Essays in Empirical Labor and Demographic Economics

Patrick Reich

Essays in Empirical Labor and Demographic Economics

Inaugural-Dissertation

zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Volkswirtschaftlichen Fakultät
der Ludwig-Maximilians-Universität München

2020

vorgelegt von

Patrick Reich

Referent:	Prof. Helmut Rainer, Ph.D
Korreferent:	Prof. Thomas Siedler, Ph.D.
Mündliche Prüfung:	22. Januar 2021
Berichterstatter:	Prof. Helmut Rainer, Ph.D
	Prof. Thomas Siedler, Ph.D.
	Prof. Dr. Joachim Winter
Promotionsabschlussberatung:	3. Februar 2021

Acknowledgments

First and foremost, I would like to thank my supervisor Helmut Rainer for his enthusiastic encouragement and valuable discussions. During the past years, I have learned a lot and especially enjoyed our joint work on one chapter of this thesis. I would also like to thank Thomas Siedler for agreeing to be my second supervisor and Joachim Winter for joining the examination committee.

I thank my coauthors Natalia Danzer and Timo Hener for the long-lasting collaboration and their genuine advice. I am grateful to my coauthor Helena Holmlund for the joint and fruitful work and for facilitating my research visit at Uppsala University. I thank Eva Mörk for hosting me at the Department of Economics and the faculty, particularly the Ph.D. students, for the warm welcome there.

Throughout my thesis, I have benefited from financial support of the Leibniz Association, which is gratefully acknowledged. Additionally, I want to thank the employees at the Research Data Center of the Statistical Offices of the Länder for their assistance in accessing the hospital register data.

Furthermore, I thank my current and former colleagues at the ifo Institute and the Munich Graduate School of Economics (MGSE) for their help and support. I am grateful to all fellow Ph.D. students who were part of the journey and who have made the challenging times less painful and the good times even better. In particular, I want to thank Andreas, Eleonora, Fabian, Marc, and Victoria.

Last but not least, I am fortunate to have the support of my family and good friends. I am very grateful for having you and would not have achieved this without you.

Contents

List of Figures

List of Tables

Preface

The advancement of econometric methods and increased availability of detailed administrative data have strongly improved the scope to answer policy-relevant research questions. Over the last decades, empirical work has become more prominent in economic research, particularly in labor and demographic economics (van der Klaauw, 2014; Angrist et al., 2017). This thesis contributes to the literature by applying microeconometric techniques to uncover causal effects and answering questions of interest for public policy. It consists of three self-contained chapters that can be read separately. In Chapter 1, which is joint work with Natalia Danzer and Timo Hener, we investigate the effect of changes in local economic conditions on the health of newborns. Chapter 2 studies the impact of seasonal time changes on work-related accidents. In Chapter 3, which is joint work with Helena Holmlund and Helmut Rainer, we aim at uncovering how cultural dimensions affect gender gaps in educational achievement. The following paragraphs provide a summary of the three chapters.

Chapter 1 assesses the relationship between changes in local economic conditions and newborn health. According to the widely accepted fetal origins hypothesis, fetal and early-life conditions and health shocks are highly important for later life outcomes (Almond and Currie, 2011a,b; Almond et al., 2018). Seminal studies for the US and other European countries report a positive impact of recessions and increasing annual unemployment rates on the health of newborns (e.g., Dehejia and Lleras-Muney, 2004; van den Berg et al., 2020; Aparicio et al., 2020). On the contrary, we show that local unemployment can lead to negative health effects at birth. Using German data on a new set of newborn health outcomes, we find that unemployment increases adverse health conditions that originate in the perinatal period. In addition to information from the birth registry, we draw on administrative hospital data observing any health condition at birth, including mild forms. We match monthly unemployment rates of local labor markets to the hospital and birth data, and construct a measure of average unemployment during pregnancy for all outcomes. Moreover, we create an unemployment measure for the three trimesters of pregnancy and the month prior to conception. Controlling for month-by-year, and local labor market fixed effects, as well as labor market-specific time trends, we estimate the effect of local unemployment deviations from national developments and long-term regional trends on newborn health. We find that health at birth is deteriorating with local unemployment. In fact, we detect an increase in perinatal health problems and a decrease in birth weight. The negative health effects materialize mostly as a response to unemployment in the third trimester. When exploring mechanisms, we cannot attribute the results to parental selection, in-utero selection, maternal hospital visits during pregnancy, hospital-capacity induced diagnoses, or environmental pollution. Instead, we observe an increase in maternal health problems during labor and delivery, potentially related to maternal stress and nutrition. Our results point towards additional health costs of economic downturns that have previously been overlooked.

Chapter 2 studies the effect of seasonal time changes on work-related accidents. Following the population's repugnance of adjusting clocks twice a year, the European Parliament decided to phase out switching between Standard Time and Daylight Saving Time (DST) in the European Union (European Parliament, 2019a). While the initial aim of introducing DST transitions was to reduce energy consumption, the general debate often centers on potentially detrimental effects in other areas, such as health, crime, and traffic collisions. I contribute to the public discussion by investigating the impact of seasonal time changes on work-related accidents in Germany. Using data from the German Social Accident Insurance, I complement the literature by considering commuting accidents as an explicit outcome. The discrete nature of the time change allows me to obtain causal estimates of the transitions into as well as out of DST. After demeaning work-related accidents by systematic calendar and weather effects, I estimate the effect of the DST transition in a regression discontinuity design. My results indicate no systematic influence of the DST regime on workplace or commuting accidents over the period 2013-2017. I identify insignificant positive estimates (for an increase in accidents) in spring when switching to DST and coefficients close to zero when reverting to Standard Time in fall. The estimates of the transition to DST on accidents during commuting stand out but lack precision. Given that the sizable estimates are stable in magnitude across a rich set of alternative specifications, there is suggestive evidence for an increase in commuting accidents during the first days of Daylight Saving Time.

Chapter 3 aims at uncovering how cultural dimensions affect gender gaps in educational achievement. Although geographical and temporal variations in gender achievement gaps have received considerable attention (Pope and Sydnor, 2010; Reardon et al., 2019; Evans et al., 2019), the role of culture in explaining this variation is not well understood. We exploit a large Swedish administrative data set to study gender gaps in education among second-generation immigrant youth with different cultural backgrounds. Guided by hypotheses we derive from the economics literature on gender differences and gender convergence (e.g., Goldin, 2006; Bertrand, 2018; Niederle and Vesterlund, 2010), we explore the predictive power of a set of cultural dimensions including achievement orientation, acceptance of inequality, risk avoidance, and long-term orientation (Hofstede et al., 2010). Our empirical strategy relies on within-family, cross-gender sibling comparisons, identifying culture's differential impact on girls relative to boys while netting out unobserved family heterogeneity and controlling, *inter alia*, for gender-specific neighborhood effects. We find that the central cultural dimension that matters for gender gaps in student achievement is the extent to which a society emphasizes ambition, competition, and achievement, which is strongly predictive of a relative achievement disadvantage of girls compared with boys. The ground lost by girls relative to boys when moving from weak to strong achievement orientation is more than half as large as the test score gap between second-generation immigrants and natives, which is strongly in favor of natives in Sweden. Cultural dimensions other than achievement orientation are less prominent in explaining gender achievement gaps. Exploring factors that may explain the results, we find that parents from achievement-oriented cultures choose higher quality schools for their children, and that boys benefit more from exposure to higher quality schools than girls do. Using PISA data to probe external validity, we find qualitatively and quantitatively remarkably similar results in a very different sample of second-generation immigrant youth.

Chapter 1

Local Labor Markets and Health at Birth*

1.1 Introduction

Economic fluctuations, and recessions in particular, can have consequences that reach far beyond job loss. A number of papers have shown that the health of adults improves in recessions (Ruhm, 2000, 2003). An emerging literature investigates accordingly, whether recessions also affect newborn health outcomes. According to the widely accepted fetal origins hypothesis, fetal and early-life conditions and health shocks are highly important for later life outcomes (Almond and Currie, 2011a,b; Almond et al., 2018). In their seminal paper, Dehejia and Lleras-Muney (2004) show that infant mortality and the incidence of low birth weight decrease with local unemployment in the United States. Common explanations for the counterintuitive positive health effects are beneficial health behavior and advantageous parental selection in economic downturns.

We show in this paper that local unemployment can lead to negative health effects at birth. Using data from Germany on a new set of newborn health outcomes different from the previous literature, we find that unemployment increases adverse health conditions that originate in the perinatal period. Our results point towards additional health costs of economic downturns that have previously been overlooked. Combining our preferred estimates with official figures on cost of illness, we calculate that a rise in the local unemployment rate during pregnancy by one percentage point increases aggregated hospital costs in the range of 15.6 to 23.9 Million Euro (in 2015). This corresponds to about 1.2% to 1.8% of all hospital costs related to conditions originating in the perinatal period, the time period between the completed 22nd week of pregnancy and the 7th day after birth.

Our estimation builds on outcome variables from two register datasets. The hospital register includes every main diagnosis from German hospitals, including information on the age and

*This chapter is based on joint work with Natalia Danzer and Timo Hener.

gender of the patient, along with an indicator of the involved hospital, the admission date, and the precise four-digit diagnosis code. While hospital register data often come with the caveat that they only contain the most severe medical cases, this caveat is not relevant for newborn health. The hospital data are particularly well suited for the description of newborn health, because almost *all* births occur in hospitals and *all* newborns are examined. Therefore, we observe any health condition, including mild forms. We use perinatal health problems, congenital defects, and neonatal mortality as main outcomes, and explore sub-diagnoses of these to investigate the underlying mechanisms.

From the birth register, we extract information on average birth weight, incidence of low birth weight, birth length, and length-adjusted birth weight as additional outcomes. We match the monthly unemployment rate of the local labor market from administrative data to the hospital and birth data, and construct a measure of average unemployment during pregnancy for all outcomes. Additionally, we use the monthly information to create an unemployment measure for the three trimesters of pregnancy and the month prior to conception. Controlling for month-by-year, local labor market fixed effects, and labor market-specific time trends, we estimate the effect of local unemployment deviations from national developments and long-term regional trends on newborn health.

Overall, we find that newborn health is not improving but deteriorating with local unemployment in the period from 2005 to 2013. Most importantly, we detect an increase in perinatal health problems and a decrease in birth weight. A one percentage point increase in the unemployment rate yields an increase in perinatal health problems by about 3.6 to 5.5 cases (per 1,000 newborns) and a 0.076% to 0.094% decline in length-adjusted birth weight. The negative health effects materialize mostly as a response to unemployment in the third trimester (and partly also in the second trimester). We establish further that the weight decrease is not accompanied by changes in length at birth. Both the late trimester effect and the lack of a birth length change are consistent with deficient nutrition hampering intrauterine growth (Kramer, 1987; Bozzoli and Quintana-Domeque, 2014). We do not see any effects on congenital defects or neonatal mortality. Using auxiliary data, we investigate a number of potential mechanisms that connect local unemployment and newborn health. Looking into the health of becoming mothers, we do not find any relation between unemployment and hospital visits during pregnancy or miscarriages. However, exposure to rising unemployment in the third trimester increases the likelihood of problems during labor and delivery (in cases per 1,000 live births). In the birth register data, we find a small reduction in fertility, but no compositional change in parental characteristics that could explain the negative health effects. Likewise, we establish a reduction in air pollution from unemployment, which does not suffice as an explanation for increasing morbidity.

We contribute to the literature on the effects of economic fluctuations on newborn health that was inspired by the work of Ruhm (2000, 2003) on the cyclicality of adult mortality and health. Dehejia and Lleras-Muney (2004) show in US data that higher state unemployment rates at conception are associated with fewer cases of low birth weight and congenital malformations, and reduced neonatal and postneonatal mortality. They also reveal that improved newborn health is most likely due to beneficial health behavior during pregnancy in bad economic times. Subsequently, Menclova (2013) confirmed the positive infant health effects for county unemployment, and Lindo (2015) showed that they are robust to different levels of aggregation from coarser regions to counties. Orsini and Avendano (2015) show that the positive health effects from state unemployment are stronger in the years 1980–1989 than in 1990–2004 and more pronounced for blacks than for whites. Recent evidence from European countries corroborates the results from the US. For Sweden, van den Berg et al. (2020) show positive health effects of unemployment on newborns that are robust to the inclusion of parental fixed effects and not driven by own unemployment. For Spain, Aparicio et al. (2020) find similar positive health effects that are mainly explained by fewer first births and positive parental selection in downturns.

