfillment before the announcement of the pension reform and to hold age constant while comparing behavioral responses across birth cohorts. I further keep only individuals born between January 1945 (hereafter, 1945m1) and January 1947 (hereafter, 1947m1) and aged between 60 and 61 years. The advantage of a small range of birth cohorts is that individuals unaffected by the pension reform can be considered close substitutes for the affected individuals. In other words, their trends in labor market behavior are likely to be comparable.

Table 5.1: Sample's Pathways to Retirement

Pension Scheme	Early (ERA)	Normal (NRA)	Waiting Period	Contribution Period
Old-Age Pension for Women	60	64-65	15	10 after age 40
Old-Age Pension for Severely Diasabled People	60	63	35	
Old-Age Pension due to Unemployment or Partial Retirement	60-62	65	15	8 in last 10 years
Old-Age Pension for Miners	-	60	25	
Disability Pension	-	63	5	3 in last 5 years

Notes: The table outlines the pension schemes relevant for the estimation sample as defined in this study. For details on all available pathways to retirement, see Appendix 5.A. ERA (NRA) indicates Early Retirement Age (Normal Retirement Age). Retirement at the ERA implies permanent pension deductions of 0.3% per month for each month before the statutory NRA. The maximum pension deductions due to retirement via disability pension before age 63 amount to 10.8% with a reference age for deductions of 63. Waiting period, also known as qualifying period, denotes the total number of insurance years. Contribution period indicates the total number of years with paid compulsory contributions for insured employment.

Next, I set the individual ERA as follows. Individuals born before 1946m1 belong to the birth cohorts unaffected by the policy measure and are therefore assigned an ERA of 60 years. On the contrary, individuals born in or after 1946m1 are assigned an ERA of 60 years if exempted from the ERA increase or a higher ERA according to the policy rule if not exempted.[7] In the same spirit, given that fulfillment of eligibility criteria for the old-age pension for women entitles them to pension benefits at age 60 at the earliest, I assign an ERA of 60 to this group of women as well. This is motivated by the retirement incentives associated with the old-age pension for women. In fact, this pension scheme offers an attractive pathway to retirement given that it subsidizes short employment careers

[7]The applied procedure to identify the exempted individuals in the data is summarized in Appendix 5.C.

of women due to e.g. child-raising periods. Therefore, based on women's employment biographies, eligibility to the old-age pension for women at age 58 determines their self-selection into the control group with an assigned ERA of 60. A correct assignment of individuals to treatment and control groups allows to compare the labor market behavior of the affected individuals to the unaffected ones, exploiting an exogenous variation in the access to early retirement induced by the reform.

From the original data set stored in episode format, I generate a monthly panel with the following labor market outcomes of interest: employment, claim of retirement benefits, claim of disability benefits, and unemployment.[8] Employment is defined as a dummy variable equal to one if an individual results to be employed s.t. social security contributions or marginally employed, and zero otherwise. Claim of retirement benefits is a dummy variable equal to one if an individual receives old-age pension benefits, and zero otherwise. Claim of disability benefits is a dummy variable equal to one if an individual receives disability pension benefits, and zero otherwise. Unemployment is a dummy variable equal to one covering registered unemployment without necessarily receiving the associated benefits, and zero otherwise. The residual category includes individuals who are neither employed, receiving retirement benefits, receiving disability benefits, nor registered unemployed.

Exploiting the detailed employment biographies recorded in BASiD over a long time period, I define variables capturing individual labor market attachment. In this respect, for each individual in the sample I compute the total number of unemployment months up to age 50, the total number of sick leave months up to age 50, the total number of years with paid compulsory contributions for insured employment up to age 50, and yearly average earnings between ages 50 and 55 adjusted for inflation. Further variables include a dummy for high education at age 50 equal to one if an individual has at least high school education, and zero otherwise, and dummies for working in 123 occupations, with missing values captured in the reference category.[9] Each of the occupational dummy variables is equal to one if an individual is employed in the respective occupation at age 50 or shortly before in case occupation is not recorded due to e.g. an unemployment spell at age 50, and zero otherwise.

Then, I define variables capturing individual characteristics. These include dummy variables for 17 regions of residence including residence abroad, with missing values captured in the reference category, a dummy variable equal to one if an individual is female, and zero otherwise, a dummy variable equal to one if an individual has at least one child, and zero otherwise, as well as a continuous variable indicating the number of children.[10] Finally, a dummy variable for being affected by the UB I reform interacted with year-month fixed

[8]Labor market state in a month is defined as a state reported on the first of month during the respective spell. Alternative definitions refer to the labor market state reported on Monday of e.g. the first week in a month or the labor market state during a reference week starting on Monday of a month week.

[9]Given that individual education level is likely to oscillate at younger ages while remaining stable with age in the middle-aged population, to align the age at which the covariates are computed the dummy variable for high education is computed at age 50 as well. Although obtaining a higher level of education at an older age is possible, this choice implies that the highest education level in the sample is assumed to have been achieved up to age 50.

[10]Information about the number of children is drawn from pension insurance accounts which are an

effects completes the list of covariates included in the regression model in Section 5.5. To be more precise, I first define a dummy capturing the relevance of the UB I reform across individuals and panel months with respect to the reform introduction and then interact this dummy variable with year-month fixed effects that capture the time trend.

5.4.3 Descriptive Statistics

The final sample includes 51,779 observations for 4,060 individuals with 39.9% of female observations and an average age of 60.5 years. Notifications on labor market states in Panel A suggest that 51.2% of observations are employed (Panel A1). For the vast majority, employment is subject to social security contributions (88%), while marginal employment plays only a minor role (12%). Retirement benefits (Panel A2) are claimed by 11.1% of observations. Among retirees, about 81.6% of observations are not engaged in any other activity, while for 12.1% of them marginal employment is an additional source of income. A negligible share of individuals is enrolled in the disability insurance program (Panel A3) relevant for only 4.7% of observations in the sample. Unlike retirees via the old-age pension scheme in Panel A2, only 31.5% of observations receive disability pension benefits and are not employed at the same time. Furthermore, 32.9% of observations are marginally employed, while 16.5% of them are employed subject to social security contributions. Registered unemployment (Panel A4) is relevant for about 13.7% of observations and is mainly associated with receipt of unemployment benefits. In this respect, 54.1% (38.9%) of unemployed receive UB I (UB II), while 1.7% of registered unemployed are engaged in job searching without receiving any kind of unemployment benefits. About 19.3% of observations are in the residual category (Panel A5). This category covers mainly partial retirement programs (78.1%) and to a lesser extent sick leave periods (3.1%) as well as unregistered unemployment (2.4%).

In Panel B, I present summary statistics for sample's background characteristics covering education level, labor marker performance, and some family information. In this respect, the total number of years with paid compulsory contributions for insured employment accumulated at age 50 is about 26.5 years, while yearly average earnings between ages 50 and 55 amount to almost €24,500 per year adjusted for inflation, on average. The number of insurance years at age 58 reaches a maximum of 44.3 years and amounts to 33.7 years on average, which well exceeds the imposed 15 years. Only 2.98% of observations with at least 15 insurance years at age 60 do not accumulate 15 years by age 58. Among observations with at least 15 insurance years at age 58, about 82.1% of them accumulate 8 years with paid compulsory contributions for insured employment in the last 10 years before age 58. This points to a sample with rather high labor market affinity and an authentic choice to participate in the labor market at an older age. This is also reflected in a relatively small number of unemployment months accumulated at age 50 (about 8.2 months, on average)

integral part of BASiD. Given that this information is recorded on December 31, 2007, it is measured at different ages (between 60 and 62) across individuals in the sample.

and that of sick leave months accumulated at age 50 (about 2.2 months, on average). Summary statistics of the analysis sample is presented in Table 5.2.

Table 5.2: Summary Statistics

	Mean	S.d.	Min	Max
A. Labor Market Outcomes				
A1. Employment	0.512	0.5	0	1
Employment s.t. social security contributions	0.880			
Marginal employment	0.120			
A2. Retirement Benefits	0.111	0.314	0	1
Pension payments only	0.816			
Parallel marginal employment	0.121			
Parallel job-seeking, unregistered unemployment	0.043			
Parallel unemployment benefits II	0.015			
Parallel employment s.t. social security contributions	/			
Parallel unemployment benefits I	/			
Parallel unemployment benefits, only credited periods	/			
Parallel job-seeking, registered unemployment	/			
A3. Disability Benefits	0.047	0.212	0	1
Pension payments only	0.315			
Parallel marginal employment	0.329			
Parallel employment s.t. social security contributions	0.165			
Parallel non-professional care-giving	0.058			
Parallel unemployment benefits II	0.039			
Parallel job-seeking, unregistered unemployment	0.037			
Parallel unemployment benefits I	0.026			
Parallel sick leave	0.012			
Parallel vocational training	/			
Parallel job-seeking, registered unemployment	/			
Parallel voluntary insurance	/			
Parallel partial retirement	/			
A4. Unemployment	0.137	0.344	0	1
Unemployment benefits I	0.541			
Unemployment benefits II	0.389			
Unemployment benefits, only credited periods	0.054			
Job-Seeking, registered unemployment	0.017			
A5. Residual Category	0.193	0.395	0	1
Partial retirement	0.781			
Voluntary insurance	0.138			
Sick leave	0.031			
Job-Seeking, unregistered unemployment	0.024			
Non-Professional care-giving	0.021			

Promotion of vocation education	0.004			
B. Background Characteristics				
High education at age 50	0.082	0.275	0	1
Unemployment at age 50, in months	8.202	19.387	0	204.616
Sick leave at age 50, in months	2.163	5.458	0	106.513
Compulsory contribution period for insured employment at age 50, in years	26.533	8.030	6.834	36.668
Yearly average earnings, ages 50-55	24,484	15,216	0	60,607
Compulsory contribution period for insured employment, ages 48-58, in years	8.879	2.384	0	10.083
Compulsory contribution period for insured employment, ages 48-58, $\geq$8 years	0.821	0.384	0	1
Insurance period at age 58, in years	33.684	8.665	15.001	44.329
Employment sector at age 58, manufac.	0.279	0.449	0	1
Employment sector at age 58, service	0.332	0.471	0	1
Employment sector at age 58, other	0.305	0.46	0	1
Employment sector at age 58, missing	0.084	0.277	0	1
Children in 2007, at least one	0.335	0.472	0	1
Children in 2007, number	0.681	1.125	0	10
Female	0.399	0.49	0	1
Age, in years	60.497	0.311	60	61
Legitimate expectations	0.067	0.251	0	1
Early retirement age (ERA)	60.189	0.327	60	61.08
Below ERA	0.375	0.484	0	1
Below ERA, treatment cohorts (1946m1-1947m1)	0.635	0.481	0	1

Notes: Sample size is 51,779 observations for 4,060 individuals. Variables for at least one child and number of children are drawn from pension insurance accounts and are recorded on December 31, 2007. Legitimate expectations refer to a group of individuals exempted from the ERA increase. Yearly average earnings in € between ages 50-55 are adjusted for inflation. / denotes suppression of small numbers for sensitive information implemented within the disclosure control by the data provider.