Our results also contribute to and are consistent with the literature on the effects of severe economic crises. In contrast to the positive health effects found in developed countries using fluctuations in unemployment rates during relative economic stability, the literature on deep recessions and severe economic shocks finds unilaterally negative effects on newborn health. The Argentinian economic crisis in 2002, as Bozzoli and Quintana-Domeque (2014) show, led to a decrease in birth weight and an increase in the incidence of low birth weight. Eiríksdóttir et al. (2013) identify an increased incidence of low birth weight and being small-for-gestational-age in newborns 6–9 months after the 2008 economic collapse in Iceland. Similarly, Olafsson (2016) finds a higher prevalence of low birth weight children from mothers who were hit by the economic collapse in the first trimester. Even for the US, Carlson (2015) uncovers that mass layoffs and plant closures as more severe shocks to local unemployment lead to decreased birth weight and more incidences of low birth weight. Kaplan et al. (2017) show that unemployment variation from the Great Recession in the Memphis, Tennessee area led to an increased incidence of low birth weight. For India, Bhalotra (2010) makes a similar account of an adverse health effect of economic shocks measured by income changes. Infant mortality increases, and health-seeking decreases, although at-risk mothers tend to have fewer children in economic downturns. Negative effects on birth weight are also observed for a different type of economic shock, a month-long blackout in Zanzibar, Tanzania (Burlando, 2014). These negative health effects at birth are consistent with the findings from the literature on the effects of own unemployment; e.g., Lindo (2011) shows that parental job loss reduces birth weight.

The only paper providing a similar negative effect of regular variation in local unemployment on newborns' health is the recent De Cao et al. (2019) paper for England, although their study period also includes the Great Recession. The authors find a significant negative effect of unemployment on birth weight, consistent with our results. They also show in contrast to Dehejia and Lleras-Muney (2004) that mothers show detrimental health behavior in downturns, increasing smoking and drinking. While there are similarities between our studies, there are also important differences. While De Cao et al. (2019) show that without mother fixed effects their results would agree with positive health effects from recessions, our results reveal a negative association even without controlling for selection into parenthood. Moreover, our sample period includes the Great Recession, which had little effect on the unemployment rate in Germany. Our study is furthermore the first to show that negative health consequences of unemployment materialize in less severe conditions than usually investigated. We provide new results on perinatal health problems, in particular neonatal jaundice, whereas most studies focus on indicators of low birth weight and forms of neonatal mortality (Dehejia and Lleras-Muney, 2004; Bozzoli and Quintana-Domeque, 2014; Aparicio et al., 2020; Olafsson, 2016; van den Berg et al., 2020; De Cao et al., 2019).

Chapter 1.2 of this paper describes the data, chapter 1.3 explains the estimation strategy, and chapter 1.4 discusses the main results, potential mechanisms, and effect heterogeneity. Chapter 1.5 estimates the economic burden of the health effects, and chapter 1.6 concludes.

1.2 Data

1.2.1 Health Outcomes and Control Variables

Our analysis is based on an panel data set that we create by combining hospital register data with birth register data, local labor market indicators, further regional control variables, and pollution data for the years 2005 to 2013.

The micro-level hospital register data contain detailed, 4-digit-level information on the main diagnosis for the universe of all German inpatient cases (17-19 million per year).[1] It also provides details on the patient's age (in months), gender, place of residence, the dates of admission and discharge (daily information), the incidence of surgery and death, the specialist department in which the patient spent most of the time, and a hospital identifier.

Since 2004, diagnoses are coded according to the German Modification (GM) of the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10). The ICD is maintained by the World Health Organization (WHO). ICD-10 diagnosis codes in the register data are grouped into 21 chapters (see Table 1.1). Official statistics distinguish between health-related diagnoses ("All diseases and sequelae of effects of external causes (A00-T98)") and the diagnosis codes related to other factors (mainly "Factors influencing health status and contact with health services (Z00-Z99)"). Importantly, for patients below the age of one month, the vast majority of these Z-codes identify cases of healthy liveborn infants (Z38 "Liveborn infants according to place of birth and type of delivery"; see Table A.9 in the Appendix). According to our data on hospital diagnoses of newborns, there are two dominant diagnosis chapters which together make up about 93% of all newborn health-related cases: "Certain conditions originating in the perinatal period, P00-P94, P96[2]" (85.0 percent) and "Congenital malformations, deformations and chromosomal abnormalities, Q00-Q99" (7.9 percent).

[1] Each stationary hospital stay results in one case, even when the person is discharged the same day. A change of the specialist department within a hospital, however, does not lead to a new case.

[2] We exclude stillbirths (P95) from this group, which indicate that a fetus is born without any signs of life. We consider stillbirths as a separate outcome variable in Section 1.4.2.

Table 1.1: Overview ICD 10 Hospital Main Diagnosis Codes of Newborns (2005-2013)

Chapter	Diagnosis codes	Description	Cases per 1000 live births	Std. Dev.	in % (of I-XIX)
A. All diseases and sequelae of effects of external causes (A00-T98)					
I	A00-B99	Certain infectious and parasitic diseases	3.2	5.2	1.1
II	C00-D48	Neoplasms	0.7	1.9	0.2
III	D50-D90	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	0.2	1.1	0.1
IV	E00-E90	Endocrine, nutritional and metabolic diseases	1.1	2.5	0.4
V	F00-F99	Mental, Behavioral and Neurodevelopmental disorders	0.2	1.3	0.1
VI	G00-G99	Diseases of the nervous system	0.6	1.8	0.2
VII	H00-H59	Diseases of the eye and adnexa	0.3	1.1	0.1
VIII	H60-H95	Diseases of the ear and mastoid process	0.3	1.5	0.1
IX	I00-I99	Diseases of the circulatory system	0.8	1.9	0.3
X	J00-J99	Diseases of the respiratory system	4.3	6.3	1.4
XI	K00-K93	Diseases of the digestive system	2.3	3.4	0.8
XII	L00-L99	Diseases of the skin and subcutaneous tissue	1.2	2.6	0.4
XIII	M00-M99	Diseases of the musculoskeletal system and connective tissue	0.1	0.7	0.0
XIV	N00-N99	Diseases of the genitourinary system	0.9	2.0	0.3
XV	O00-O99	Pregnancy, childbirth and the puerperium			
XVI	P00-P94, P96	Certain conditions originating in the perinatal period	256.4	62.2	85.0
XVII	Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities	23.8	17.3	7.9
XVIII	R00-R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	3.8	4.7	1.3
XIX	S00-T98	Injury, poisoning and certain other consequences of external causes	1.7	2.8	0.6
I - XIX		**All health-related diagnoses**	**301.7**	**70.9**	**100.0**
B. Other codes*					
XXI	Z00-Z99	Factors influencing health status and contact with health services	**717.8**		
	Z38	*Liveborn infants according to place of birth and type of delivery*	*704.2*	*94.8*	
XXII	U00-U99	Codes for particular cases/purposes	0		
XXI-XXII		**All other factors**	**717.8**		
All diagnoses/reasons for contact with health services			**1,019.5**		

Notes: Summary statistics of the German hospital register data of main diagnoses for newborns (zero months old patients) for the years 2005 to 2013. *ICD-10 chapter XX (V01-Y84, External causes of morbidity and mortality) is not included in the register data.

Building on the 4-digit alphanumeric main diagnosis codes, we define several health related outcome variables (see Section A.3 for all constructed outcome measures and the respective ICD-10 codes.). Our most general health measure is a binary variable indicating all newborn hospital cases with a health-related diagnosis (*Health-related diagnosis*) (A00-T98) (all other, not health-related cases, i.e., healthy live births, are coded as 0). In addition, we focus on two more refined indicators capturing (i) health-related cases that have their origin in the perinatal period[3] (*Perinatal health*) and (ii) congenital malformations (*Congenital defects*). These two diagnosis chapters were the leading causes of infant deaths (under the age of 1 year) in Germany in 2015 (see Table A.9 in the Appendix). Furthermore, we construct indicators for neonatal mortality (death in the first 28 days of life, *Neonatal mortality*) and pre-term births (*Perinatal: Pre-term* which is a subcategory of *Perinatal health* problems). In further analyses, we exploit the data even further and analyze the incidence of even more specific diagnoses (see Appendix A.3).

Apart from health-related diagnoses of newborns, we also construct measures of maternal health during pregnancy and delivery as well as miscarriages based on the hospital register data. Health problems of the mother during pregancy or delivery can potentially affect a newborn's health. Unfortunately, the data do not allow us linking mothers to children.

While the hospital register data have the advantage of providing detailed health information for the universe of hospital cases, there are also some shortcomings and limitations. First, we do not observe individuals but cases. Since the data do neither contain exact date of birth nor individual identifiers, we cannot account for the possibility that some individuals appear in the sample more than once. However, since we only consider diagnoses during the first month of life, a relatively short period, the likelihood of repeated hospital stays should not be extensive.[4] A second important point to note is that the register data do only contain main diagnoses, but not secondary or additional diagnoses. Hence, we only capture and assess the main health problem, not all health problems.[5]

In addition to the hospital register data, we use the German birth registry for the years 2005 to 2013 as further source on newborn health, fertility, and parental characteristics. As we were only allowed to merge the birth and hospital registers via regional identifiers, we aggregated the birth register micro data on the regional level. Variables extracted from the birth register are monthly, regional data on the average birth weight, birth length and ponderal index ($\frac{\text{weight}}{\text{height}^3}$ in

[3]The perinatal period refers to the time period between the completed 22nd week of pregnancy and the 7th day after birth.

[4]In principle, a rise in health-related diagnoses does not necessarily imply a higher incidence of newborns with health problems (extensive margin), but could also stem from worsened health problems of unhealthy newborns (intensive margin) which translate into repeated hospital stays during the first month of life. However, our results and tests do not indicate that our findings are merely caused by repeated hospital visits.)

[5]For instance, in the US, 91.3 percent of infant deaths in the year 2002 that were classified as "Newborn affected by maternal complications of pregnancy (P01)" were also born preterm (Callaghan et al., 2006).

$\frac{\text{kg}}{\text{m}^3}$)[6], as well as the number of births[7] and the average age of the mother. Moreover, the data contain quarterly information on the number of newborns with low birth weight ($<$2,700 g)[8] and further measures related to the parity of newborns and the age and marital status of parents. All count variables are expressed as number of cases per 1,000 live births or 1,000 women between 15 and 44 years of age (or the respective age sub-group).

Moreover, we merge regional statistics from the Federal and Regional Statistical Offices.[9] These include information on the demographic structure of a county, in particular, data on population size and its composition with respect to age, gender, migration background (fraction of population with German citzienship), internal migration across counties, and on the share of school leavers with university-entrance degree and without any degree.

As cyclical variations in industrial production might affect the degree of air pollution, we include several environmental indicators. We extracted daily measures of temperature, precipitation and ambient air pollution (carbon monoxide, sulfur dioxide, nitrogen dioxide and particulate matter) from monitors across the country. Weather and pollution measures are spatially interpolated for the centroid of each county using inverse distance weighting. Using this data, we construct measures of air pollution and weather for local labor markets at the level of days for auxiliary regressions.

1.2.2 Local Labor Markets and Regional Unemployment

The regional level of our analysis is the local labor market (LLM), following the 2014 classification of the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR).[10] A LLM consists of one or more counties linked by a commuter flow, thereby offering jobs for at least 65% of the labor force and allowing for a one-way commuting-time up to 45 minutes. This aggregation level should allow us to capture spillover effects between counties while retaining statistical power (Lindo, 2015). Three East-German regions conducted major district reforms during our observation period.[11] To generate consistent regional classification over time, we group these particular counties into one broader region. That is why our number of regions reduces from 258 to 246 LLMs. Additionally, we merge two very small LLMs in Thuringia (data protection regulations of the German Birth Register require a minimum cell size and number of cases), leaving us with a total of 245 LLMs. Section A.4 in the Appendix provides a list of the LLMs. The corresponding map is shown in Figure A.4.

[6] In contrast to the more well-known Body-Mass-Index, PI is better suited for very small (and very large) body heights. PI-values between 11 and 14 are considered normal weight for adults.)

[7] This number refers to births to mothers aged 15 to 44 years. The age restriction is due to data protection guidelines. The share of births born to mothers between 15 to 44 years of age in Germany, however, is at least 99.69% per year (in the years 2009 to 2013).

[8] The standard cut-off for low-birth weight is 2500g; however, we had to choose 2,700 g due to data protection reasons due to too few cases.

[9] The data were collected between March and May 2016 and in May 2017 from the following websites: `https://www.regionalstatistik.de/genesis/online/`, `https://www.destatis.de/`

[10] In official documents LLM are referred to as travel-to-work-areas.