5.5 Econometric Framework

5.5.1 Regression Kink Design

In a regression kink design (hereafter, RKD), a term coined by Nielsen et al. (2010), the policy rule is defined by the presence of a kink in the formula relating the assignment variable to the policy variable. In such settings, the kinked assignment rule allows to identify the policy variable effect.[11] This study exploits the kinked pension reform schedule, as illustrated in Figure 5.1, linking birth month (assignment variable) and early retirement age (policy variable) to measure the average labor market response to the increased ERA. The strategy is to use the corresponding kink in the relationship between birth month and labor market outcome and relate the slope change in the outcome relationship to the slope change in the policy rule defined by the reform. Given that ERA is a function of birth month, it is likely to be correlated with other unobserved individual characteristics that determine labor supply and retirement decisions. The RKD circumvents this endogeneity problem by using the quasi-experimental variation in the ERA induced by the pension reform.

To define the estimator formally, let Y be the outcome of interest, ERA the early retirement age as determined by the policy rule, and V the birth month. Under the smoothness conditions (Card et al., 2015), the regression kink estimand

$$\tau = \frac{\lim\limits_{v_0 \to 0^+} \frac{dE[Y|V=v]}{dv}\Big|_{v=v_0} - \lim\limits_{v_0 \to 0^-} \frac{dE[Y|V=v]}{dv}\Big|_{v=v_0}}{\lim\limits_{v_0 \to 0^+} \frac{dE[ERA|V=v]}{dv}\Big|_{v=v_0} - \lim\limits_{v_0 \to 0^-} \frac{dE[ERA|V=v]}{dv}\Big|_{v=v_0}} \tag{5.1}$$

identifies a weighted average of the marginal effects of ERA on Y at the kink point $v = 0$.[12] The RKD estimate is obtained by dividing the estimated slope change in the numerator $E[Y|V = v]$ (reduced form) by the estimated slope change in the denominator $E[ERA|V = v]$ (first stage) of equation (5.1). The numerator and denominator are estimated by running the following parametric polynomial regressions, respectively:

$$Y_{it} = \mu_y + \left[\sum_{p=1}^{\bar{p}} \gamma_p v_i^p + \alpha_p v_i^p * D_i\right] + \theta_y X_{it} + \epsilon_{it}, \qquad \text{with } |v_i| \leq h \tag{5.2}$$

$$ERA_i = \mu_b + \left[\sum_{p=1}^{\bar{p}} \sigma_p v_i^p + \beta_p v_i^p * D_i\right] + \theta_b X_i + \epsilon_i, \qquad \text{with } |v_i| \leq h \tag{5.3}$$

[11] The kink-based identification strategy has been employed in a variety of research fields. For example, Guryan (2001) uses the discontinuous state education aid formula as an instrumental variable for the increase in per-pupil spending to causally estimate the effect of spending on student test scores. In the same spirit, Dahlberg et al. (2008) use the discontinuous formula for the distribution of funds to estimate the causal effects of intergovernmental grants on local spending and taxes. Similarly to studies on the impact of government grant policies, Nielsen et al. (2010) use a a kinked student aid scheme to identify the effect of direct costs on college enrollment. More recently, Simonsen et al. (2016) estimate price sensitivity of demand for prescription drugs, exploiting variation in prices induced by the kinked reimbursement scheme.

[12] In the RKD setting, the marginal effect refers to the derivative of the outcome with respect to the continuous endogenous regressor and should not be confused with the marginal treatment effect in Heckman and Vytlacil (2005), where the treatment is binary.

where v is the assignment variable centered around the first treated birth cohort (i.e. 1946m1) chosen as a kink point, $D_i = \mathbb{1}[v_i \geq 0]$ is an indicator for being at or above the kink, p is the polynomial order, h is the bandwidth size. X_{it} includes the control variables defined in Section 5.4.2: total number of unemployment months up to age 50, total number of sick leave months up to age 50, total number of years with paid compulsory contributions for insured employment up to age 50, yearly average earnings between ages 50 and 55 adjusted for inflation, total number of children in 2007; dummies for at least one child in 2007, high education at age 50, occupation at age 50, female, region of residence in month t, and finally a dummy equal to one if an individual is affected by the UB I reform in month t, and zero otherwise, interacted with year-month fixed effects. Standard errors are clustered by month of birth allowing for correlation between individuals born in the same month.

I adopt a fuzzy RKD design and estimate the slope change of the first stage function $E[ERA|V = v]$ to account for variation in the eligibility for early retirement pension as well as in the exemption from the ERA increase at the individual level. In fact, individual number of insurance years and legitimate expectations laid down by the reform might be measured with some error and can thus be treated as unobserved determinants of ERA. Therefore, allowing for deviations of the observed value for ERA from the predicted ERA based on the policy rule, a fuzzy RKD is more appropriate compared to a sharp RKD.[13]

The estimated change in the slope of Y around the kink is given by α_1 in equation (5.2), while the estimated change in the slope of ERA around the kink is given by β_1 in equation (5.3). For estimation, I use a uniform kernel, which is a common practice in the Regression Discontinuity literature (Imbens and Lemieux, 2008), and choose a maximum symmetric bandwidth of 12 months around the kink point. I first present estimates from a parametric estimation of RKD using local linear polynomial estimation based on the preferred bandwidth (see Section 5.6.2) and then explore the estimates sensitivity to alternative bandwidths (see Section 5.7.1).

5.5.2 Threats to Identification

Before discussing the estimation results in the next Section 5.6, I assess the validity of the adopted research design via testable implications on identifying assumptions. There are two key assumptions for identification in the RKD (Card et al., 2015). The first identifying assumption implies a continuously differentiable density of the assignment variable V. In other words, the distribution of V for each individual is required to be sufficiently smooth. This condition rules out the possibility that an individual can precisely manipulate V.[14]

[13]In a setting with full compliance with the policy rule and no measurement error, a sharp RKD is more appropriate. In a sharp RKD, the treatment variable is a deterministic function of the assignment variable, implying that the slope change of the first stage function is a known constant.

[14]In general, the smooth density condition rules out deterministic sorting, though mild forms of endogeneity are allowed. These cover, for instance, situations with endogenous sorting but also small optimization errors (Chetty, 2012).

In this study, ERA is a function of birth month V, which is exogenous and not subject to manipulation for individuals around the cutoff birth dates. Furthermore, it is not plausible that the policy makers selected particular cohorts for the examined reform given that adjacent birth cohorts were affected by similar pension reforms directed to gradually postpone entry into retirement. This practice rather indicates a long-run perspective of the pension policy. In support of the validity of these arguments, I present the frequency of observations by birth month in Table 5.3. The pattern shows a relatively stable number of observations across birth cohorts. Slight fluctuations in frequencies, if anything, might rather reflect variations in the population birth rates around World War II unrelated to the pension reform under study.

Table 5.3: Distribution of Monthly Birth Cohorts

Birth Cohort (1)	Observations (2)	Individuals (3)
1945m1	2,640	206
1945m2	2,224	172
1945m3	2,535	195
1945m4	1,955	153
1945m5	1,934	151
1945m6	1,882	149
1945m7	1,877	145
1945m8	1,986	155
1945m9	2,227	175
1945m10	2,084	162
1945m11	1,336	104
1945m12	1,443	114
1946m1	1,622	127
1946m2	1,640	128
1946m3	1,950	151
1946m4	2,090	163
1946m5	2,361	185
1946m6	2,014	160
1946m7	2,347	183
1946m8	2,328	182
1946m9	2,223	176
1946m10	2,279	178
1946m11	2,100	164
1946m12	2,260	177
1947m1	2,442	205
Total	51,779	4,060

Notes: Estimation sample.

The second identifying assumption requires the smoothness of conditional expectation of pre-determined covariates around the kink point. This is to assure that the observed kinks in the outcome variables are not caused by the kink in the sample's characteristics. Panel A from Table 5.4 shows estimation results from local linear regressions for a set of characteristics: dummy for high education at age 50, total number of unemployment months accumulated at age 50, total number of sick leave months accumulated at age 50, total number of compulsory contribution years for insured employment accumulated at age 50, yearly average earnings in € adjusted for inflation between ages 50-55, dummy for at least one child in 2007, number of children in 2007, and dummy for female. Large p-values across all variables point to insignificant coefficient estimates, which in turn implies no slope changes in the relationship between individual characteristics and birth months. In Panel B from Table 5.4, I combine multiple covariates to estimate composite "covariate indices" by predicting labor market outcomes using covariates based on rich information on employment and earnings histories from Panel A and adding multiple dummies for occupation at age 50 as regressors. In this respect, the estimated covariate index functions can be understood as best linear predictions of the outcomes of interest, given the vector of pre-determined variables. The coefficient estimates point again to insignificant kinks in the relationship between birth months and predicted outcomes. By and large, the results from both tests suggest that the covariates evolve smoothly around the kink point, implying that kinks in covariates are not likely to be the driving force of the main results. Taken these pieces of evidence together, I conclude that the identifying assumptions necessary for a valid RKD hold in the context under study.

Table 5.4: Smoothness around the Kink Point

	Coeff. (1)	(SE) (2)	P-Value (3)	Mean (4)	S.d. (5)
			A. Covariates		
High education, at age 50	-0.002	(0.003)	0.632	0.082	0.275
Unemployment months, at age 50	-0.073	(0.167)	0.667	8.202	19.387
Sick leave months, at age 50	-0.021	(0.0227)	0.442	2.163	5.458
Compulsory contribution years for insured employment, at age 50	0.034	(0.094)	0.725	26.533	8.030
Yearly average earnings in € between ages 50-55	49.502	(110.049)	0.657	24,483.930	15,215.960
At least one child, in 2007	0.005	(0.005)	0.292	0.335	0.472
Number of children, in 2007	0.008	(0.010)	0.418	0.681	1.125
Female	0.005	(0.004)	0.238	0.399	0.490
			B. Covariate Indices		
Predicted Employment	0.001	(0.001)	0.539	0.512	0.148
Predicted Retirement Benefits	0.000	(0.001)	0.619	0.111	0.096
Predicted Disability Benefits	0.000	(0.000)	0.716	0.047	0.063

Predicted Unemployment	-0.002	(0.001)	0.103	0.137	0.108

Notes: The table reports estimates of slope changes from fitting linear polynomials in the month of birth to the left and the right of the kink threshold (1946m1) to the outcome variables indicated in each line. The estimates are based on a sample with a maximum symmetric bandwidth of 12 months around the kink point. Yearly average earnings in € between ages 50-55 are adjusted for inflation. The estimated covariate index functions in Panel B are best linear predictions of the labor market outcomes using pre-determined covariates from Panel A and occupation at age 50 as regressors. Standard errors in parentheses are clustered at the birth month level. Sample size is 51,779 observations for 4,060 individuals.

5.6 Regression Kink Analysis

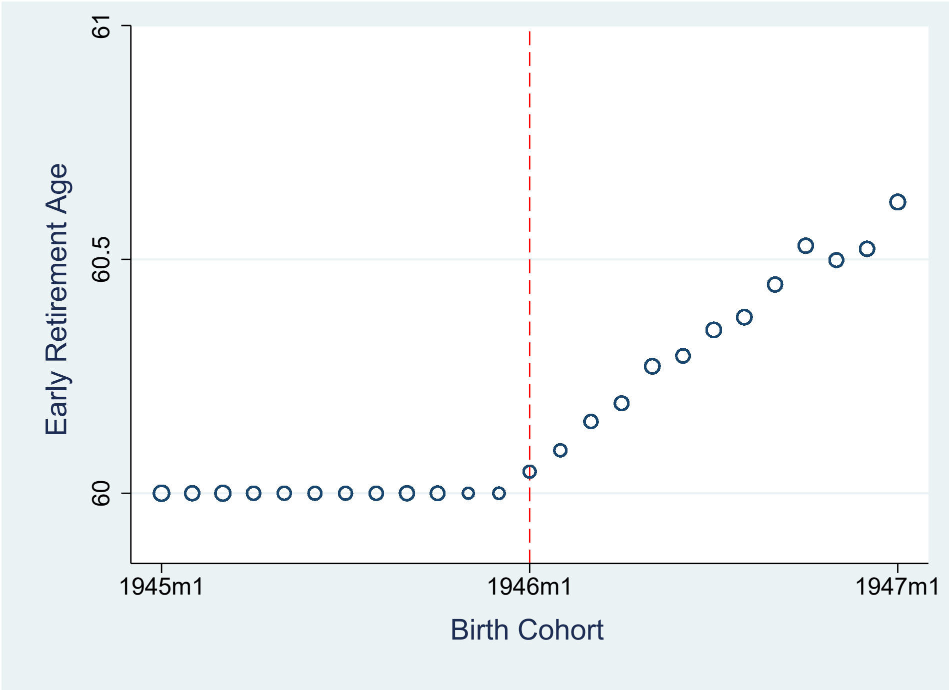
5.6.1 Graphical Evidence

I first present the graphical results of the regression kink analysis showing the first stage and reduced forms for the outcomes of interest, and then turn to the regression analysis in the next Section 5.6.2. Figure 5.3 plots the first stage illustrating the change in the average ERA by monthly birth cohort. Although Figure 5.3 largely follows the pattern arising from the policy rule (compare Figure 5.1), the highest average ERA among the treated birth cohorts does not reach the maximum individual ERA of 61 years in the sample. This is mainly because the average ERA per birth month to the right of the kink point includes both individuals affected by the reform and those unaffected due to legitimate expectations, for whom the pre-reform ERA remains constant at age 60 till the pension scheme completely phases out.[15]

Figure 5.4 shows the reduced form results for average labor market outcomes by monthly birth cohort around the kink point. The idea underlying these graphs is that if ERA exerts a causal effect on Y and there is a kink in the relationship between ERA and V at $v = 0$, then one should expect an induced kink in the relationship between Y and V at $v = 0$. Indeed, in spite of large 95% confidence intervals linked to a small number of observations per bin, these figures present first evidence of raw slope changes in the relationship between labor market outcomes and birth months around the kink point. In this respect, employment (Panel A) is likely to increase and claim of retirement benefits (Panel B) to decrease at $v = 0$. Slope change around the kink point for claim of disability benefits (Panel C) seems negligible, while unemployment (Panel D) is likely to decrease though its slope change at $v = 0$ is less pronounced compared to Panels A and B. Keeping

[15]This is in line with the motivation to adopt a fuzzy RKD. Similar to the LATE interpretation in a standard instrumental variables setting, the fuzzy RKD estimand upweights types of individuals with a larger kink at the threshold $v = 0$, while individuals whose policy schedule is not kinked at $v = 0$ do not contribute to the estimand. For details on the factor components by which the average marginal effects are weighted in a fuzzy RKD, see Card et al. (2015, p.2,467).

Figure 5.3: Early Retirement Age by Monthly Birth Cohort



Notes: Markers are weighted by number of observations. The vertical line marks the beginning of the pension reform.

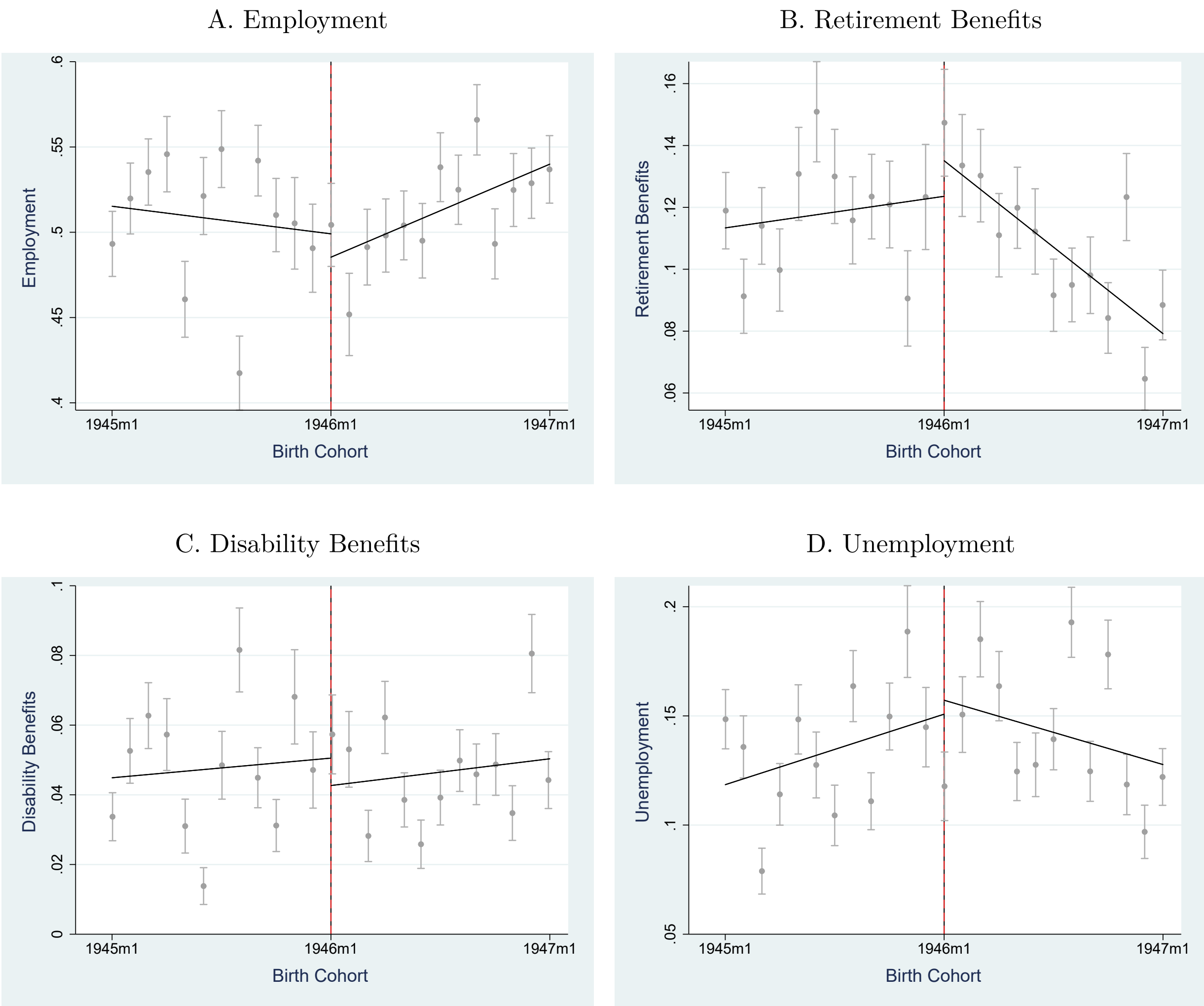
in mind the concurrent institutional changes in the UI system described in Section 5.2.4, the presented graphical evidence should be interpreted with caution. In fact, the plotted reduced form functions ignore the coincidence of the timing of legislative changes in the pension insurance and unemployment insurance systems, for which I explicitly control in the regression analysis in the next Section 5.6.2.

5.6.2 Estimation Results

In this section, I present the main regression kink results for reduced form and fuzzy estimates based on a local linear regression model and a maximum symmetric bandwidth of 12 months around the kink point. Robustness checks based on alternative bandwidth selectors are provided in Section 5.7.1.

Regression results in Table 5.5 quantify the changes in slopes of the average ERA and labor market outcomes. The estimates suggest that according to the reform schedule the ERA increases by 0.046 of a year (column (1)), while employment increases by 0.5 percentage points (column (2)) and claim of retirement benefits decreases by 0.7 percentage points (column (3)), on average. Moving to the reduced form estimates for program substitution via disability insurance (column (4)) and unemployment insurance (column (5)), it follows that both coefficients are relatively small while none of them is statistically significant.

Figure 5.4: Labor Market Outcomes by Monthly Birth Cohort



Notes: Plots are implemented by `rdplot`. The vertical line marks the beginning of the pension reform.