[11] Saxony-Anhalt in July 2007, Saxony in August 2008, and Mecklenburg-Vorpommern in September 2011

Our measure of local labor market conditions is the monthly regional unemployment rate, released by the German Federal Agency for Employment (BA).[12] It relates the number of registered unemployed to the dependent civilian labor force.[13] This measure is consistently available from January 2005 onwards, when the labor market reform Hartz IV came into effect.[14] Originally all unemployment rates were based on the civilian workforce in dependent employment. Since January 2009, however, the BA publishes primarily unemployment rates related to the total civilian labor force. This renders all gender- and age-specific unemployment rates time-inconsistent. Only the general unemployment rate (related to the dependent civilian labor force) is continuously provided in a consistent way.

Figure A.1 in the Appendix plots the change in the annual unemployment rate for the years 2005-2013. The German labor market experienced a notable improvement between 2005 and 2008, with the average annual unemployment rate falling from 12.85% to 8.67%. After a small increase in 2009, the unemployment rate decreased further in the following years, stabilizing around 7.7% in 2012 and 2013. Additionally, the figure reveals a compression of the distribution, indicating that high-unemployment regions experienced a steeper decline in unemployment than low-unemployment regions. Figure A.2 depicts a strong regional gradient in the unemployment rate in our sample. While the labor market regions in East Germany are in the upper deciles of the unemployment distribution, labor markets with low unemployment are typically located in south of Germany. As we account in our estimation for time-invariant variation between labor markets and common variation over time (see Section 1.3 for details), this pattern gets more dispersed. Figure A.3 shows the mean absolute residual variation in the unemployment rate as exploited in our preferred specification which additionally accounts for LLM-specific trends. Summary statistics of our sample reveal that the average monthly local unemployment rate is on average at 8.99 percent (with a standard deviation of 4.04 pp, see Table A.1 in the Appendix).

1.2.3 Sample

We base our analysis on the period from 2005 to 2013 for which we have access to harmonized unemployment data and birth register data. In order to assess the effect of local economic conditions on newborn health accurately, we take advantage of the monthly information in our data and link health outcomes at birth to regional unemployment rates that prevailed before pregnancy and during different stages of gestation (three trimesters).

This differentiation with respect to time is important for two reasons. First, the decision to procreate may be influenced by economic conditions around or before potential conception. This might affect the total number of births as well as the composition of births in terms of socioeconomic parental background. Thus, variation in local health outcomes at birth might simply reflect differential selection of parents into pregnancy. Second, fetus development and

[12] Downloaded from `https://statistik.arbeitsagentur.de/` in May 2017

[13] The size of the labor force is updated once a year.

[14] There are missing values for two regions (Düsseldorf and Mönchengladbach) in February 2010, which we impute by the average of the preceding and the following month.

survival may be affected by in-utero exposure to economic conditions due to maternal stress, maternal health-related behavior, and environmental conditions.[15] The impact of these might vary over the period of gestation. This is why our empirical setup aims at disentangling effects due to selection into pregnancy from those caused by in-utero exposure to local economic shocks.

As the hospital register data do not contain exact birth date information, we classify all hospital cases with zero months of age at the time of admission as newborn cases. We define the nine months preceding the month of birth as the *pregnancy period* and the month before this as the *"selection period"*.[16] We also consider a more refined classification in which we further split the *pregnancy period* into trimesters (three time periods with three months each). We merge average local unemployment rates prevailing at each of these different time periods.[17]

Hence, our sample contains all newborn cases in the period of November 2005 to December 2013 (the sample starts in November 2005 due to the 10-month lag of the unemployment rate prevailing in the month before conception). We drop observations with missing information on the main diagnosis, age, gender or date of admission. In total, our sample includes 98 months of observation for 245 LLMs. It covers 5,531,003 newborn hospital cases with a patient aged 0 months at the time of admission.[18]

Table 1.2 shows summary statistics on the main outcome and composition variables. We observe on average 1005.9 newborn hospital cases per 1,000 registered live births. 30.2 percent of these cases reflect health-related diagnoses (301.7 cases per 1,000 live births), while 70.4 percent of these cases are "Liveborn infants according to place of birth" (704.2 cases[19] per 1,000 live births). As discussed above, the number of hospital cases of newborns slightly exceeds the number of live births - by 5.9 cases per 1,000 live births - and arises from newborns with multiple hospital admissions in their first month of life.

The largest fraction of main newborn health diagnoses refer to conditions originating in the perinatal period (256.4 cases per 1,000 live births). Out of these, the most common main diagnoses are children with low birth weight, respiratory health problems, and neonatal jaundice (49.0, 34.3 and 31.6 cases per 1,000 live births respectively). Around 23.8 cases per 1,000 live births are related to congenital defects. Neonatal deaths occur on average at a rate of 2.4 newborns per 1,000 live births. Health outcomes in the birth registry reveal that the average regional birth weight is at about 3,333 grams (average regional birth weight) and the number of children with less than 2,700 g amounts to 108.4 per 1,000 live births . This number is much

[15]See Section 1.4.2 for a more detailed discussion of potential transmission channels.

[16]The register data do not contain length of gestation which we would ideally use to classify these pregnancy periods. On average, a pregnancy lasts around 38 weeks after conception or 40 weeks after the first day of the woman's last menstrual period. Thus, our *pregnancy period* includes the month of conception. We are aware that our classification based on months since birth introduces some measurement error. However, the results are relatively similar, when the *pregnancy period* includes the month of birth or the *"selection period"* spans over 3 months.

[17]We exclude the month of birth from our definition of the *pregnancy period* to be sure that we not assess the effect of unemployment prevailing postpartum (i.e., for children born at the beginning of a month).

[18]We drop hospital cases with Z diagnoses other than Z38 which are neither classified as "health-related diagnoses" (A-T) nor "liveborn infant" (Z38).

[19]This number includes newborns who are born outside of a hospital.

Table 1.2: Descriptive Statistics of Outcome Variables

	Mean	Std. Dev.	Min	Max	N
1. Newborn health: Hospital register data					
1.A. Main hospital diagnosis: Individual case level data (per 1,000 newborn cases)					
Health-related diagnosis	299.5	458.0	0	1	5,531,003
Perinatal health (P)	254.6	435.7	0	1	5,531,003
P-Subcategories:					5,531,003
Perinatal: Neon. Jaundice	31.1	173.7	0	1	5,531,003
Perinatal: Maternal	8.8	93.5	0	1	5,531,003
Perinatal: Low b.w.	48.7	215.2	0	1	5,531,003
Perinatal: Pre-term	25.7	158.2	0	1	5,531,003
Perinatal: Respiratory	34.1	181.4	0	1	5,531,003
Perinatal: Infectious	18.9	136.3	0	1	5,531,003
Congenital defects	23.6	151.8	0	1	5,531,003
Neonatal mortality	2.4	49.1	0	1	5,531,003
1.B. Main hospital diagnosis: Ratio of number of diagnoses (per 1,000 live births/LLM)					
Health-related diagnosis	301.7	70.9	20.4	820.5	24,010
Perinatal health (P)	256.4	62.2	0	704.2	24,010
P-Subcategories:					
Perinatal: Neon. Jaundice	31.6	20.2	0	314.3	24,010
Perinatal: Maternal	8.9	12.8	0	276.3	24,010
Perinatal: Low birth weight	49.0	20.4	0	243.9	24,010
Perinatal: Pre-term	25.9	15.5	0	178.9	24,010
Perinatal: Respiratory	34.3	17.4	0	282.1	24,010
Perinatal: Infectious	19.0	13.8	0	277.8	24,010
Congenital defects	23.8	17.3	0	265.6	24,010
Neonatal mortality	2.4	3.5	0	69.8	24,010
Liveborn infant	704.2	0.50			24,010
2. Newborn health: Birth registry data (per 1,000 live births/LLM)					
Birth weight (in g)	3332.8	51.6	2,996.4	3,662.6	24,010
Low birth weight (<2,700g)*	108.4	15.3	36.5	223.2	8,085
Average Ponderal index (in kg/m^3)	24.9	0.6	22.5	32.3	24,010
Average birth length (in cm)	51.1	0.5	48.7	53.0	24,010
3. Maternal health: Ratio of hospital diagnoses of mothers (per 1,000 live births/LLM)					
Delivery	926.3	80.2	252.5	2,136.0	24,010
Delivery (Complications)	572.3	76.3	116.3	1,476.0	24,010
Delivery (Particular maternal care)	354.0	73.8	47.6	1,051.0	24,010
Pregnancy	234.7	78.4	0	1,028.0	23,275
Miscarriage	15.1	10.9	0	133.3	22,540
Stillbirth	0.3	1.3	0	39.2	23,275
Infectious, Parasitic diseases (A00-B99)	71.3	31.2	0	647.5	23,275
Mood [affective] disorders (F30-F39)	76.8	34.2	0	423.1	23,275

Notes: Sample descriptives. Data sources: German register of hospital diagnoses (2005-2013); German Birth registry data (2005-2013). LLM statistics are weighted by the average number of live births. * The smaller sample size of this outcome arises due to data confidentiality restrictions which require aggregration at the quarterly, not the monthly level.

higher than the one in the hospital diagnosis data, because some "low-birth weight" children will have another main diagnosis (a more severe, or specific health problem) and because the threshold of 2,700 g is higher than in the hospital data. The average birth length is 51 cm and the mean Ponderal Index is 24.9.

The summary table furthermore contains maternal hospital diagnosis data related to pregnancy and delivery. Per 1,000 live births there are on average 926.3 hospital health-related diagnoses related to delivery (this measure does not include "delivery" as main diagnosis). 572.3 of these cases are due to complications during labor or delivery, while the remaining 354 cases indicate "maternal care related to the fetus and amniotic cavity and possible delivery problems". On average, there are 234.7 pregnancy-related diagnoses, 15.1 cases of miscarriages (i.e., spontaneous abortions), and 0.3 cases of stillbirth as main diagnosis per 1,000 live births.

1.3 Estimation Strategy

We estimate the effect of local unemployment on the health of newborns in two distinct models and in several specifications. In the first model, the units of observation are individual hospital cases for every main diagnosis of zero months old children in German hospitals. This framework allows us to asses gender-specific health effects and to account for time-invariant hospital characteristics by controlling for hospital fixed effects. In contrast, the units of observation in the second model are 24,010 month-LLM cells, containing aggregated health outcomes of newborns (relative to the number of live born children) at the regional level. This way, we can relate hospital diagnoses to number of overall live births occurring in a specific region and month (as indicated by the birth registry). Furthermore, in our analysis and data, newborn health information based on the birth registry (average weight, height) is only available on the regional level.

In particular, when focusing on individual hospital cases, we estimate the following model

$$Diag_{i,c,h,t} = \alpha + \delta UR_{c,\tilde{t}} + X'_{i,c,h,t}\kappa + X'_{c,\tilde{t}}\beta + \eta_c + \nu_h + \theta_t + \gamma_c T_c + \epsilon_{i,c,h,t}, \tag{1.1}$$

where $Diag_{i,c,h,t}$ indicates diagnosis i in LLM c per 1,000 newborn hospital cases, hospital h, and month-by-year t.[20] $Diag_{i,c,h,t}$ takes on the value one for all health-related hospital diagnoses of newborns and is zero otherwise (i.e., if the main diagnosis indicates a healthy newborn.) The in-utero unemployment rate at the LLM level $UR_{c,\tilde{t}}$ is the average of the monthly unemployment rates during the *pregnancy period*, i.e., 1 to 9 months prior to the registration at the hospital.[21] Our coefficient of interest δ therefore measures the effect of a one percentage point increase in the

[20] We define month-by-year as the fully identified combination of year and month, i.e., 12 months X 9 years. We refer to calendar months or month of the year as 12 binary variables for the months January to December.

[21] We match regional unemployment rates during pregnancy according to place of living at the time of birth. Hence, our mapping of local economic conditions during different stages at gestation according to place at the time of birth does not account for the possibility that mothers move across regions during the pregnancy. This potentially introduces measurement error in our measure of unemployment during pregnancy. To account for migration patterns across regions, all our analyses control for regional migration.

unemployment rate on the prevalence of health problems of newborns (per 1,000 hospital cases of healthy and sick newborns). Control variables at the case level $X_{i,c,h,t}$ are indicators of the patient's gender. In-utero regional control variables $X_{c,\tilde{t}}$ include population size and composition with respect to age, gender, migration and educational background as well as migration flows across counties. We include fixed effects for the LLM in η_c, for the hospital in ν_h, and for the month-by-year in θ_t. In an additional specification, we include linear LLM-specific time trends T_c to account for potential secular, region-specific trends. In all estimations (all models and specifications) we cluster standard errors at the level of LLM.