Table 5.5: Reduced Form Regression Kink Estimates

	ERA	Employment	Retirement Benefits	Disability Benefits	Unemployment
	(1)	(2)	(3)	(4)	(5)
Birth Month	0.046***	0.005***	-0.007***	-0.001	-0.003
	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)
Mean	60.189	0.512	0.111	0.047	0.137
S.d.	0.327	0.500	0.314	0.212	0.344

Notes: The table reports estimates of slope changes from fitting linear polynomials in the month of birth to the left and the right of the kink threshold (1946m1) to the outcome variables indicated in each column. The estimates are based on a sample with a maximum symmetric bandwidth of 12 months around the kink point. All regressions control for high education at age 50 (dummy), unemployment months at age 50, sick leave months at age 50, compulsory contribution years for insured employment at age 50, yearly average earnings in € adjusted for inflation between ages 50-55, at least one child in 2007 (dummy), number of children in 2007, and female (dummy) and include region fixed effects, occupation at age 50 fixed effects, and a dummy for relevance of the UB I reform interacted with year-month fixed effects. Standard errors in parentheses are clustered at the birth month level. Sample size is 51,779 observations for 4,060 individuals. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.6: Fuzzy Regression Kink Estimates

	Employment	Retirement Benefits	Disability Benefits	Unemployment
	(1)	(2)	(3)	(4)
ERA	0.116***	-0.150***	-0.012	-0.055
	(0.039)	(0.035)	(0.030)	(0.057)
Mean	0.512	0.111	0.047	0.137
S.d.	0.500	0.314	0.212	0.344

Notes: The table reports IV estimates of slope changes from fitting linear polynomials in the ERA to the left and the right of the kink threshold (1946m1) and instrumenting the ERA with birth month. The estimates are based on a sample with a maximum symmetric bandwidth of 12 months around the kink point. All regressions control for high education at age 50 (dummy), unemployment months at age 50, sick leave months at age 50, compulsory contribution years for insured employment at age 50, yearly average earnings in € adjusted for inflation between ages 50-55, at least one child in 2007 (dummy), number of children in 2007, and female (dummy) and include region fixed effects, occupation at age 50 fixed effects, and a dummy for relevance of the UB I reform interacted with year-month fixed effects. Standard errors in parentheses are clustered at the birth month level. Sample size is 51,779 observations for 4,060 individuals. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

I now turn to the fuzzy regression kink estimates in Table 5.6. The fuzzy regression kink estimates are based on the ratio of reduced form and first stage estimates and represent therefore the effect of increasing the ERA by one year on average labor market outcomes. The results reported in columns (1)-(2) of Table 5.6 point to a positive employment effect and a negative retirement effect of the pension reform, with both effects being highly statistically significant. In particular, I find that a one-year increase in the ERA increases employment by 11.6 percentage points (22.7%). This is in line with Giesecke and Kind

(2013) who analyze the same reform and find evidence of increased reemployment probability among older unemployed men. The estimated employment response reflects the short-run effect of the ERA increase. The longer-term effect may differ to the extent that younger birth cohorts, being aware of the higher ERA well in advance, smooth their consumption earlier on. Therefore, the expected employment response of the ERA increase in the longer-run is likely to be lower due to this anticipation behavior.

Next, a one-year increase in the ERA reduces claim of retirement benefits by 15 percentage points (135%), on average. The large negative impact on claim of retirement benefits is plausible given that it reflects the mechanical effect of the ERA increase induced by the pension reform. Reason is that, in line with sample definition, this outcome variable includes benefits granted to the old-age pensioners due to unemployment or partial retirement who are forced to delay their benefits claims until the new ERA. Therefore, changes in retirement benefits claims are mechanically related to changes in the ERA. All together, these findings hint at the effectiveness of the adopted policy tool to increase labor force participation among older workers and are in line with previous studies (Geyer and Welteke, 2019; Staubli and Zweimüller, 2013, among other).

The estimates in columns (3)-(4) of Table 5.6 show no evidence of program substitution in response to the ERA increase, suggesting that disability insurance and unemployment insurance programs are not used as alternative pathways to retirement. This is in contrast with Duggan et al. (2007) finding that an increase in the NRA in the U.S. context significantly increased enrollment into the disability insurance program. However, keeping in mind that individual age in the sample ranges between 60 to 61 years, insignificant coefficient estimates for claim of disability benefits (column(3)) are not surprising if disability pension has been claimed already before age 60. This interpretation seems plausible in the light of the official statistics on payments of disability pensions from the German Pension Insurance (Deutsche Rentenversicherung, 2019c). In this respect, official data report that nearly all disability pensions are reduced by the maximum deduction rate of 10.8%. It follows that most recipients claim disability pensions before age 60, implying that most workers with poor health exit the labor market before reaching the ERA through the disability insurance program (see Appendix 5.A). In addition, health-related eligibility criteria are relatively demanding, with about 50 % of applications being rejected. Even though the formal eligibility criteria for disability insurance did not change with the ERA increase, it might still be possible that the insignificant coefficient estimate is due to applications for disability benefits being screened more rigorously after the increase in the ERA. In this case, there would be little variation in disability enrollment despite the increased number of applications.

Differently from Engels et al. (2017) who show that introduced deductions for early retirement for women in Germany caused women aged 60-65 to use unemployment as a bridge to retirement, I do not find evidence for increased program substitution with the UI scheme in my sample. In this respect, it is interesting to discuss how the institutional context is taken into account in the analysis of the labor market effects of the ERA increase. As mentioned in Section 5.2.4, two major reforms of the UI system were enacted almost contemporaneously with the pension reform under study. To begin with, the UB II reform was introduced in January 2005. Given that the observation period in the analysis sample is between January 2005 and December 2007, both control and treatment groups are equally affected by this institutional change. Hence, the introduction of the UB II reform is unlikely to confound the estimates. On the contrary, the UB I reform, which shortened the maximum duration of unemployment benefits for workers who lost their jobs starting from February 2006, affected the control and treatment groups over time to a different extent. To illustrate this point, I present the sample composition in Table 5.7. This table shows that the treatment cohorts (starting from individuals born in or after February 1946) are affected during the whole observational period, while the relevance of the UB I reform for the control cohorts (individuals born in or after January 1945) varies across panel months.

Table 5.7: Relevance of the UB I Reform for Treatment and Control Groups in the Analysis Sample

	Month	2005												2006												2007											
		1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
1945	1	C	C	C	C	C	C	C	C	C	C	C	C	C																							
	2		C	C	C	C	C	C	C	C	C	C	C	C	C																						
	3			C	C	C	C	C	C	C	C	C	C	C	C	C																					
	4				C	C	C	C	C	C	C	C	C	C	C	C	C																				
	5					C	C	C	C	C	C	C	C	C	C	C	C	C																			
	6						C	C	C	C	C	C	C	C	C	C	C	C	C																		
	7							C	C	C	C	C	C	C	C	C	C	C	C	C																	
	8								C	C	C	C	C	C	C	C	C	C	C	C	C																
	9									C	C	C	C	C	C	C	C	C	C	C	C	C															
	10										C	C	C	C	C	C	C	C	C	C	C	C	C														
	11											C	C	C	C	C	C	C	C	C	C	C	C	C													
	12												C	C	C	C	C	C	C	C	C	C	C	C	C												
1946	1													T	T	T	T	T	T	T	T	T	T	T	T	T											
	2														T	T	T	T	T	T	T	T	T	T	T	T	T										
	3															T	T	T	T	T	T	T	T	T	T	T	T	T									
	4																T	T	T	T	T	T	T	T	T	T	T	T	T								
	5																	T	T	T	T	T	T	T	T	T	T	T	T	T							
	6																		T	T	T	T	T	T	T	T	T	T	T	T	T						
	7																			T	T	T	T	T	T	T	T	T	T	T	T	T					
	8																				T	T	T	T	T	T	T	T	T	T	T	T	T				
	9																					T	T	T	T	T	T	T	T	T	T	T	T	T			
	10																						T	T	T	T	T	T	T	T	T	T	T	T	T		
	11																							T	T	T	T	T	T	T	T	T	T	T	T	T	
	12																								T	T	T	T	T	T	T	T	T	T	T	T	T
1947	1																									T	T	T	T	T	T	T	T	T	T	T	T

Notes: Observation period is from 2005m1 till 2007m12. Control group includes individuals born between 1945m1 and 1945m12, while treatment group includes individuals born between 1946m1 and 1947m1. The grey area marks the period 2006m2 to 2007m12 during which the UB I reform was in effect.

The exposed changes to the UB I scheme are likely to affect individual behavior across birth cohorts differently. This implies that trends in labor market outcomes could be changing across birth cohorts over time for reasons unrelated to the ERA increase, which would violate the identifying assumption. I explicitly control for the changes in benefits generosity by including in the model a dummy variable equal to one if an individual is affected by the UB I reform in month t, and zero otherwise, interacted with year-month fixed effects as explained in Section 5.4.2. This allows to identify the effects of the ERA increase separately from the effects of changes in the UI incentives. Reconciling my findings with Manoli and Weber (2016) who do not detect any program substitution in their sample highly attached to the labor market, I conclude that individuals with relatively long employment careers in my sample adjust to the increased ERA by delaying withdrawals from the labor market rather than substituting early retirement with unemployment benefits or disability pensions.

5.7 Robustness Checks

In this section, I analyze the sensitivity of the RKD estimates and provide further evidence for the validity of the research design. To this end, I present estimates based on alternative bandwidths in Section 5.7.1, a higher polynomial order in Section 5.7.2, and a placebo sample composed of individuals unaffected by the pension reform in Section 5.7.3.

5.7.1 Alternative Bandwidths

In Table 5.8, I explore the sensitivity of the main estimates obtained from parametric estimation of RKD (Section 5.6.2) to alternative bandwidths. As in the main specification, the fuzzy RKD estimates are based on local linear polynomial estimation and a uniform kernel. While I can not choose an arbitrarily large bandwidth given a ceiling of a symmetric bandwidth of 12 months around the kink point, I apply smaller symmetric bandwidths ranging from 11 to 7 months around the kink point.