The use of monthly data enables us to disentangle the effect of the in-utero unemployment $UR_{c,\tilde{t}}$ during the entire *pregnancy period* into exposure during different trimesters of pregnancy. Additionally, we include the unemployment rate for the month prior to conception to capture selection into childbearing and potential compositional effects on the health outcomes. This new specification results in the following equation

$$\begin{aligned} Diag_{i,c,h,t} =& \alpha + \sum_{tri=1}^{3} \delta_{tri} UR_{c,tri} + \delta_{t-10} UR_{c,t-10} \\ &+ X'_{i,c,h,t}\kappa + X'_{c,\tilde{t}}\beta + \eta_c + \nu_h + \theta_t + \gamma_c T_c + \epsilon_{i,c,h,t}, \end{aligned} \tag{1.2}$$

where δ_{tri} is the trimester-specific effect of average regional monthly unemployment prevailing during each of the three trimesters tri of pregnancy (3 months in each). Additionally, we include the unemployment rate for the month prior to conception (*"selection period"*), $UR_{c,t-10}$, to capture economic conditions that might affect selection into childbearing and potential compositional effects on the health outcomes. All else remains as in equation 1.2.

The second model is estimated for diagnosis rates at the aggregate LLM and month-by-year level, the same level that constitutes the variation in the treatment. Whereas the model for $Diag_{i,c,h,t}$ relies on identifying single birth incidents,[22] the aggregate model uses the rate of diseases per births in the LLM as the outcome variable in the following equation

$$Y_{c,t} = \alpha + \delta UR_{c,\tilde{t}} + X'_{c,\tilde{t}}\beta + \eta_c + \theta_t + \gamma_c T_c + \epsilon_{c,t}. \tag{1.3}$$

$Y_{c,t}$ is the number of health-related diagnoses of zero months old children per 1,000 registered live births in LLM c and month-by-year t. The specification is as before in equation 1.1, only without individual control variables and hospital fixed effects that cannot be included at the aggregate level. We estimate the trimester-specific effects of unemployment in the aggregate model in equation

$$Y_{c,t} = \alpha + \sum_{tri=1}^{3} \delta_{tri} UR_{c,tri} + \delta_{t-10} UR_{c,t-10} + X'_{c,\tilde{t}}\beta + \eta_c + \theta_t + \gamma_c T_c + \epsilon_{c,t}. \tag{1.4}$$

[22]There is a subtle difference between the two approaches due to measurement of the underlying variables. In principle, the same person can have more than one health diagnosis in both models. In the case based model, disease diagnoses are contrasted with healthy births registered as diagnoses in hospitals. In the aggregate model, we use all births from birth register as the denominator to compute the disease rate.

We further apply the models in equation 1.3 and 1.4 using additional other outcome variables at the LLM level, when we investigate potential mechanisms.

The unemployment rate as the main treatment variable varies at the LLM and month-by-year level. With the inclusion of fixed effects for the LLM and for the month-by-year, the identifying variation in all estimation approaches stems from variation of unemployment rates within LLMs beyond national monthly time variation. In other words, the time fixed effects eliminate all common variation over time and the region fixed effects eliminate time-constant differences between LLMs. What is left is the within LLM variation in the unemployment rate that is not explained by national fluctuations. The variation can be interpreted as local unemployment shocks that include mass layoffs and idiosyncratic firm closures, which should be as good as random if the model correctly specifies all common factors. The estimate is unconfounded if no unobserved, time-variant variable correlates with unemployment net of the fixed effects and determines the outcome.

As discussed in Section 1.2.2, the German labor market experienced a general but heterogeneous improvement over the observation period. The extend of this improvement varied decisively between the different regions, with high-unemployment regions experiencing a stronger reduction in unemployment than low-unemployment regions. To account for these different long-term developments, we additionally phase in LLM-specific linear trends. What is left is the within-LLM variation that is not explained by national fluctuations or regional secular trends.

1.4 Results

1.4.1 Effects of Unemployment on Newborn Health

Our main results indicate that unemployment rates in the local labor market are negatively associated with newborn health. Table 1.3 shows results from models of individual hospital cases in Panel A and aggregated diagnosis rates in Panel B. We present results without and with linear region-specific time trends. As the German labor market experienced a general improvement in most regions during the period of our analysis, the latter specification accounting for these general patterns is our preferred specification.

A rise in the average unemployment rate during pregnancy by one percentage point increases the number of a health-related diagnosis at the time of birth by 3.1 and 6.2 cases (per 1,000 hospital cases) in the specification without and with linear LLM-specific time trends, respectively (panel A). These estimates correspond roughly to an increase in the probability of health problems at birth by 0.3 or 0.6 percentage points.[23] Compared to the mean and standard deviation of this health-related diagnosis, this coefficient implies an increase of 1.1% and 2.1% at the mean or 0.7% and 1.3% of a standard deviation, respectively. Similarly, the results in Panel B using

[23]Our outcome variable is measured as number of diagnoses per 1,000 hospital cases of newborns (including cases with the diagnosis of a healthy live birth). Hence, dividing the coefficient by 10 thus provides a measure of "additional cases" per 100 cases which can be interpreted as increase in percentage points.

aggregated diagnosis rates per 1,000 liveborns, suggest that a rise in local unemployment by one percentage point increases the number of health problems of newborns by 3.4 or 3,9 cases, respectively (an increase of about 1.1% or 1.3 % at the mean). In our preferred specification with trends (column 2), the point estimates are statistically significant from zero (at the 1- and 10-percent level, respectively). The estimates without the trend controls are somewhat smaller, but not statistically different from the point estimates with trends; however, they are not statistically different from zero at conventional thresholds.

As we noted earlier, the hospital diagnosis data reflect cases rather than individuals. This introduces a minor measurement error due to multiple counting of individuals who are admitted to hospitals during the first month of life more than once. To test whether our results merely reflect such multiple treatments, we use as alternative outcome the ratio of cases coded as "Liveborn infant" (ICD-10 code Z38) in all liveborn infants (based on the birth registry as in Panel B). Since an individual can only be coded once per life as "Liveborn infant", this measure does not suffer from potential multiple counting. The results using this outcome reveal that a one percentage point rise in the average local unemployment during pregnancy *decreases* the number of healthy newborns by 4.0 or 7.9 cases per 1,000 liveborns, in the specification without and with trends, respectively (see Appendix, Table A.2). The latter coefficient is significantly different from zero at the 5-% level. It suggests that, per 1pp increase in the unemployment rate, almost one (0.8) in one hundred newborns suffers from adverse health consequences due to in-utero exposure to worsened local economic conditions.

Overall, our findings of pro-cyclical impacts on newborn health, i.e., adverse consequences of unemployment, stand in contrast to much of the literature that found positive health effects of regional unemployment. Turning to the results of the more refined trimester specification, we furthermore see that this negative health effect is not merely driven by a selection into motherhood. Dividing the *pregnancy period* into trimesters reveals that there is a marked negative health impact of unemployment during the third trimester of pregnancy. A one percentage point increase in unemployment towards the end of pregnancy increases the number of health-related diagnoses at birth by 2.9 to 4.1 cases depending on the specification. The estimates are significant at the 5% and 1% level, respectively. In contrast, elevated levels of unemployment prior to conception and during the first two trimesters do not appear to affect health outcomes at birth significantly. This suggests, that in-utero exposure to increased levels of local unemployment towards the end of pregnancy translates into worse health-outcomes at birth. These baseline results are confirmed in the diagnosis rates model in Panel B. Estimates are generally less precise, but the overall pattern and size is very similar. According to the aggregate level analysis, a one percentage point increase in the third trimester unemployment rate increases the incidence of health problems by about 2.2 to 2.9 cases (per 1,000).[24]

[24]Since we lack consistent gender-specific unemployment rates, we perform a background analysis in which we approximate female and male unemployment rates. We combine unemployment statistics with information on employment subject to social insurance and assume the gender shares in dependent employment (needed for the denominator) to be similar to the ones in employment subject to social insurance. The results using these rough gender specific unemployment rates are in line with our main findings (results are reported in the Appendix, Table A.8). Moreover, they do not point towards significant differences in the impact of male or female unemployment

Table 1.3: Impact of Regional Unemployment on Hospital Diagnoses of Newborns

	Newborn health problems							
	Any		Perinatal health		Congenital defects		Neonatal mortality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Hospital cases								
$UR_{\tilde{t}}$	3.1490	6.1591***	3.2173	5.4967***	0.0885	0.5852	-0.0704	-0.0766
	(2.5295)	(2.2848)	(2.2633)	(2.0236)	(0.4625)	(0.4667)	(0.0589)	(0.0660)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
UR_{t-10}	0.7676	1.7960	1.0031	1.6791	0.0079	0.2813	-0.0249	-0.0345
	(1.3655)	(1.2076)	(1.2486)	(1.0820)	(0.2721)	(0.2682)	(0.0646)	(0.0694)
Tri_1	0.7180	1.5121	0.3022	1.0435	0.2495	0.2900	-0.0247	-0.0167
	(1.0292)	(1.0834)	(0.9366)	(0.9922)	(0.2912)	(0.3032)	(0.0746)	(0.0767)
Tri_2	-0.7094	0.3061	-0.3767	0.4484	-0.2261	-0.0712	-0.0831	-0.0820
	(1.1683)	(1.1709)	(1.0673)	(1.0587)	(0.2517)	(0.2523)	(0.0604)	(0.0607)
Tri_3	2.9587**	4.1478***	2.8828**	3.7451***	0.1014	0.3168	0.0713	0.0637
	(1.2687)	(1.2280)	(1.1186)	(1.0801)	(0.2696)	(0.2845)	(0.0605)	(0.0627)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
Mean	299.46	299.46	254.65	254.65	23.6	23.6	2.42	2.42
Standard Deviation	458.02	458.02	435.66	435.66	151.8	151.8	49.1	49.1
Observations	5531003	5531003	5531003	5531003	5531003	5531003	5531003	5531003
Panel B: Diagnosis rates								
$UR_{\tilde{t}}$	3.3811	3.8730*	3.4365	3.5958*	0.2005	0.4886	-0.0724	-0.1001
	(2.5564)	(2.3183)	(2.2594)	(2.0517)	(0.4811)	(0.4996)	(0.0574)	(0.0687)
R^2	.53	.61	.5	.58	.39	.47	.03	.04
UR_{t-10}	1.4732	1.4202	1.5738	1.3541	0.0941	0.2821	-0.0182	-0.0334
	(1.4107)	(1.3098)	(1.2570)	(1.1523)	(0.2880)	(0.2868)	(0.0663)	(0.0706)
Tri_1	0.8647	1.2143	0.4504	0.8035	0.3040	0.3216	-0.0259	-0.0259
	(1.1831)	(1.2247)	(1.0321)	(1.0820)	(0.3108)	(0.3229)	(0.0780)	(0.0809)
Tri_2	-0.5060	-0.3646	-0.1592	-0.0412	-0.2164	-0.1558	-0.0849	-0.0932
	(1.3191)	(1.3299)	(1.1645)	(1.1602)	(0.2694)	(0.2792)	(0.0642)	(0.0651)
Tri_3	2.2229	2.9396**	2.2232*	2.6418**	0.0831	0.2870	0.0670	0.0613
	(1.3798)	(1.3769)	(1.1886)	(1.1916)	(0.3004)	(0.3100)	(0.0611)	(0.0650)
R^2	.53	.61	.5	.58	.4	.47	.03	.04
Mean	301.67	301.67	256.35	256.35	23.79	23.79	2.44	2.44
Standard Deviation	70.87	70.87	62.17	62.17	17.32	17.32	3.49	3.49
Observations	24010	24010	24010	24010	24010	24010	24010	24010
Reg. cont.	X	X	X	X	X	X	X	X
LLM trend		X		X		X		X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. The regressions in Panel A additionally control for hospital fixed effects and gender. The regressions in Panel B are weighted by the average number of live births in a LLM.

The nature of these health problems are better understood by analyzing the impact of unemployment on the prevalence of more specific diagnoses. The results for the diagnoses perinatal health problems and congenital defects - the two leading causes of neonatal deaths - are reported in columns 3 to 6. It turns out that local unemployment affects newborn health primarily through its impact on perinatal health diagnoses. We do not detect any significant effects on congenital defects. According to Panel A, the number of newborns with *health conditions originating in the perinatal period* increases by about 3.2 to 5.5 cases with each additional percentage point increase in unemployment during pregnancy. Again, the estimates in our preferred specification using regional trends are significant at the 1% level. The trimester results resembles the earlier pattern suggesting that in-utero exposure during the third trimester has the largest adverse impact. Conditional on prevailing unemployment before pregnancy and during the first two trimesters, a one percentage point increase in local unemployment towards the end of pregnancy increases perinatal health problems in the range of about 2.2 to 3.7 cases (depending on the specification). Importantly, this coefficient is statistically significant in all four specifications (Panel A and B). In contrast, the estimated coefficients on congenital defects are positive, but their absolute size is much smaller and none of the estimates is statistically different from zero.