Table 5.8: Fuzzy Regression Kink Estimates for Alternative Bandwidths

	Local Bandwidth					
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Employment			
ERA	0.116***	0.147***	0.139*	0.069	0.153*	0.131
	(0.039)	(0.052)	(0.069)	(0.083)	(0.087)	(0.107)
Mean	0.512	0.511	0.510	0.507	0.506	0.504
S.d.	0.500	0.500	0.500	0.500	0.500	0.500
			B. Retirement Benefits			
ERA	-0.150***	-0.163***	-0.126***	-0.120***	-0.105**	-0.061

	(0.035)	(0.038)	(0.039)	(0.041)	(0.049)	(0.054)
Mean	0.111	0.112	0.115	0.115	0.118	0.119
S.d.	0.314	0.315	0.320	0.320	0.323	0.324
			C. Disability Benefits			
ERA	-0.012	0.012	-0.024	-0.038	-0.081*	-0.075
	(0.030)	(0.034)	(0.033)	(0.044)	(0.045)	(0.076)
Mean	0.047	0.048	0.046	0.045	0.044	0.045
S.d.	0.212	0.213	0.209	0.208	0.206	0.208
			D. Unemployment			
ERA	-0.055	-0.094	-0.071	0.005	-0.150	-0.014
	(0.057)	(0.065)	(0.066)	(0.065)	(0.074)	(0.101)
Mean	0.137	0.137	0.140	0.145	0.144	0.146
S.d.	0.344	0.344	0.347	0.352	0.351	0.353
Bandwidth	12	11	10	9	8	7
Individuals	4,060	3,649	3,300	2,941	2,610	2,283
Observations	51,779	46,697	42,213	37,578	33,344	29,187

Notes: The table reports IV estimates of slope changes from fitting linear polynomials in the ERA to the left and the right of the kink threshold (1946m1) and instrumenting the ERA with birth month. The estimates are based on samples with alternative symmetric bandwidths including a bandwidth of 12 months around the kink point (column (1)) applied to obtain the main estimates in Table 5.6. All regressions control for high education at age 50 (dummy), unemployment months at age 50, sick leave months at age 50, compulsory contribution years for insured employment at age 50, yearly average earnings in € adjusted for inflation between ages 50-55, at least one child in 2007 (dummy), number of children in 2007, and female (dummy) and include region fixed effects, occupation at age 50 fixed effects, and a dummy for relevance of the UB I reform interacted with year-month fixed effects. Standard errors in parentheses are clustered at the birth month level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The estimates reported in column (1) are based on the maximum symmetric bandwidth of 12 months around the kink point from the main analysis in Table 5.6 and are included in the table for the ease of comparison. Starting from column (2), the applied bandwidth decreases by one month around the kink point. Overall, the obtained estimates are relatively stable across bandwidths and outcome variables with one exception (Panel C, column (5)). Taken together, I conclude that the main results are not sensitive to the choice of bandwidth in my setting.[16]

[16] Additionally, I apply methods for nonparametric estimation of RKD producing bias-corrected estimates and the associated bandwidth selectors. These include Imbens and Kalyanaraman (2012) bandwidth

5.7.2 Polynomial Order

In Table 5.9, I present regression kink estimates from a quadratic model adopting the same bandwidth of 12 months around the kink point as in the linear model in Table 5.6. Across all outcome variables, applying a higher polynomial order leads to statistically insignificant kink estimates. A closer look at the estimated kinks in columns (1)-(2) reveals, however, that these are similar in size to the corresponding statistically significant coefficient estimates from the linear model, although the estimated standard errors are substantially larger. In line with Card et al. (2015), this finding indicates that using local quadratic polynomials in the RKD is connected to a substantial cost in terms of variance compared to local linear polynomials.[17] In fact, even though a local quadratic regression is likely to have smaller bias than a local linear regression under the same bandwidth, the asymptotic variance for the quadratic regression is much higher than that for the linear specification. In this respect, similar kink estimates from linear and quadratic specifications in this study are interpreted as a reassuring result.

Table 5.9: Fuzzy Regression Kink Estimates from Quadratic Model

	Employment	Retirement Benefits	Disability Benefits	Unemployment
	(1)	(2)	(3)	(4)
ERA	0.169	-0.155	-0.084	-0.009
	(0.181)	(0.119)	(0.101)	(0.170)
Mean	0.512	0.111	0.047	0.137
S.d.	0.500	0.314	0.212	0.344

Notes: The table reports IV estimates of slope changes from fitting quadratic polynomials in the ERA to the left and the right of the kink threshold (1946m1) and instrumenting the ERA with birth month. The estimates are based on a sample with a maximum symmetric bandwidth of 12 months around the kink point. All regressions control for high education at age 50 (dummy), unemployment months at age 50, sick leave months at age 50, compulsory contribution years for insured employment at age 50, yearly average earnings in € adjusted for inflation between ages 50-55, at least one child in 2007 (dummy), number of children in 2007, and female (dummy) and include region fixed effects, occupation at age 50 fixed effects, and a dummy for relevance of the UB I reform interacted with year-month fixed effects. Standard errors in parentheses are clustered at the birth month level. Sample size is 51,779 observations for 4,060 individuals. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

for fuzzy RKD and the "rule-of-thumb" bandwidth based on Fan and Gijbels (1996). By and large, estimates from procedures selecting small bandwidths yield relatively unstable point estimates and have larger standard errors. This is in line with Card et al. (2015), who find in their setting that point estimates stabilize for larger bandwidths but are highly volatile at smaller bandwidth levels.

[17] Card et al. (2015) show that for the same bandwidth an improvement in mean squared error from using a quadratic specification would require the bias to be at least $\sqrt{15}$ times as large as the standard error in the linear specification.

5.7.3 Placebo Sample

To further exclude the functional dependence between the assignment variable and the outcome variables, I perform a placebo test of the regression kink estimates. To this end, I define a placebo sample with less than 15 insurance years. The underlying idea is that older people with few insurance years are not eligible for early retirement via the reformed pension scheme and thus the impact of the policy change for this population group is expected to be zero. Table 5.10 reports the kink estimates from the placebo test. Across all outcome variables, the coefficient estimates based on a relatively small sample are statistically insignificant. In this respect, a small magnitude of the estimates is more reassuring. Altogether, statistical and economic insignificance suggest that the estimation strategy adopted in this study is not simply picking up trends in differences across birth cohorts.

Table 5.10: Fuzzy Regression Kink Estimates for Placebo Sample

	Employment	Retirement Benefits	Disability Benefits	Unemployment
	(1)	(2)	(3)	(4)
ERA	-0.023	0.008	-0.027	-0.008
	(0.103)	(0.012)	(0.035)	(0.106)
Mean	0.577	0.004	0.021	0.275
S.d.	0.494	0.063	0.144	0.446

Notes: The table reports IV estimates of slope changes from fitting linear polynomials in the ERA to the left and the right of the kink threshold (1946m1) and instrumenting the ERA with birth month. The estimates are based on a placebo sample with less than 15 insurance years at age 58 and a maximum symmetric bandwidth of 12 months around the kink point. All regressions control for high education at age 50 (dummy), unemployment months at age 50, sick leave months at age 50, compulsory contribution years for insured employment at age 50, yearly average earnings in € adjusted for inflation between ages 50-55, at least one child in 2007 (dummy), number of children in 2007, and female (dummy) and include region fixed effects, occupation at age 50 fixed effects, and a dummy for relevance of the UB I reform interacted with year-month fixed effects. Standard errors in parentheses are clustered at the birth month level. Sample size is 15,145 observations for 1,310 individuals. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.8 Treatment Effect Heterogeneity

In this section, I investigate the treatment effect heterogeneity along multiple dimensions. First, I examine whether behavioral responses vary by economic sector (Section 5.8.1) and past earnings (Section 5.8.2) as potential determinants of retirement decisions. Then, I test for differential responses by sex (Section 5.8.3). Finally, I examine the persistence of the reform effects, analyzing whether the reform mainly affects labor market behavior prior to reaching the ERA or its effects tend to persist also beyond the ERA (Section 5.8.4).

5.8.1 Differences by Economic Sector

Empirical research suggests that retirement patterns are likely to differ by occupation and sector, mostly for reasons related to workplace injury or illness and job productivity (Mitchell et al., 1988, among others). In line with this evidence, I separate the main sample by economic sector to test whether individuals working in manufacturing sector are to some extent differently affected by the pension reform compared to those working in service sector. The rationale behind this type of heterogeneity analysis is that differences in production process or work-related health may be responsible for potential differences in behavioral responses. In fact, production in the manufacturing sector is often physically demanding, while jobs in the service sector are relatively psychologically demanding, implying a high degree of social interaction and communication skills. Therefore, the relative productivity of older workers in the manufacturing sector is likely to decline at a faster pace compared to that of service sector, suggesting that the age-productivity profile may differ by sector (Skirbekk, 2008). At the same time, differences in health status related to workplace injury or illness may be an alternative mechanism underlying potential differential retirement patterns across sectors. In fact, health plays an important role in determining early retirement behavior as shown in Bound et al. (2010). Building on these pieces of evidence, I analyze which economic sector is more sensitive to a reduction in the retirement options and therefore potentially more likely to seek benefits from other government programs.

Table 5.11: Fuzzy Regression Kink Estimates by Economic Sector

	Employment (1)	Retirement Benefits (2)	Disability Benefits (3)	Unemployment (4)
		A. Service Sector		
ERA	-0.004	-0.188**	-0.047	-0.001
	(0.124)	(0.090)	(0.073)	(0.104)
Mean	0.564	0.101	0.051	0.151
S.d.	0.496	0.301	0.219	0.358
		B. Manufacturing Sector		
ERA	0.054	-0.002	0.075**	-0.023
	(0.094)	(0.047)	(0.028)	(0.079)
Mean	0.501	0.144	0.023	0.116
S.d.	0.500	0.351	0.150	0.320

Notes: The table reports IV estimates of slope changes from fitting linear polynomials in the ERA to the left and the right of the kink threshold (1946m1) and instrumenting the ERA with birth month. The estimates are based on a sample with a maximum symmetric bandwidth of 12 months around the kink point. Panel A (Panel B) includes workers in service (manufacturing) sector at age 58 or shortly before. All regressions control for high education at age 50 (dummy), unemployment months at age 50, sick leave months at age 50, compulsory contribution years for insured employment at age 50, yearly average earnings in € adjusted for inflation between ages 50-55, at least one child in 2007 (dummy), number of children in 2007, and female (dummy) and include region fixed effects, occupation at age 50 fixed effects, and a dummy for relevance of the UB I reform interacted with year-month fixed effects. Standard errors in parentheses are clustered at the birth month level. Sample size is 17,168 observations for 1,343 individuals in Panel A and 14,463 observations for 1,124 individuals in Panel B. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.11 presents the regression kink estimates for service sector in Panel A and manufacturing sector in Panel B. It follows that across all outcome variables except for retirement in column (2), workers in service sector are rather unresponsive to the enacted pension reform. Small in size and statistically insignificant coefficients in this population group further indicate that its labor supply and retirement decisions including the choice to retire through other programs such as disability insurance or unemployment insurance program are unaffected by the pension policy. The large statistically significant negative effect on claim of retirement benefits in column (2) suggests, however, that the implemented ERA increase has a strong impact on retirement via the associated old-age pension scheme, which reflects the mechanical effect of the policy change.

The estimates in Panel B suggest that among workers in manufacturing sector, postponement of entry into retirement via the old-age pension scheme statistically significantly affects their claim of disability pension benefits. Given that these benefits are granted to individuals with health problems, the large response in column (3) indicates that workers in manufacturing sector with poor health are likely to be to some extent protected from the potential negative effects of the ERA increase by claiming benefits from the disability insurance program.

5.8.2 Differences by Previous Earnings

There is a strand of literature evaluating the importance of economic variables in determining transitions in and out of work with implications for exits to retirement (Boldrin et al., 1999, among others). For example, Meghir and Whitehouse (1997) show that increased earnings in work delay job exit, while increased social security benefits delay the return to work. Taken together, the two effects imply that economic incentives may be important determinants of the retirement age. Based on empirical studies documenting a strong relationship between past earnings and retirement decisions (Staubli and Zweimüller, 2013, among others), I investigate how past earnings interact with the policy change under study. To this end, I separate the analysis sample in a subsample with earnings below the median and a subsample with earnings above the median in Table 5.12.

Table 5.12: Fuzzy Regression Kink Estimates by Earnings

	Employment (1)	Retirement Benefits (2)	Disability Benefits (3)	Unemployment (4)
		A. Below Median		
ERA	0.114	-0.174***	-0.109	-0.045
	(0.093)	(0.060)	(0.088)	(0.132)
Mean	0.526	0.063	0.077	0.192
S.d.	0.500	0.244	0.267	0.393
		B. Above Median		
ERA	0.112**	-0.055	0.010	-0.066
	(0.047)	(0.033)	(0.020)	(0.047)
Mean	0.497	0.159	0.017	0.083
S.d.	0.500	0.366	0.128	0.276

Notes: The table reports IV estimates of slope changes from fitting linear polynomials in the ERA to the left and the right of the kink threshold (1946m1) and instrumenting the ERA with birth month. The estimates are based on a sample with a maximum symmetric bandwidth of 12 months around the kink point. Panel A (Panel B) includes individuals with below (above) median yearly average earnings in € adjusted for inflation between ages 50-55. All regressions control for high education at age 50 (dummy), unemployment months at age 50, sick leave months at age 50, compulsory contribution years for insured employment at age 50, at least one child in 2007 (dummy), number of children in 2007, and female (dummy) and include region fixed effects, occupation at age 50 fixed effects, and a dummy for relevance of the UB I reform interacted with year-month fixed effects. Standard errors in parentheses are clustered at the birth month level. Sample size is 25,884 observations for 2,051 individuals in Panel A and 25,882 observations for 2,008 individuals in Panel B. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Th estimates in Panel A suggest that the ERA reform is unlikely to foster employment (column (1)) among low-wage workers. The only statistically significant coefficient estimate is again for claim of retirement benefits in column (2), which suggests that postponement of entry into retirement via the old-age pension scheme mechanically decreases claims of the associated pension benefits. Noteworthy are, however, statistically insignificant kink estimates in columns (3)-(4), indicating that low-wage workers in the sample do not seek benefits from other insurance programs.

Unlike low-wage workers in Panel A, high-wage workers in Panel B respond to the ERA increase by delaying exits from employment. Their employment response is relatively large and statistically significant. This finding might reflect stronger incentives of workers with high past earnings to work longer compared to low-wage workers.

5.8.3 Differences by Sex

Labor market responses to the ERA increase may differ by sex given the different employment biographies of men and women. Moreover, public pension systems often subsidize short employment careers of women due to e.g. child-raising periods offering attractive pathways to retirement exclusively for women. This applies also to the German case. In fact, women in Germany may apply for the old-age pension for women, which allows the qualified women to retire earlier at the ERA of 60 years. In 2018, about 3,550,000 (ca. 20% of all old-age pensions) claimants were women retired via the old-age pension for women with an average benefits level of about €870 per month (Deutsche Rentenversicherung, 2019b). I investigate how this factor interacts with the policy change under study. To this end, I separate the analysis sample in a subsample with only women and a subsample with only men in Table 5.13.

The estimates in Panel A suggest that the ERA reform is unlikely to foster employment (column (1)) among women. The only statistically significant coefficient estimate is again for claim of retirement benefits in column (2), which suggests that postponement of entry into retirement via the old-age pension scheme mechanically decreases claims of the associated pension benefits for women. Furthermore, statistically insignificant kink estimates in columns (3)-(4), indicate that women in the sample do not seek benefits from other insurance programs. A potential explanation for these results is that women might receive benefits from the old-age pension for women or rely on other sources of income, e.g. spouse's or partner's income. On the contrary, men in Panel B respond to the ERA increase by delaying exits from employment. This finding might reflect stronger incentives of men to work longer due to more stable economic careers.

Table 5.13: Fuzzy Regression Kink Estimates by Sex

	Employment (1)	Retirement Benefits (2)	Disability Benefits (3)	Unemployment (4)
		A. Women		
ERA	0.358	-0.392***	0.000	-0.363
	(0.243)	(0.137)	(0.132)	(0.217)
Mean	0.579	0.052	0.053	0.142
S.d.	0.494	0.222	0.223	0.350
		B. Men		
ERA	0.069*	-0.094**	-0.023	-0.010

	(0.038)	(0.036)	(0.030)	(0.039)
Mean	0.467	0.150	0.043	0.134
S.d.	0.500	0.358	0.204	0.340

Notes: The table reports IV estimates of slope changes from fitting linear polynomials in the ERA to the left and the right of the kink threshold (1946m1) and instrumenting the ERA with birth month. The estimates are based on a sample with a maximum symmetric bandwidth of 12 months around the kink point. Panel A (Panel B) includes only women (men). All regressions control for high education at age 50 (dummy), unemployment months at age 50, sick leave months at age 50, compulsory contribution years for insured employment at age 50, at least one child in 2007 (dummy), number of children in 2007, and female (dummy) and include region fixed effects, occupation at age 50 fixed effects, and a dummy for relevance of the UB I reform interacted with year-month fixed effects. Standard errors in parentheses are clustered at the birth month level. Sample size is 20,640 observations for 1,626 women in Panel A and 31,139 observations for 2,434 men in Panel B. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.8.4 Persistence of Reform Effects

The main regression kink estimates (see Section 5.6.2) reflect a composed effect of the ERA increase combining the reform effects before and after the ERA. However, empirical evidence suggests that the increase in ERA may affect individual labor market behavior prior to and upon reaching the ERA differently (Staubli and Zweimüller, 2013). In this respect, it is interesting to investigate to what extent the main estimates are driven by individual responses below the ERA (hereafter, BELOW ERA), or, in other words, to test whether the labor market effects caused by the pension reform persist upon reaching the ERA (hereafter, ABOVE ERA).

Table 5.14: Fuzzy Regression Kink Estimates, Above ERA vs Below ERA

	Employment (1)	Retirement Benefits (2)	Disability Benefits (3)	Unemployment (4)
		A. Above ERA		
ERA	0.109*	-0.140**	-0.059	0.007
	(0.054)	(0.061)	(0.036)	(0.079)
Mean	0.504	0.122	0.047	0.135
S.d.	0.500	0.328	0.212	0.342
		B. Below ERA		
ERA	0.117**	-0.140***	-0.022	-0.084
	(0.046)	(0.036)	(0.031)	(0.060)

Mean	0.516	0.106	0.047	0.136
S.d.	0.500	0.306	0.212	0.343

Notes: The table reports IV estimates of slope changes from fitting linear polynomials in the ERA to the left and the right of the kink threshold (1946m1) and instrumenting the ERA with birth month. The estimates are based on a sample with a maximum symmetric bandwidth of 12 months around the kink point. Panel A (Panel B) includes observations prior to (upon) reaching the ERA. All regressions control for high education at age 50 (dummy), unemployment months at age 50, sick leave months at age 50, compulsory contribution years for insured employment at age 50, yearly average earnings in € adjusted for inflation between ages 50-55, at least one child in 2007 (dummy), number of children in 2007, and female (dummy) and include region fixed effects, occupation at age 50 fixed effects, and a dummy for relevance of the UB I reform interacted with year-month fixed effects. Standard errors in parentheses are clustered at the birth month level. Sample size is 41,689 observations for 4,052 individuals in Panel A and 34,213 observations for 3,651 individuals in Panel B. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The estimates reported in Table 5.14 suggest that decisions to enroll into disability insurance (column (3)) or unemployment insurance (column (4)) programs are unaffected by the pension reform in both subsamples, while their labor supply and retirement decisions follow an interesting pattern. In this respect, while the negative impact on claim of retirement benefits across panels is very similar, there is a small, though not statistically significant, difference in the employment response. In other words, the positive employment response might be stronger for individuals not yet eligible to the old-age pension benefits (Panel B). A longer time span for a larger sample would help gain a more precise picture of the effects difference in this respect, but this type of analysis is beyond the scope of this paper.

5.9 Conclusion

In this paper, I evaluate a German pension reform to gain novel insights into labor supply and retirement decisions of older people. The reform gradually raised the early retirement age over multiple birth cohorts, thus providing an exogenous variation in the access to early retirement. The estimation is based on a Regression Kink Design using high-quality administrative data from the Federal Employment Agency.

I show that that the reform had large effects on employment and retirement benefits claims. I find that a one-year increase in the early retirement age leads to an increase in employment by 11.6 percentage points (22.7%) and a decrease in retirement benefits claims by 15 percentage points (135%), on average. I do not find evidence for substitution with other social insurance programs. Workers highly attached to the labor market in my sample are not more likely to claim unemployment benefits or disability pensions. They are rather likely to adjust to the postponed access to early retirement by delaying employment exits. My results also show that some population groups are more sensitive to a reduction in retirement options and more likely to seek benefits from other government programs. In this respect, I find that workers in manufacturing sector respond to the raised early retirement age by claiming benefits from the disability insurance program designed to compensate for reduced earnings capacity due to severe health problems. The treatment heterogeneity analysis further suggests that high-wage workers are more likely to delay exits from employment, which is in line with incentives but might also indicate an increased inequality within the affected birth cohorts induced by the reform. Finally, women seem to rely on alternative sources of income such as retirement benefits for women, or spouse's or partner's income not observed in the data.

Public pension systems in many countries are confronted with the question of a long-term financing of pension payments. This paper shows that raising the early retirement age is an effective policy tool to increase employment among older people who have the real choice to delay employment exits. This implies that such a policy measure might yield unintended heterogeneous effects given that the ability to stay longer at work is likely to depend on socio-economic status. Therefore, reforms that raise statutory ages should ensure social support for workers only marginally attached to the labor market or not able to work longer due to potential health problems or other circumstances. Finally, my results are based on analysis of the short-run effects of the increase in the early retirement age. At the same time, the reform effects in the longer-run may differ if younger birth cohorts are forward looking and internalize changes in pension policy in their life-time decisions. This is a potential avenue for future research.

References

Atalay, K. and G. 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*(1), 71–87.