Perinatal health problems and congenital defects make up 93% of all newborn health-related diagnoses and are therefore in the focus of our analysis. However, we also assess the effect of unemployment on all other main ICD-10 diagnosis chapters. The corresponding results (based on the specification with LLM trends) are reported in Table A.3 in the Appendix. Focusing on the effect of average unemployment during the pregnancy period, this overview reveals that indeed the coefficient on perinatal health problems is by far the largest and most significant effect (column 1). Local unemployment significantly affects 5 of the 18 different diagnosis chapters. Rising unemployment increases the prevalence of four health-related diagnoses (i.e., endocrine, nutritional and metabolic diseases; ear and mastoid process diseases; diseases related to the respiratory system; perinatal health problems). In contrast, unemployment is significantly negatively related to the prevalence of cases diagnosed with diseases related to the digestive system. However, the coefficient is very small (-0.12 cases per 1,000 cases) and only significant at the 10% level. This overview thus reveals that in-utero exposure to unemployment adversely affects newborn health as reflected in the rising prevelance of 4 groups of diseases, the most relevant of which - in absolute terms - are related to the perinatal period.

Apart from health-related diagnoses at birth, we also explore the impact of economic downturns on neonatal mortality (columns 7 and 8). None of the estimates shows a significant impact of unemployment prior or during pregnancy on neonatal mortality. While most coefficients are negative, suggesting that higher unemployment reduces the incidence of neonatal mortality, the point estimates are very small and none of them is significant. Hence, the rise in newborn health problems due to in-utero exposure to unemployment reflects higher morbidity which, however, does not translate into higher neonatal mortality.

(even though the estimates on the first trimester effect are significantly larger for men when looking at diagnosis rates).

Table 1.4: Impact on Hospital Diagnoses of Newborns by Gender

	Any		Perinatal health		Congenital defects		Neonatal mortality	
	Female (1)	Male (2)	Female (3)	Male (4)	Female (5)	Male (6)	Female (7)	Male (8)
Panel A: Hospital cases								
$UR_{\tilde{t}}$	6.3707***	5.9341**	5.6177***	5.3898**	0.5865	0.5585	-0.0234	-0.1230
	(2.3384)	(2.4159)	(2.1133)	(2.1211)	(0.5475)	(0.4806)	(0.1007)	(0.0897)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
UR_{t-10}	2.4143*	1.2604	2.1367*	1.2859	0.5131	0.0660	-0.0383	-0.0308
	(1.4204)	(1.2301)	(1.2504)	(1.1379)	(0.3368)	(0.2894)	(0.0790)	(0.0963)
Tri_1	0.7455	2.1337*	0.5034	1.4892	0.1035	0.4420	-0.0435	0.0054
	(1.2630)	(1.2346)	(1.0810)	(1.1997)	(0.3610)	(0.3408)	(0.0948)	(0.1059)
Tri_2	1.3738	-0.6600	1.1532	-0.1875	0.2118	-0.3372	-0.0102	-0.1450*
	(1.3298)	(1.2908)	(1.2378)	(1.1517)	(0.2987)	(0.3179)	(0.0819)	(0.0840)
Tri_3	3.3838**	4.8625***	3.2302***	4.2450***	-0.0052	0.6134*	0.0554	0.0726
	(1.3437)	(1.3417)	(1.2259)	(1.2249)	(0.3509)	(0.3298)	(0.0952)	(0.0791)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
Mean	281.03	316.8	238.36	269.96	23.08	24.1	2.19	2.63
Standard Deviation	449.5	465.23	426.08	443.94	150.15	153.35	46.77	51.19
Observations	2680399	2850507	2680399	2850507	2680399	2850507	2680399	2850507
Reg. cont.	X	X	X	X	X	X	X	X
LLM trend	X	X	X	X	X	X	X	X

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, hospital fixed effects, LLM fixed effects, and month-by-year fixed effects.

As medical studies report gender differences between morbidity and mortality rate of newborns (Zhao et al., 2017), we test next whether unemployment has differential health effects on newborn girls and boys. Table 1.4 shows results separately by sex of newborn. Overall, we find a very similar pattern for boys and girls. The effects of in-utero exposure to unemployment during the entire pregnancy period on any health-related diagnosis and for perinatal health diagnosis are hardly different from the baseline results for the pooled sample (first line); for instance, the estimates reveal that a rise in unemployment by 1 percentage point increases the prevalence of health problem among newborn girls by 6.4 cases and among newborn boys by 5.9 cases (the coefficients are statically indistinguishable from each other).[25] Turning to the trimester specification, the results reveal that the significantly adverse effect of unemployment during the third trimester on health, and particularly on perinatal health, is also of similar size for boys and girls (even though the point estimates are somewhat larger for boys, they are not statistically different from the coefficients for girls). However, the results point towards potential gender differences in the effect of unemployment during the selection period (conception); the coefficient is slightly larger and marginally statistically significant for female newborns. However, again, the point estimates of girls and boys are not statistically different from each other. This significant rise in health-related diagnoses in response to unemployment shocks around the time of conception might indicate that some of the negative health outcomes at birth are caused by

[25] The underlying regression is based on the hospital cases-specification only, since it allows differentiation by gender.

adverse selection into pregnancy during downturns (among girls). However, in terms of size and significance the impact is much stronger during the third trimester.[26] Effects on congenital defects are generally insignificant for both sexes, with the exception of a marginally significant positive effect for boys in the third trimester. Furthermore, there are no general effects of unemployment on neonatal mortality of neither newborn girls nor boys; the exception is a marginally significant coefficient for boys, suggesting that exposure to higher unemployment during the second trimester reduces neonatal mortality of male newborns.

To better understand which particular diagnoses drive the increase in perinatal health problems, we present results on the incidence of subcategories within the group of "perinatal health problems" in Table 1.5. Among the different sub-diagnoses of perinatal health problems, neonatal jaundice (columns 1 and 2) shows the most profound reaction to unemployment rates. Between one third (comparing the coefficients 1.97/6.16) and one half (1.67/3.39) of the additional health problems of newborns can be attributed to neonatal jaundice, a common yellow discoloration of eyes and skin in newborns. Depending on the blood level of bilirubin, jaundice is often treated with phototherapy or exchange transfusions, and long-term consequences can usually be prevented. The results indicate that the highly significant effect of unemployment on neonatal jaundice is caused by in-utero exposure during pregnancy, in particular during the first and third trimester, showing significant positive effects with similar effect sizes. We also find weak evidence for increases in perinatal health problems which are related to maternal health factors and complications (columns 3 and 4). Again, it is mainly exposure to unemployment during the third trimester which increases the risk of these health issues. Similarly, there is also indication of a significantly higher incidence of low birth weight from unemployment in the third trimester (columns 5 and 6). In contrast, the results do not suggest any significant impact of in-utero unemployment exposure on pre-term birth incidence. This result suggests that late fetal growth is negatively affected by unemployment during pregnancy without shortening of gestational length. However, pre-term births might be affected by unemployment via selection effects: higher local unemployment rates in the month before conception significantly increase the incidence of pre-term cases. However, since pre-term births occur before the 37th week of gestation, our mapping of unemployment rates before and during pregnancy is much more imprecise for these cases. A caveat of the hospital data is that we can only identify low birth weight as a *main* diagnosis for defined cut-off weights, but we do not have information on the actual birth weight. Respiratory diseases, another common health issue of newborns, are not affected by unemployment before conception or during pregnancy as shown in columns 9 and 10. There is some evidence of reduced incidence of infectious diseases among newborns (see columns 11 and 12) due to unemployment during the pregnancy. However, this estimate is not robust across specifications and is insignificant in our preferred specification with regional trends.

To investigate further the effect of unemployment on birth weight, we draw on another source of data. The birth registry includes information on the birth weight of each child along with

[26] Note that our analysis is informative about fetuses surviving until birth. Hence, we cannot rule out the possibility that the sex of the fetus affects the likelihood of terminations of pregnancy by death of fetus (Trivers and Willard, 1973).

Table 1.5: Assessment of the Impact on Sub-Diagnoses of Newborns

	Perinatal: Neon. jaundice		Perinatal: Maternal		Perinatal: Low b.w.		Perinatal: Pre-term		Perinatal: Respiratory		Perinatal: Infectious	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Hospital cases												
$UR_{\tilde{t}}$	2.2010***	1.9726***	0.6396	0.8751*	0.3544	0.5663	0.0679	0.4601	-0.2622	-0.0163	-0.6680*	-0.1057
	(0.6559)	(0.6364)	(0.6316)	(0.4494)	(0.4944)	(0.5246)	(0.3748)	(0.4184)	(0.5132)	(0.4567)	(0.3714)	(0.3562)
R^2	.02	.02	.03	.04	.05	.05	.02	.02	.02	.02	.02	.02
UR_{t-10}	-0.2228	-0.2752	0.0311	0.0102	0.2295	0.1761	0.6311**	1.0187***	0.1746	0.1551	-0.1802	0.0157
	(0.3151)	(0.3061)	(0.4292)	(0.2887)	(0.3611)	(0.3230)	(0.2525)	(0.2485)	(0.3162)	(0.3058)	(0.2225)	(0.2106)
Tri_1	1.0171***	0.9641***	0.2897	0.3321	-0.1591	0.0873	-0.4721	-0.4421	-0.3825	-0.2534	-0.0576	0.0527
	(0.3599)	(0.3636)	(0.2331)	(0.2617)	(0.4142)	(0.4314)	(0.3054)	(0.3221)	(0.3256)	(0.3353)	(0.2489)	(0.2438)
Tri_2	0.3058	0.2949	0.1646	0.1987	-0.4557	-0.2574	-0.0284	0.1384	-0.1448	-0.0582	-0.1621	-0.0465
	(0.2705)	(0.2915)	(0.2703)	(0.2107)	(0.3709)	(0.3644)	(0.2525)	(0.2599)	(0.2785)	(0.2816)	(0.2417)	(0.2556)
Tri_3	1.1344***	0.9877**	0.1659	0.3741**	0.9395**	0.8374**	0.1123	0.2703	0.1609	0.2556	-0.3305	-0.1236
	(0.4333)	(0.4080)	(0.1741)	(0.1826)	(0.3801)	(0.3791)	(0.2588)	(0.2556)	(0.3205)	(0.3084)	(0.2519)	(0.2305)
R^2	.02	.02	.03	.04	.05	.05	.02	.02	.02	.02	.02	.02
Mean	31.14	31.14	8.84	8.84	48.68	48.68	25.7	25.7	34.05	34.05	18.92	18.92
Standard Deviation	173.69	173.69	93.59	93.59	215.2	215.2	158.23	158.23	181.36	181.36	136.25	136.25
Observations	5531003	5531003	5531003	5531003	5531003	5531003	5531003	5531003	5531003	5531003	5531003	5531003
Panel B: Diagnosis rates												
$UR_{\tilde{t}}$	2.4954***	1.6703***	0.7375	0.8109*	0.2676	0.1496	0.0854	0.3463	-0.3357	-0.3901	-0.8109**	-0.3117
	(0.6028)	(0.6189)	(0.6319)	(0.4607)	(0.4667)	(0.5319)	(0.3576)	(0.4022)	(0.5116)	(0.4760)	(0.4077)	(0.3565)
R^2	.47	.53	.45	.65	.23	.26	.32	.37	.27	.33	.41	.46
UR_{t-10}	-0.1168	-0.3653	0.0995	0.0353	0.3659	0.2219	0.6193**	0.9102***	0.2196	0.0869	-0.1651	-0.0132
	(0.2909)	(0.3088)	(0.4380)	(0.3003)	(0.3634)	(0.3524)	(0.2419)	(0.2546)	(0.3090)	(0.3078)	(0.2249)	(0.2120)
Tri_1	1.1248***	0.9337**	0.2939	0.3072	-0.2005	-0.0413	-0.4631	-0.4478	-0.3628	-0.2925	-0.1140	-0.0090
	(0.3626)	(0.3617)	(0.2333)	(0.2678)	(0.4274)	(0.4417)	(0.3185)	(0.3358)	(0.3338)	(0.3546)	(0.2663)	(0.2524)
Tri_2	0.3594	0.2373	0.2051	0.1890	-0.4720	-0.3960	0.0487	0.1408	-0.1376	-0.1408	-0.1763	-0.1027
	(0.2691)	(0.2940)	(0.2824)	(0.2229)	(0.3889)	(0.3906)	(0.2566)	(0.2635)	(0.2770)	(0.2849)	(0.2529)	(0.2692)
Tri_3	1.1900***	0.7972**	0.1638	0.3259*	0.7936**	0.6578*	0.0330	0.1963	0.0123	0.0144	-0.4205	-0.2016
	(0.4095)	(0.3950)	(0.1748)	(0.1857)	(0.3833)	(0.3886)	(0.2699)	(0.2664)	(0.3419)	(0.3330)	(0.2679)	(0.2413)
R^2	.47	.53	.45	.65	.23	.26	.32	.37	.27	.33	.41	.46
Mean	31.55	31.55	8.89	8.89	48.96	48.96	25.87	25.87	34.28	34.28	19.01	19.01
Standard Deviation	20.16	20.16	12.77	12.77	20.41	20.41	15.49	15.49	17.41	17.41	13.75	13.75
Observations	24010	24010	24010	24010	24010	24010	24010	24010	24010	24010	24010	24010
Reg. cont.	X	X	X	X	X	X	X	X	X	X	X	X
LLM trend		X		X		X		X		X		X

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. The regressions in Panel A additionally control for hospital fixed effects and gender. The regressions in Panel B are weighted by the average number of live births in a LLM.

birth length and parental characteristics. We estimate the same model as for the diagnosis rates described in equations 1.3 and 1.4 and show results in Table 1.6. The effect of unemployment during pregnancy on the average birth weight is negative and significant in the specification without trends (column 1). A 10 percentage point increase in regional unemployment decreases the average birth weight by 22.51 grams. We also find a 0.9 percentage point increase in low birth weight incidence[27] in the standard specification for a 10 percentage point increase in unemployment. The effects are mostly driven by unemployment in the second trimester (birth weight) and in the third trimester (low birth weight). While both results for birth weight lose significance with the inclusion of LLM-specific trends, the combined evidence with the hospital diagnoses results points towards a negative effect of exposure to in-utero unemployment on birth weight. An even clearer picture emerges when considering the ponderal index. We observe a significant decline in both specifications - with and without regional trends - of 0.234 and 0.189 points per 10 percentage points increase in the local unemployment rate during pregnancy, which corresponds to 0.94% or 0.76% of the mean or 38.4% and 31% of an LLM-level standard deviation. The non-effect on birth length confirms the indication from the hospital data that the birth weight decline is related to fetal growth rather than gestational length.

Table 1.6: Weight and Length from Birth Registry

	Avg. birth weight		Low b.w. (<2700g)		Ponderal index		Avg. birth length	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$UR_{\bar{t}}$	-2.2512***	-1.4344	0.9026**	0.0481	-0.0234***	-0.0189**	0.0022	0.0066
	(0.7140)	(0.9026)	(0.3517)	(0.4042)	(0.0082)	(0.0078)	(0.0069)	(0.0071)
R^2	.36	.37	.31	.34	.84	.86	.71	.72
UR_{t-10}	-0.4094	0.1472	-0.1393	-0.3409	-0.0133**	-0.0112**	0.0057	0.0081
	(0.8372)	(0.8632)	(0.6846)	(0.7269)	(0.0053)	(0.0055)	(0.0050)	(0.0052)
Tri_1	0.3335	0.2928	0.5449	0.3499	0.0075	0.0068	-0.0011	-0.0001
	(1.1214)	(1.1266)	(0.9879)	(1.0081)	(0.0056)	(0.0058)	(0.0066)	(0.0068)
Tri_2	-1.5556*	-1.4284	-0.5013	-0.5179	-0.0120**	-0.0115**	-0.0018	-0.0007
	(0.8836)	(0.9019)	(0.7805)	(0.8121)	(0.0053)	(0.0053)	(0.0053)	(0.0054)
Tri_3	-0.5990	-0.0538	1.1406**	0.6304	-0.0085*	-0.0071	0.0012	0.0044
	(0.7739)	(0.8177)	(0.5374)	(0.5542)	(0.0050)	(0.0050)	(0.0050)	(0.0052)
R^2	.36	.37	.31	.34	.84	.86	.71	.72
Mean	3332.6	3332.6	108.52	108.52	24.91	24.91	51.05	51.05
Standard Deviation	51.81	51.81	15.38	15.38	.61	.61	.45	.45
Observations	24010	24010	8085	8085	24010	24010	24010	24010
Reg. cont.	X	X	X	X	X	X	X	X
LLM trend		X		X		X		X

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered at the LLM level (in parentheses). Regressions are weighted by the average number of live births in a LLM. Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects.

The fact that the negative effect on the ponderal index is driven by unemployment during the second and third trimester has interesting implications. Birth weight is a function of intrauterine growth and gestational length (Kramer, 1987). While intrauterine growth happens mostly in

[27] Low birth weight here is defined as births below 2,700 grams and measured at the quarterly instead of the monthly level.

the third trimester of pregnancies, stressful events like terrorist attacks (Camacho, 2008) and earthquakes (Torche, 2011) affect birth weight most profoundly in the first two trimesters. The gestational length is mainly determined by circumstances early in pregnancy, while the late pregnancy environment and nutrition is critical for intrauterine growth. Accordingly, Bozzoli and Quintana-Domeque (2014) interpret their effect from the first trimester as a sign of maternal stress and their effect from the third trimester as indication of improper maternal nutrition. Following this line of argument, our results could indicate that regional unemployment affects newborn health via maternal nutrition during late pregnancy due to behavioral changes towards less healthy food intake or credit constraints.

1.4.2 Mechanisms

The negative health impact of regional unemployment has a number of potential underlying mechanisms. Our findings on reduced birth weight or increased likelihood of low birth weight birth are mainly driven by in-utero exposure to unemployment in the late pregnancy, which points towards deteriorating food quality as a potential explanation (Bozzoli and Quintana-Domeque, 2014). However, it is less clear whether that also holds true for other health problems related to perinatal health. When investigating potential mechanisms, we focus on the role of parental composition, maternal health, and environmental pollution.

Parental composition

Our analysis of the hospital diagnoses did not indicate any systematic selection effects of pre-conception unemployment rates. However, to investigate potential selection effects in more detail, we additionally draw on the birth registry data containing several relevant background characteristics of parents.[28]

First, we assess whether local economic conditions affect fertility behavior. In Table 1.7, columns 1 and 2, we see a significant reduction in fertility rate in response to a rise in local unemployment ten months prior to birth (measured by number of live births per 1,000 women aged 15 to 44). The decrease in fertility amounts to between 0.48% and 0.67% for a one percentage point increase in unemployment around conception. A smaller birth cohort is in and of itself an interesting result, but more important as a potential explanation of our results is the composition of parents. The decrease in fertility stretches over all maternal age groups under 35 years as shown in columns 9 through 16. The resulting decrease in the average age in columns 3 and 4 is tiny (about 0.01 years, i.e., less than 4 days) and does not indicate a shift in the overall age distribution of mothers. The proportion of parents who are married at the time of birth does not change either (columns 5 and 6). Last, but not least, we do not find significant effects on the share of first born children either, indicating that also family composition

[28] Unfortunately, the birth registry does not contain information on educational attainment. Due to data protection guidelines, we could extract some of the outcomes only on the quarterly level. This concerns the share of births to married mothers, the share of first births, and all age-specific fertility rates.

is unaffected. These results corroborate our previous findings on selection effects and indicate that the decrease in fertility rates due to regional unemployment does not translate into large changes of parental composition. Hence, taken together, our results based on the birth registry as well as on the hospital register reveal that self-selection into parenthood is unlikely to be the main mechanism behind the negative health effects among newborns. For the reduction in the fertility rate to be responsible for the birth weight effect of -2.25g, we would need that the missing newborns are on average between 336g and 468g heavier than the average newborn. The implied selection corresponding to 10% or more of the mean birth weight seems implausible as a mechanism. Moreover, the fertility response could only explain an effect of unemployment prior to conception, but it is not a valid explanation of the third trimester effect. In line with these findings and considerations, Table A.4 in the Appendix shows that our main results are very robust to the inclusion of additional controls characterizing the regional composition of births (in particular, monthly or quarterly indicators of regional age-specific fertility rates, average age of mother at birth, fraction of male births, share of births according to marital status, religious denomination, and birth parity).[29]

Maternal health

Evidence for the US reveals that maternal health behavior in the form of increased prenatal care visits improves with rising unemployment (Dehejia and Lleras-Muney, 2004). If the same effect was at work in our setting, it would bias our results towards zero. But apart from maternal health behavior, maternal health itself could be affected by local labor market shocks and cause a deterioration of newborn health. Similar to the results on health behavior, US studies suggests positive or more recently no effect of state unemployment on adult health (Ruhm, 2000, 2003; Miller et al., 2009; Ruhm, 2015). We investigate the potential mediating role of maternal health in Tables 1.8 and 1.9, building on information provided in the hospital registry on diagnoses related to delivery and pregnancy. The outcome variables in these analyses are the ratios of number of diagnoses per 1,000 live births.

Table 1.8 reports the results on delivery-related health problems. We assess the impact on three different outcome measures: one capturing any health problems related to delivery and two subgroup indicators, differentiating between complications during delivery and other types of problems during delivery (e.g., health problems of the fetus requiring particular care of the mother). As regards the general measure of health problems at delivery, there is no significant effect of average local unemployment during pregnancy. While the estimated coefficients are positive, they are very small compared to the mean and statistically insignificant. In contrast, the results suggest that higher unemployment during the *selection period* significantly increases the likelihood of maternal health problems at delivery. Note, however, that in our preferred specification with regional trends, this effect is only marginally significant. Overall, while expo-

[29]We are aware that these controls might be affected by regional unemployment themselves (and might thus constitute "bad" controls). Even though our analysis does not point towards such compositional effect, this aspect should be generally kept in mind when interpreting the findings.

Table 1.7: Fertility Outcomes and Maternal Characteristics

	Fertility rate		Maternal age (avg.)		Married		First child	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UR_{t-10}	-0.2160***	-0.3009***	-0.0121**	-0.0109**	0.3413	0.0124	-0.7495	0.6389
	(0.0455)	(0.0570)	(0.0049)	(0.0052)	(0.3876)	(0.4662)	(0.7432)	(0.8208)
R^2	.69	.69	.84	.84	.98	.98	.58	.61
Mean	44.72	44.72	30.39	30.39	673.12	673.12	450.21	450.21
Standard Deviation	5.4	5.4	0.93	0.93	136.47	136.47	41.82	41.82
Observations	24010	24010	24010	24010	8085	8085	8085	8085
	Fertility rate: maternal age <25		Fertility rate: maternal age 25-29		Fertility rate: maternal age 30-34		Fertility rate: maternal age>34	
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
UR_{t-10}	-0.0495	-0.1279**	-0.5530***	-0.9207***	-0.3388**	-0.5816***	-0.0135	0.0001
	(0.0463)	(0.0532)	(0.1462)	(0.1928)	(0.1466)	(0.1807)	(0.0383)	(0.0461)
R^2	.83	.84	.79	.80	.82	.83	.92	.93
Mean	24.17	24.17	84.69	84.69	83.67	83.67	21.64	21.64
Standard Deviation	5.27	5.27	13.64	13.64	13.27	13.27	5.94	5.94
Observations	8085	8085	8085	8085	8085	8085	8085	8085
Reg. cont.	X	X	X	X	X	X	X	X
LLM trend		X		X		X		X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Regressions are weighted by the average number of live births in a LLM. Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects.

sure to higher levels of unemployment during pregnancy does not seem to translate into health problems at the time of delivery, the results suggest that higher unemployment prior to conception might induce some selection effects. This pattern does not seem to help explain our findings on newborn health in the previous section.

Columns 3 to 6 report estimates by subgroups of delivery-related diagnoses. Unemployment during different phases before or during pregnancy has no significant effect on delivery-problems related to care of the mother due to health problems of the fetus ("Delivery/other", columns 5 and 6). In contrast, we find significant and adverse effects of higher unemployment during the third trimester on the prevalence of complications at delivery (columns 3 and 4). One additional percentage point in unemployment increases the number of complications at delivery by about 3 cases per 1,000 liveborns. This pattern corresponds to our earlier findings on health-related diagnoses of newborns. Hence, higher unemployment rates towards the end of pregnancy appear to worsen health outcomes of newborns and increase the likelihood of complications during delivery.