Boldrin, M., S. Jiménez-Martín, and F. Peracchi (1999). Social security and retirement in Spain. In *Social Security and Retirement around the World*, pp. 305–353. University of Chicago Press.

Borghans, L., A. C. Gielen, and E. F. Luttmer (2014). Social support substitution and the earnings rebound: Evidence from a regression discontinuity in disability insurance reform. *American Economic Journal: Economic Policy 6*(4), 34–70.

Börsch-Supan, A. and R. Schnabel (1999). Social security and retirement in Germany. In *Social Security and Retirement around the World*, NBER Chapters, pp. 135–180. National Bureau of Economic Research, Inc.

Börsch-Supan, A. H. and C. B. Wilke (2004). The German public pension system: How it was, how it will be. *NBER Working Paper 10525, National Bureau of Economic Research*.

Bound, J., T. Stinebrickner, and T. Waidmann (2010). Health, economic resources and the work decisions of older men. *Journal of Econometrics 156*(1), 106–129.

Brown, K. M. (2013). The link between pensions and retirement timing: Lessons from California teachers. *Journal of Public Economics 98*, 1–14.

Burtless, G. (1986). Social security, unanticipated benefit increases, and the timing of retirement. *The Review of Economic Studies 53*(5), 781–805.

Card, D., A. Johnston, P. Leung, A. Mas, and Z. Pei (2015). The effect of unemployment benefits on the duration of unemployment insurance receipt: New evidence from a regression kink design in Missouri, 2003-2013. *American Economic Review 105*(5), 126–30.

Card, D., D. S. Lee, Z. Pei, and A. Weber (2015). Inference on causal effects in a generalized regression kink design. *Econometrica 83*(6), 2453–2483.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica 80*(3), 969–1018.

Dahlberg, M., E. Mörk, J. Rattsø, and H. Ågren (2008). Using a discontinuous grant rule to identify the effect of grants on local taxes and spending. *Journal of Public Economics 92*(12), 2320–2335.

Deutsche Rentenversicherung (2019a). https://www.deutsche-rentenversicherung.de/DRV/DE/Ueber-uns-und-Presse/Presse/Meldungen/2019/191216_aenderungen_2020.html, last accessed on January 13, 2020.

Deutsche Rentenversicherung (2019b). Rente 2018. *Statistik der Deutschen Rentenversichrung, Band 215*.

Deutsche Rentenversicherung (2019c). Rentenversicherung in Zeitreihen 2019. *DRV-Schriften 22*.

Dlugosz, S., G. Stephan, and R. A. Wilke (2014). Fixing the leak: Unemployment incidence before and after a major reform of unemployment benefits in Germany. *German Economic Review 15*(3), 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*(7-8), 1327–1350.

Engels, B., J. Geyer, and P. Haan (2017). Pension incentives and early retirement. *Labour Economics 47*, 216–231.

Fan, J. and I. Gijbels (1996). Local polynomial modelling and its applications. *London: Chapman and Hall*.

French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. *The Review of Economic Studies 72*(2), 395–427.

Geyer, J. and C. Welteke (2019). Closing routes to retirement for women: How do they respond? *Journal of Human Resources*. doi:10.3368/jhr.56.1.0717-8947R2.

Giesecke, M. and M. Kind (2013). Bridge unemployment in Germany: Response in labour supply to an increased early retirement age. *Ruhr Economic Paper Nr. 410*.

Gruber, J. and D. A. Wise (2007). Introduction to" Social security programs and retirement around the world: Fiscal implications of reform". In *Social Security Programs and Retirement around the World: Fiscal Implications of Reform*, pp. 1–42. University of Chicago Press.

Guryan, J. (2001). Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts. *NBER Working Paper 8269, National Bureau of Economic Research*.

Gustman, A. L. and T. L. Steinmeier (1985). The 1983 social security reforms and labor supply adjustments of older individuals in the long run. *Journal of Labor Economics 3*(2), 237–253.

Gustman, A. L. and T. L. Steinmeier (2005). The social security early entitlement age in a structural model of retirement and wealth. *Journal of Public Economics 89*(2-3), 441–463.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Hochfellner, D., D. Müller, and A. Wurdack (2012). Biographical data of social insurance agencies in Germany: Improving the content of administrative data. *Schmollers Jahrbuch: Zeitschrift für Wirtschafts-und Sozialwissenschaften 132*(3), 443–451.

Hunt, J. (1995). The effect of unemployment compensation on unemployment duration in Germany. *Journal of Labor Economics 13*(1), 88–120.

Imbens, G. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies 79*(3), 933–959.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

Karlström, A., M. Palme, and I. Svensson (2008). The employment effect of stricter rules for eligibility for DI: Evidence from a natural experiment in Sweden. *Journal of Public Economics 92*(10-11), 2071–2082.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*, 435–464.

Kleven, H. J. and M. Waseem (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *The Quarterly Journal of Economics 128*(2), 669–723.

Landais, C. (2015). Assessing the welfare effects of unemployment benefits using the regression kink design. *American Economic Journal: Economic Policy 7*(4), 243–78.

Manoli, D. S. and A. Weber (2016). The effects of the early retirement age on retirement decisions. *NBER Working Paper 22561, National Bureau of Economic Research.*

Mastrobuoni, G. (2009). Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities. *Journal of Public Economics 11*(93), 1224–1233.

Meghir, C. and E. Whitehouse (1997). Labour market transitions and retirement of men in the UK. *Journal of Econometrics 79*(2), 327–354.

Mitchell, O. S., P. B. Levine, and S. Pozzebon (1988). Retirement differences by industry and occupation. *The Gerontologist 28*(4), 545–551.

Mitchell, O. S. and J. W. Phillips (2000). Retirement responses to early social security benefit reductions. *NBER Working Paper 7963, National Bureau of Economic Research.*

Nielsen, H. S., T. Sørensen, and C. Taber (2010). Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform. *American Economic Journal: Economic Policy 2*(2), 185–215.

OECD (2017). Pensions at a glance 2017: OECD and G20 indicators. OECD Publishing, Paris.

OECD (2018). Database on average effective retirement age. http://www.oecd.org/employment/emp/average-effective-age-of-retirement.htm, last accessed on January 16, 2020.

Oguzoglu, U., C. Polidano, and H. Vu (2016). Impacts from delaying access to retirement benefits on welfare receipt and expenditure: Evidence from a natural experiment. *IZA Discussion Paper 10014, Institute of Labor Economics.*

Pellizzari, M. (2006). Unemployment duration and the interactions between unemployment insurance and social assistance. *Labour Economics 13*(6), 773–798.

Rust, J. and C. Phelan (1997). How social security and medicare affect retirement behavior in a world of incomplete markets. *Econometrica 65*(4), 781–831.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy 180*, 212.

Seibold, A. (2017). Statutory ages as reference points for retirement: Evidence from Germany. *International Institute of Public Finance.*

Simonsen, M., L. Skipper, and N. Skipper (2016). Price sensitivity of demand for prescription drugs: Exploiting a regression kink design. *Journal of Applied Econometrics 31*(2), 320–337.

Skirbekk, V. (2008). Age and productivity capacity: Descriptions, causes and policy options. *Ageing Horizons 8*, 4–12.

Staubli, S. (2011). The impact of stricter criteria for disability insurance on labor force participation. *Journal of Public Economics 95*(9-10), 1223–1235.

Staubli, S. and J. Zweimüller (2013). Does raising the early retirement age increase employment of older workers? *Journal of Public Economics 108*, 17–32.

Ye, H. (2018). The effect of pension subsidies on retirement timing of older women: Evidence from a regression kink design. *IZA Discussion Paper 11831, Institute of Labor Economics.*

Appendix 5.A

Pathways to Retirement

Disability Pension

Disability pension *(Erwerbsminderungsrente)* allows people unable to work due to severe health problems covering illness and disability to apply for benefits compensating for their reduced earnings capacity.[18] Based on a medical assessment of individual employability and following the principle of "rehabilitation before pension", the German Pension Insurance thoroughly examines all applications to evaluate whether a rehabilitation treatment can maintain or improve a claimant's earnings capacity. Upon fulfillment of the applicable requirements, entitled individuals can claim the associated benefits. Given that no age threshold applies, disability pension is the only pathway to retirement available before age 60. Eligibility conditions include at least 5 insurance years, also known as waiting period or qualifying period, and 3 years with paid compulsory contributions for insured employment in the last 5 years prior to retirement. Before the disability pension reform in January 2001, eligibility for disability pension was tied to individual earnings capacity. Full benefits were granted to individuals with a substantial limitation in their ability to work defined as earnings capacity of 15% or less of average gross earnings *(Erwerbsunfähigkeitsrente)*. Partial benefits equivalent to 66% of full benefits were granted to individuals unable to find a job within the same occupational profile and with a substantial reduction in earnings of at least 50% when changing occupation *(Berufsunfähigkeitsrente)*.

After the reform in January 2001, the only relevant parameter for benefits calculation is the number of daily hours an individual can work for at least 6 months under the normal labor market conditions. In this respect, claimants unable to work more than 6 hours per day are entitled to full benefits, while claimants unable to work more than 3 hours per day are entitled to partial benefits. Based on the principle of "legitimate expectations", cohorts born on or before January 1, 1964 and unable to work in their previous occupation can exceptionally apply for partial benefits as well and are therefore the only population group for whom occupational attachment remains preserved under the new regime. In both pension schemes, benefits are calculated based on the previous employment biography and are equal to an amount paid as if a claimant continued to work until age 60. Moreover, pension benefits are subject to actuarial deductions of a maximum of 10.8% with a reference age for deductions of 63. As individuals mainly claim disability pension before age 60, most pensions are exactly reduced by 10.8% (Deutsche Rentenversicherung, 2019c). In 2018, about 97,000 disability pensions were paid to compensate for partially reduced earnings capacity and amounted to about €528 on average, while more than 1,715,000 disability pensions with full benefits werde paid with an average benefits level of €812 (Deutsche Rentenversicherung, 2019b). Upon reaching the statutory retirement age, the disability

[18] For institutional details, see § 43 SGB VI in combination with §§ 240 and 241 SGB VI.

pension is converted into an old-age pension, usually with the same benefits level.

Old-Age Pension for Women

Old-age pension for women *(Altersrente für Frauen)* offers an attractive pathway to retirement given that it subsidizes short employment careers of women due to e.g. child-raising periods. In fact, it is the most common old-age pension claimed following the regular old-age pension. In 2018, about 3,550,000 (ca. 20% of all old-age pensions) claimants were women retired via the old-age pension for women with an average benefits level of about €870 per month (Deutsche Rentenversicherung, 2019b). This scheme is available to women with at least 15 insurance years, also known as waiting period or qualifying period, and 10 years with paid compulsory contributions for insured employment upon reaching age 40.[19] This program is phased out by 2016 implying that the birth cohort December 1951 is the last eligible one. For women born in or after January 1940 the NRA gradually increased from 60 to 65 years. As a result, benefits eligibility requires women born in or after December 1944 to be at least 65 years old. Old-age pension for women also allows to retire earlier, with an ERA of 60 years applied to all birth cohorts. Given that early retirement entails permanent pension deductions of 0.3% per month for each month before the statutory NRA, this implies that pension benefits can be permanenetly reduced by at most 18%.

Old-Age Pension for Severely Disabled People

Old-age pension for severely disabled people *(Altersrente für Schwerbehinderte)* allows workers with at least 35 insurance years, also known as waiting period or qualifying period, to retire due to severe disability, i.e. at least 50% disability at retirement entry.[20] In 2018, among all old-age pensions the take-up of the old-age pension for severely disabled people was nearly 10% corresponding to 1,835,000 pensions with an average benefits level of about €1,150 per month (Deutsche Rentenversicherung, 2019b). This pension scheme enables disabled people to receive benefits at the NRA or the ERA with permanent pension deductions of 0.3% per month for each month before the statutory NRA. The retirement age thresholds vary across birth cohorts, with younger cohorts experiencing gradual increases in both parameters. Hence, for claimants born before January 1952, the NRA is 63 years, while the ERA is 60 years. This implies that pension benefits can be permanently reduced by at most 10.8%. Both age thresholds gradually increase for claimants born in or after January 1952, reaching 65 years and 62 years for the NRA and the ERA, respectively, starting from cohorts born in January 1964.

[19]For institutional details, see § 237a SGB VI in combination with Annex 20 SGB VI.
[20]For institutional details, see § 37 SGB VI in combination with § 236a SGB VI.

Old-Age Pension for Miners

Old-age pension for miners *(Altersrente für langjährig unter Tage beschäftigte Bergleute)* was introduced to account for miners' hard working conditions. This pension scheme is not as relevant as other old-age pensions and was claimed by about 35,500 miners with an average benefits level of about €2,075 per month in 2018 (Deutsche Rentenversicherung, 2019b). Upon fulfillment of the applicable requirements, this pension allows miners to retire at age 60 at the earliest. Thus, claimants born before January 1952 are entitled to pension benefits at age 60, while for claimants born in or after January 1952 the earliest retirement age gradually increased in monthly steps reaching age 62. As a result, workers born in or after January 1964 can be entitled to this pension at age 62 at the earliest. Apart from the miner occupation, the only eligibility condition is a waiting period, also known as qualifying period, of 25 insurance years. Noteworthy is also another kind of pension available to miners with reduced capacity to work, which is marked by relaxed qualifying conditions *(Rente für Bergleute)*.[21] This pension scheme is even less relevant and was claimed by almost 10,00 miners with an average benefits level of only €530 per month in 2018 (Deutsche Rentenversicherung, 2019b).

Old-Age Pension for People with a Long Insurance Record

Old-age pension for people with a long insurance record *(Altersrente für langjährig Versicherte)* is granted to individuals with long employment biographies and requires a waiting period, also known as qualifying period, of 35 insurance years. Among almost 18,250,000 old-age pensioners, almost 2,000,000 claimants retired via the old-age pension for the long-term insured with an average benefits level of about €1,100 per month in 2018 (Deutsche Rentenversicherung, 2019b). This pension scheme allows to retire at the NRA or the ERA with pension deductions of 0.3% per month for each month before the statutory NRA. While the ERA remains constant at age 63 across birth cohorts, the NRA differs. In this respect, claimants born before January 1949 can apply for pension benefits without deductions at age 65, while those born in or after January 1949 experience a gradual increase in the NRA 67 for claimants born in or after January 1964.[22] Given the required long-term insurance period, this pension has a major impact on male retirement behavior. Nevertheless, as the sample age is between 60 and 61 years, this pension is irrelevant for the main analysis. The same applies to the old-age pension for people with an exceptionally long insurance record *(Altersrente für besonders langjährig Versicherte)* introduced with the pension reform 2012. This pension scheme allows selected birth cohorts (claimants born before January 1953) to retire at age 63 without deductions upon fulfillment of the waiting period, also known as qualifying period, of 45 insurance years.[23] Due to a very

[21] For institutional details, see § 40 SGB VI in combination with § 238 SGB VI and § 45 SGB VI in combination with § 242 SGB VI.

[22] For institutional details, see § 36 SGB VI in combination with § 236 SGB VI.

[23] For institutional details, see § 38 SGB VI in combination with § 236b SGB VI.

long insurance period required, only about 1,160,000 retirees took up this pension scheme though with a higher average benefits level of about €1,300 per month in 2018 (Deutsche Rentenversicherung, 2019b).

Regular Old-Age Pension

Regular old-age pension *(Regelaltersrente)* is granted to individuals upon reaching the statutory retirement age and fulfillment of the waiting period, also known as qualifying period, of 5 insurance years. Cohorts born before January 1947 can apply for pension benefits at age 65, while those born in or after January 1947 experience a gradual increase in the retirement age threshold reaching age 67 for claimants born in or after January 1964.[24] Early retirement via this pension scheme is not possible. Regular old-age pension was claimed by almost 7,700,000 retirees with an average benefits level of about €640 per month in 2018 (Deutsche Rentenversicherung, 2019b). This is the only pension scheme for which additional earnings on top of the received pension benefits are irrelevant for benefits level. Contrary to other pensions paid up to the statutory retirement age, this rule implies no pension deductions due to potentially high supplementary income. Given that sample age is definitely lower than the required age threshold, this pension is irrelevant for the main analysis.

[24] For institutional details, see § 35 SGB VI in combination with § 235 SGB VI.

Appendix 5.B

Table 5.B1: Normal Retirement Age (NRA)

Birth Cohort		NRA	
1937	January	60	1
	February	60	2
	March	60	3
	April	60	4
	May	60	5
	June	60	6
	July	60	7
	August	60	8
	September	60	9
	October	60	10
	November	60	11
	December	61	0
1938	January	61	1
	February	61	2
	March	61	3
	April	61	4
	May	61	5
	June	61	6
	July	61	7
	August	61	8
	September	61	9
	October	61	10
	November	61	11
	December	62	0
1939	January	62	1
	February	62	2
	March	62	3
	April	62	4
	May	62	5
	June	62	6
	July	62	7
	August	62	8
	September	62	9
	October	62	10
	November	62	11
	December	63	0
1940	January	63	1

	February	63	2
	March	63	3
	April	63	4
	May	63	5
	June	63	6
	July	63	7
	August	63	8
	September	63	9
	October	63	10
	November	63	11
	December	64	0
1941	January	64	1
	February	64	2
	March	64	3
	April	64	4
	May	64	5
	June	64	6
	July	64	7
	August	64	8
	September	64	9
	October	64	10
	November	64	11
	December	65	0
1942 -1951		65	0

Source: Annex 19 SGB VI, own illustration.

Table 5.B2: Early Retirement Age (ERA)

Birth Cohort		ERA	
1946	January	60	1
	February	60	2
	March	60	3
	April	60	4
	May	60	5
	June	60	6
	July	60	7
	August	60	8
	September	60	9
	October	60	10
	November	60	11
	December	61	0

1947	January	61	1
	February	61	2
	March	61	3
	April	61	4
	May	61	5
	June	61	6
	July	61	7
	August	61	8
	September	61	9
	October	61	10
	November	61	11
	December	62	0
1948	January	62	1
	February	62	2
	March	62	3
	April	62	4
	May	62	5
	June	62	6
	July	62	7
	August	62	8
	September	62	9
	October	62	10
	November	62	11
	December	63	0
1949 - 1951		63	0

Source: Annex 19 SGB VI, own illustration.

Appendix 5.C

Details on Sample Construction

Identification of Individuals Exempted from the ERA Increase

For individuals with legitimate expectations the pension reform framed an exemption from the ERA increase. This implies that among the affected cohorts born in or after 1946m1, there is a small population group with an ERA of 60. Given that individuals exempted from the ERA increase are not directly detectable because not classified as such in BASiD, I exploit detailed individual employment biographies to recognize this population group in the data. To all individuals classified as exempted from the ERA increase according to the following method, I assign an ERA of 60.

To detect i) individuals unemployed on January 1, 2004, I mark the respective unemployment spells which cover registered unemployment without necessarily receiving the associated benefits among the affected birth cohorts. Then, to detect ii) individuals whose employment relationship ends before January 1, 2004 and who are non-employed (i.e. unemployed but not jet registered unemployed) on January 1, 2004, I mark among the affected birth cohorts individuals whose employment relationship ends between December 18, 2003 and December 31, 2003 with the subsequent employment relationship starting between January 2, 2004 and January 14, 2004. The underlying idea is that unregistered unemployment is not observed in the process data and an unemployment registration at the Employment Office within a minimum of two weeks and a maximum of four weeks after a job loss is assumed.[25]

In the same spirit, to detect iii) individuals whose employment relationship ends after December 31, 2003 due its temporary nature (including fixed-term training measures) or notice given before January 1, 2004, I apply the following step-wise procedure. Given that training measures are inherently temporary, I mark individuals participating in training measures which start in or before December 31, 2003 and end after December 31, 2003 among the affected birth cohorts. To identify the relevant employment relationships, more steps are required. First, for each employment relationship I calculate the respective tenure, with vocational training spells increasing tenure duration. To this end, gaps in employment at the same plant up to six months are allowed. Then, based on tenure duration I calculate the respective statutory period of notice and mark among the affected birth cohorts individuals whose employment relationship ends after December 31, 2003 with the computed

[25]The choice of this time span is motivated by statutory disqualification periods used as a major sanctioning instrument in the unemployment compensation system. The duration of a disqualification period ranges between one week in case of failure to report to the job office or delay to register as unemployed and twelve weeks in case of voluntary unemployment or a repeated decline of a job offer or a repeated refusal to participate in a job integration program. For details on disqualification periods for infringement of the rules, consult the German Social Security Legislation *(Sozialgesetzbuch, SGB)*, § 148 and § 159 SGB III.

period of notice due on or before December 31, 2003.[26] Unfortunately, I can not observe in the data whether an employment relationship is open-ended or temporary. Finally, I also mark iv) individuals ever been employed as miners or receiving pension benefits for miners.

[26]For details, consult the German Civil Code *(Bürgerliches Gesetzbuch, BGB)*, §622.