Table 1.8: Maternal Health Problems at Delivery

	Delivery (all)		Delivery / Complications (O60, O62-O75)		Delivery / Other (O31-O48)	
	(1)	(2)	(3)	(4)	(5)	(6)
$UR_{\bar{t}}$	2.8022	1.6861	3.3961	3.5863	-0.5939	-1.9002
	(2.2941)	(2.2529)	(2.5226)	(2.2449)	(2.3142)	(2.4326)
R^2	.47	.53	.50	.59	.55	.61
UR_{t-10}	3.4095***	2.1391*	1.2956	0.6799	2.1139*	1.4592
	(1.2498)	(1.1785)	(1.1737)	(1.0621)	(1.1532)	(1.0671)
Tri_1	-1.4444	-0.6429	0.2298	1.0640	-1.6743	-1.7068
	(1.4747)	(1.4922)	(1.1625)	(1.2298)	(1.2094)	(1.2809)
Tri_2	-0.1345	-0.0004	-0.4256	-0.2028	0.2910	0.2024
	(1.2899)	(1.2921)	(1.0964)	(1.0816)	(1.1225)	(1.1014)
Tri_3	2.0525	1.5013	2.9667**	2.9428**	-0.9143	-1.4415
	(1.3089)	(1.3997)	(1.3785)	(1.2819)	(1.3344)	(1.4579)
R^2	0.47	0.53	0.50	0.59	0.55	0.61
Mean	926.33	926.33	572.30	572.30	354.03	354.03
Standard Deviation	80.18	80.18	76.24	76.24	73.8	73.8
Observations	24,010	24,010	24,010	24,010	24,010	24,010
Reg. cont.	X	X	X	X	X	X
LLM trend		X		X		X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Regressions are weighted by the average number of live births in a LLM. Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects.

Table 1.9 reports the results on maternal health problems during pregnancy and diagnoses indicating the termination of pregnancy due to miscarriage or stillbirth. Since we lack information on week of gestation at the time when these diagnoses are recorded, we cannot match local unemployment rates to respective trimesters. We thus restrict our analysis to the specification including average unemployment during the last 9 months, which we interpret as a general indicator for local economic conditions.

Table 1.9: Maternal Health Problems related to Pregnancy, Miscarriages and Stillbirths

	Pregnancy		Infectious and Parasitic Diseases		Mood (affective) disorders		Miscarriages		Stillbirths	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$UR_{\tilde{t}}$	-1.4054	-0.4664	0.7297	0.7830	-0.3846	0.4066	0.1336	0.0483	-0.0183	-0.0168
	(1.2975)	(1.3583)	(0.7290)	(0.7534)	(0.6635)	(0.5625)	(0.2349)	(0.2624)	(0.0216)	(0.0319)
R^2	.68	.70	.43	.45	.6	.63	.33	.36	.10	.12
Mean	234.69	234.69	71.25	71.25	76.78	76.78	15.10	15.10	0.28	0.28
Std. Dev.	78.43	78.43	31.16	31.16	34.2	34.2	10.91	10.91	1.26	1.26
Obs.	23,275	23,275	23,275	23,275	23,275	23,275	22,540	22,540	23,275	23,275
Reg. cont.	X	X	X	X	X	X	X	X	X	X
LLM trend		X		X		X		X		X

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered at the LLM level (in parentheses). Regressions are weighted by the average number of live births in a LLM. Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects.

While the estimates relating unemployment to hospital visits during pregnancy have a negative sign, they are small and not statistically significantly different from zero (columns 1 and 2). Hence, we find no evidence indicating that severe health problems during pregnancy were the prime reason for the adverse effects on newborn health.[30]

Apart from diagnoses directly related to pregnancy and delivery itself, unemployment might harm the health of pregnant women (and the health of newborns) via its impact on the prevalence of infectious diseases. Unemployment changes how and when people get in contact at work, on commuting trips, and at leisure activities and may therefore alter the spread of infections. If higher regional unemployment and perceived higher risk of becoming unemployed induced individuals to go to work while sick, this may increase the rate of infections in the general population and might also put pregnant women at higher risk of disease. To test this potential general health channel of unemployment, we regress the prevalence of infectious and parasitic diseases in women aged 15 to 44 years on regional unemployment. The results in columns 3 and 4 do not reveal any significant effects. Thus, it is unlikely that infectious diseases are the driving force behind the worsened health outcomes of newborns.

Similarly, regional unemployment may affect the mental health of the population and of pregnant women in particular. Adverse behavior related to mental health issues is a potential channel to newborn health. In columns 5 and 6, we show results for mood (affective) disorders in women aged 15 to 44 years. The estimates are small and statistically insignificant in both specifications and, thus, not indicative of maternal mental health explaining the newborn health results.

Finally, we assess the relationship between unemployment and unexpected termination of pregnancy due to miscarriages or stillbirths. Average health conditions at birth might also be determined by the selection occurring during the pregnancy (i.e., conditional on conception). If

[30]In general, hospital visits in Germany reflect comparatively severe indications. Thus, we cannot entirely rule out that local unemployment affects the health during pregnancy in a milder form.

local economic conditions positively affected the survival probability of certain at risk fetuses (potentially those with disadvantaged health), such a selection mechanism might contribute to a deterioration of health of newborns. However, the results in columns 7 to 10 do not support this potential explanation. None of the coefficients is significantly different from zero, suggesting that the main results are not driven by changes in prenatal survival.

Air pollution

A common candidate explanation for positive health impacts of unemployment is the reduction in air pollution that follows economic downturns and the reduction of economic activity. Research has shown that economic cycles are predictive of air pollution and that newborn health is among the most prevalent consequences (Chay and Greenstone, 2003; Currie et al., 2014). However, if air pollution, as commonly regarded, is negatively associated with unemployment due to less production, it cannot explain our negative health results. In Table 1.10, we test how four criteria air pollutants, particulate matter (PM_{10}), nitrogen dioxide (NO_2), carbon monoxide (CO), and sulfur dioxide (SO_2), are responding to unemployment in our setting. We find a strong negative association of all four pollutants with unemployment that vanishes when we include LLM-specific linear trends. The fact that the association with air pollution is negative or zero excludes them from the list of possible channels to health problems of newborns. Air pollution is also unlikely to yield a strong bias toward zero as the relationship of air pollution with unemployment is, unlike our newborn health regressions, not robust to the inclusion of LLM-specific trends. Nevertheless, as a sensitivity check and to rule out any direct health effects of air pollution on newborn health, we include these measures of air pollution as additional control variables in our regression. The results remain very robust (see Table A.5 in the Appendix).

Table 1.10: Air Pollutants

	PM_{10}		NO_2		CO		SO_2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UR_t	-0.103***	0.055	-0.215***	0.003	-0.007***	-0.001	-0.100***	-0.000
	(0.039)	(0.066)	(0.042)	(0.071)	(0.001)	(0.001)	(0.017)	(0.019)
Mean	21.44		24.78		0.41		3.18	
S.D.	12.52		12.54		0.19		2.42	
N	898,336		898,336		898,336		898,336	
LLM trend		X		X		X		X

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered at the LLM level (in parentheses). Regressions are done at the LLM-day level. Control variables include LLM fixed effects, day-of-the-week fixed effect, and month-by-year fixed effects.

1.4.3 Heterogeneity Analysis

Regional characteristics related to (female) labor market participation

Depending on the underlying mechanism, the impact of economic downturns on newborn health might differ across population subgroups. For instance, local unemployment might affect the health of babies mainly through direct unemployment shocks of their parents (generating negative income effects and/or stress). If so, a deterioration of local labor markets should only affect employed subgroups, but not the unemployed or inactive population. If instead local unemployment affects newborn health mainly through increased general stress levels, e.g., because of perceived risks of future own unemployment or worsened future job prospects, expected income loss or a general sense of deprivation, we would expect to see effects in all subgroups (irrespective of own labor market status). Furthermore, if women worry mostly about their *own* job security or income prospects, the effect should be stronger in regions with a higher female labor force participation.

Against this background we test whether the effect of unemployment on newborn health is related to specific characteristics of local labor markets. We split all regions into groups according to whether their characteristics along a certain dimension are below or above the respective sample median. We modify our main specification (with regional trends) by interacting the measures of regional unemployment with binary indicators for particular regions (e.g., unemployment times indicator for regional value above sample median and unemployment times indicator for regional value below sample median). The coefficients of these interactions thus reflect the effect of local unemployment in regions with certain characteristics.[31] We test for regional differences along four dimensions: First, urban versus rural regions (proxied by larger versus smaller population size); second, regions with low versus high shares of school leavers with university entrance qualification (as a proxy for general educational attainment)[32]; third, regions with low versus high average regional female employment rates (to differentiate between regions having traditionally a lower or higher female labor market attachment); fourth, East versus West German regions (to assess differences in female labor force participation due to social norms and child-care infrastructure).

The results of these tests are reported in Table 1.11. Considering the results in Panel A on the effect of average unemployment during pregnancy (top row) we find surprisingly consistent effects in all regional subgroups. As indicated by the reported p-values, the effect of unemployment on newborn health does not differ significantly between rural and urban regions. Similarly, there are no significant differences in the effects with respect to the average educational attainment, female labor market attachment or East and West German regions. Turning to the results of

[31] The classification of regions in subgroups is fixed over time.

[32] In Germany, education at the secondary level commences in grade 5 (approximately at age 10) and is divided into several tracks. Only school graduates of the higher track with 8 or 9 years of additional schooling, obtaining the so-called "Allgemeine Hochschulreife" (general higher education entrance qualification), are entitled to enter university. Hence, the share of school leavers with such a qualification are correlated with the share of university graduates.

Table 1.11: Estimated Health Effects by Characteristics of Local Labor Markets

	Population size		University graduates		Female employment rate		East/West-Germany	
	Low	High	Low	High	Low	High	East	West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Hospital cases								
$UR_{\tilde{t}}$	6.6136***	5.0365*	5.8231**	6.6160**	6.5557**	5.7934**	3.6897	7.3485***
	(2.3263)	(2.9774)	(2.4559)	(2.6685)	(2.6592)	(2.6806)	(2.7088)	(2.6134)
R^2	.16		.16		.16		.16	
P-val. $\beta_l = \beta_h$	.5098		.7288		.7820		.2127	
UR_{t-10}	1.9991*	1.2494	1.2299	2.7974*	0.2383	2.6640*	1.8161	1.8159
	(1.1392)	(1.9633)	(1.2038)	(1.6041)	(1.2245)	(1.4776)	(1.4419)	(1.4798)
Tri_1	1.5333	1.6951	0.9289	1.8424	0.6920	1.5528	1.4446	0.0683
	(1.1548)	(1.3273)	(1.2068)	(1.2579)	(1.2602)	(1.1300)	(1.1516)	(1.3465)
Tri_2	0.3009	0.1198	0.5912	-0.1779	1.1410	0.3883	-0.2086	1.3783
	(1.1693)	(1.7119)	(1.3252)	(1.2792)	(1.1510)	(1.3556)	(1.4366)	(1.2618)
Tri_3	4.4771***	3.4136**	4.3821***	4.0143***	5.6545***	3.2870**	3.0678***	5.1745***
	(1.3296)	(1.4504)	(1.3625)	(1.3907)	(1.6077)	(1.2974)	(1.1366)	(1.7923)
R^2	.16		.16		.16		.16	
UR_{t-10} $\beta_l = \beta_h$	.6292		.2511		.0737		.9999	
Tri_1 $\beta_l = \beta_h$	.8769		.3830		.4650		.2810	
Tri_2 $\beta_l = \beta_h$	.8973		.5381		.5249		.2785	
Tri_3 $\beta_l = \beta_h$	.4130		.7655		.1152		.1859	
N	5,531,003		5,531,003		5,531,003		5,531,003	
Panel B: Diagnosis rates								
$UR_{\tilde{t}}$	3.6395	3.9614	5.6510*	2.8912	1.7389	5.7486**	1.7359	4.9021*
	(3.0755)	(2.4241)	(2.9354)	(2.4328)	(2.5482)	(2.7734)	(2.5827)	(2.6552)
R^2	.61		.61		.61		.61	
P-val. $\beta_l = \beta_h$	.9061		.2716		.1525		.2416	
UR_{t-10}	1.2387	1.5542	2.7027*	0.5690	-0.5604	2.6888*	0.6380	2.5769*
	(1.3196)	(1.6029)	(1.4403)	(1.5311)	(1.3393)	(1.5886)	(1.5326)	(1.5348)
Tri_1	1.2520	1.1346	1.2194	1.4760	0.3115	1.5907	1.4276	-0.7636
	(1.4497)	(1.3862)	(1.4784)	(1.3219)	(1.3931)	(1.3300)	(1.2995)	(1.4901)
Tri_2	-0.7114	-0.1672	-0.7381	-0.4162	-0.9552	0.4281	-0.9335	0.2493
	(1.4012)	(1.5941)	(1.4257)	(1.5235)	(1.3243)	(1.5263)	(1.5149)	(1.4314)
Tri_3	3.2641**	2.7198*	4.1688**	2.3682	3.9433**	3.0603**	1.7237	4.1426**
	(1.5648)	(1.5759)	(1.6956)	(1.4684)	(1.6579)	(1.4712)	(1.3707)	(1.7381)
R^2	.61		.61		.61		.61	
UR_{t-10} $\beta_l = \beta_h$	.8270		.1600		.0331		.2518	
Tri_1 $\beta_l = \beta_h$	.9308		.8484		.3587		.1267	
Tri_2 $\beta_l = \beta_h$	.7104		.8141		.3028		.4282	
Tri_3 $\beta_l = \beta_h$	.7212		.2183		.5651		.1251	
N	24,010		24,010		24,010		24,010	
Reg. cont.	X		X		X		X	
LLM trend	X		X		X		X	

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. The regressions in Panel A additionally control for hospital fixed effects and gender. The regressions in Panel B are weighted by the average number of live births in a LLM. The dependent variable is the ratio of any health-related diagnosis per 1,000 newborn hospital cases (Panel A) or per 1,000 live births (Panel B).

the trimester specification in Panel A, the same picture emerges: The strong and significant detrimental effect of in-utero exposure to unemployment during third trimester is pervasive and significant across all regional subgroups. This pattern is also confirmed in our alternative regression model in Panel B and thus corroborates the robustness of the third trimester effect.

The general conclusion from this subgroup analysis is that the adverse effect of local unemployment on newborn health in general and in particular during the last trimester is pervasive and does not depend on certain characteristics of the regional labor market and in particular female attachment to the labor market. Furthermore, this evidence points in the direction of a general effect of unemployment on stress and worries which is not necessarily related to the personal experience of own job loss.

Differences by General Levels of Regional Unemployment

It is possible that the perceived severity and the induced stress of a rise of local unemployment by one percentage point depends on the general local labor market conditions. For instance, a rise in the unemployment rate by one percentage point in a region with 18% unemployment might have a different impact than the same rise in a region with only 1% unemployment. A priori, it is unclear as to whether the effect in the region with low or high unemployment should be larger or smaller. It could be that a further rise in unemployment in a region already suffering from high levels of unemployment might be perceived as especially dramatic. A possible explanation could be more pessimistic expectations regarding the prospects of finding a new job. However, it is also possible that what matters is the *relative* rise in unemployment levels. In this case, a rise in unemployment by one percentage point in a region with generally low unemployment might be perceived as much more scaring since the relative increase is much larger than the relative increase in a high-unemployment region. This latter scenario would fit well with empirical evidence on unemployment as a social norm which demonstrates that the negative effect of (individual) unemployment on subjective well-being is much smaller the higher the regional unemployment, i.e., the more people in the region share the same experience (Chadi, 2014; Clark, 2003; Clark et al., 2010).

Table 1.12 reports the results of regressions estimating the effect of a one percentage point rise in unemployment in regions with average unemployment levels at the first (bottom), second, third, and fourth quartile of the regional panel (estimated by interacting unemployment with four indicator variables, one for each group of regions).[33] The central insight from this table is that a rise in unemployment per one percentage point has a similar effect on the incidence of adverse health conditions of newborns in regions with very low, medium or very high levels of average unemployment. While the point estimates in Panel A are the smallest for regions with the lowest unemployment, neither the results for Panel B support this relationship nor the joint p-values suggest significant differences between all quartiles. Again, the strong third

[33]The classification of regions in subgroups is fixed over time, according to the average unemployment rate between 2005 and 2013.

Table 1.12: Estimated Health Effects by Average Labor Market Conditions

	Average unemployment rate			
	1. Quartile	2. Quartile	3. Quartile	4. Quartile
	(1)	(2)	(3)	(4)
Panel A: Hospital cases				
$UR_{\tilde{t}}$	0.8586	7.2618**	6.4929**	6.1607**
	(3.3220)	(2.9878)	(3.0995)	(2.9110)
R^2	.16			
P-value $UR_{\tilde{t}}$ $\beta_{q1} = \beta_{q2} = \beta_{q3} = \beta_{q4}$	.2288			
UR_{t-10}	1.8318	3.8778**	2.2825	1.6052
	(1.9895)	(1.6736)	(1.8181)	(1.3689)
Tri_1	-1.8928	-0.6856	0.9576	2.2137*
	(1.7557)	(1.5148)	(1.6843)	(1.2164)
Tri_2	-1.6414	1.0164	1.7400	0.0886
	(1.7224)	(1.4127)	(1.5193)	(1.4077)
Tri_3	4.1039*	4.7248**	2.7977	4.0463***
	(2.3162)	(1.9016)	(2.0026)	(1.3064)
R^2	.16			
P-value UR_{t-10} $\beta_{q1} = \beta_{q2} = \beta_{q3} = \beta_{q4}$	.6321			
P-value Tri_1 $\beta_{q1} = \beta_{q2} = \beta_{q3} = \beta_{q4}$	.0622			
P-value Tri_2 $\beta_{q1} = \beta_{q2} = \beta_{q3} = \beta_{q4}$	.2586			
P-value Tri_3 $\beta_{q1} = \beta_{q2} = \beta_{q3} = \beta_{q4}$	.7744			
N	5,531,003			
Panel B: Diagnosis rates				
$UR_{\tilde{t}}$	2.8471	3.2617	0.0443	6.2916**
	(3.5141)	(3.6855)	(2.8302)	(2.9446)
R^2	.61			
P-value $UR_{\tilde{t}}$ $\beta_{q1} = \beta_{q2} = \beta_{q3} = \beta_{q4}$	.2872			
UR_{t-10}	2.5998	2.6534	-0.5495	2.1429
	(2.0762)	(1.8886)	(1.8421)	(1.5882)
Tri_1	-2.0029	-2.2377	0.2563	2.4698*
	(2.2424)	(1.8494)	(2.0673)	(1.2622)
Tri_2	-1.2433	-0.1841	0.0707	0.1069
	(2.0932)	(1.6613)	(1.6682)	(1.5168)
Tri_3	5.2426**	4.5465**	1.0060	3.2970**
	(2.2846)	(2.0713)	(2.1647)	(1.5187)
R^2	.61			
P-value UR_{t-10} $\beta_{q1} = \beta_{q2} = \beta_{q3} = \beta_{q4}$	.4125			
P-value Tri_1 $\beta_{q1} = \beta_{q2} = \beta_{q3} = \beta_{q4}$	.0342			
P-value Tri_2 $\beta_{q1} = \beta_{q2} = \beta_{q3} = \beta_{q4}$	.9252			
P-value Tri_3 $\beta_{q1} = \beta_{q2} = \beta_{q3} = \beta_{q4}$	.2492			
N	24,010			
Reg. cont.	X			
LLM trend	X			

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. The regressions in Panel A additionally control for hospital fixed effects and gender. The regressions in Panel B are weighted by the average number of live births in a LLM.The dependent variable is the ratio of any health-related diagnosis per 1,000 newborn hospital cases (Panel A) or per 1,000 live births (Panel B).

trimester effect is prevalent irrespective of underlying general levels of local unemployment or labor market conditions. Some evidence points towards a different impact of unemployment in the first trimester. It is estimated to increase hospital cases in regions with the highest average unemployment while having non-significant negative estimates for the other regions.

1.4.4 Further Sensitivity Analyses

Hospital capacity and diagnoses of newborns

Thus far our explanations of deteriorating health of newborns have mainly focused on the health production process of fetuses and on the potential role played by selection into pregnancy and parenthood. We implicitly assumed that the quality of hospital treatments as well as coding of diagnosis by clinicians was unrelated to changes in local economic conditions. Theoretically, however, fluctuations in unemployment might be related to the degree of capacity utilization and the quality of diagnosis coding and treatments in hospitals. As regards the funding of hospitals in Germany, operating expenses (i.e., costs related to the medical treatment of patients) are covered by public and private health insurers via reimbursement according to the diagnosis related groups (DRGs) payment system (Quentin et al., 2010)[34]. Investment costs of hospitals are covered by the federal states. Hence, short-term fluctuations in local economic conditions and unemployment are unlikely to affect hospital funding in the short-run, as only long-run investments are related to state budgets and taxes, while the operating costs are funded at the national level. Health insurance coverage of the local population is not related to unemployment either, because the unemployed remain insured via the social security system.

However, our results suggest that fertility rates are influenced by local economic conditions. Higher unemployment is related to decreases in birth rates, i.e., fewer women giving birth and requiring assistance in hospitals. Hence, fluctuations in birth rates in concordance with fluctuations in unemployment might affect the degree of capacity utilization in obstetrics wards and clinics counter-cyclically and potentially spill-over into diagnosis coding and/or quality of treatment.

To assess the sensitivity of our results to this potentially confounding mechanism, we pursue the following two strategies. First, we re-estimate our main model replacing the main explanatory variable monthly local unemployment by monthly local birth rates (number of live births to mothers aged 15 to 44). In particular, we now estimate the effect of average monthly birth rates during the pregnancy period or - in the trimester specification - the average monthly birth rates during the trimesters and in the selection period. If our main results were simply driven by capacity induced factors, we would expect to see that higher (lower) birth rates are related to improved (worsened) health of newborns. However, the results (reported in Table A.6 in the Appendix) do not reveal any significant impact of birth rates on health outcomes at birth.

[34] The DRG system was gradually introduced in Germany starting in the year 2000.

As a second test we re-estimate our original specification now additionally *controlling* for local birth rates. We thus investigate to what extent our results are sensitive to the inclusion of these capacity-related numbers. As the results in Table A.7 in the Appendix indicate, our main findings in Table 1.3 remain virtually unaffected by this amendment.

To conclude, the combined evidence suggests that the negative newborn health impact of unemployment is not caused by capacity induced changes in hospitals.

Robustness to Aggregation

Most previous studies that assess the effect of unemployment and recessions on health have used annual unemployment rates (matched to year of conception) instead of monthly unemployment rates (matched relative to month of birth). Since our findings stand in contrast to the counter-cyclical health effects reported in some of these studies, we report results based on yearly unemployment rates in Table 1.13 (Panel B). Furthermore, we assess the sensitivity of our results to the level of geographic aggregation (across columns in Table 1.13). In particular, we contrast our results based on the local labor market level (245 LLMs) with results on a more disaggregated county level (385 counties) and a more aggregated level (Raumordnungsregionen) covering 96 regions.

Table 1.13: Impact of Aggregation on Newborn Health Results

	Counties		LLM		ROR	
	(1)	(2)	(3)	(4)	(5)	(6)
Hospital cases: monthly UR						
YM+Region FE	0.7112	4.0997**	3.1490	6.1591***	4.0394	7.0727**
	(1.9971)	(1.9512)	(2.5295)	(2.2848)	(3.1015)	(3.0578)
Hospital cases: yearly UR						
YM+Region FE	1.0969	4.8324**	4.4252	7.6237***	6.3079*	9.2525***
	(2.1977)	(2.3256)	(2.7460)	(2.7742)	(3.3350)	(3.4295)
Observations	5,531,003	5,531,003	5,531,003	5,531,003	5,531,003	5,531,003
Reg. cont.	X	X	X	X	X	X
LLM trend		X		X		X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Regressions are weighted by the average number of live births in a LLM. Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. All regressions control for hospital fixed effects and gender. The dependent variable is the ratio of any health-related diagnosis per 1,000 newborn hospital cases (Panel A) or per 1,000 live births (Panel B).

While our results are qualitatively robust to the degree of geographic aggregation (see Panel A), the magnitude of the estimated coefficients increases monotonically with the level of aggregation. This pattern corresponds with the findings by Lindo (2015) for the US, who argues that it is due to spillover effects of local unemployment on health outcomes of other counties. However, despite the differences in magnitude, the estimated effects do not differ statistically from each other. Similarly, a comparison of the impact of *monthly* unemployment rates during pregnancy versus *annual* unemployment rates around conception (Panel A and B, respectively) reveals a sightly larger magnitude of the coefficients based on annual unemployment (however, the estimates do not differ significantly from each other).[35]

Summing up, these additional tests imply that our finding of a health deteriorating effect of rising unemployment - which is in contrast to many studies - is not simply caused by our refined unemployment measure. Moreover, it seems that our estimated effects are potentially more conservative than those estimated in studies relying on annual unemployment rates.

1.5 Estimated Hospital Costs

To put our results in context, we estimate the hospital costs which accrue due the rising share of newborns with adverse health conditions originating in the perinatal period (P-diagnoses) when the economy worsens. The latest official figures on cost of illness by ICD-10 diagnosis and by provider in Germany exist for 2015 (e.g., hospitals, ambulatory health care, residential long-term care). These diagnosis-specific costs are based on the universe of patients, i.e., covering patients of all ages. We report these figures on costs of illness for all main diagnosis chapters in Table A.9 in the Appendix (the table also contains a breakdown of the official corresponding numbers of hospital patients per diagnosis chapter for all patients and for patients below age 1.) Since only (newborn) infants receive P-diagnoses, the national cost numbers on P-diagnoses - in contrast to the other diagnosis chapters - are age-specific and hence informative about the costs of these illnesses among the youngest patients. We will focus the following calculations exclusively on P-diagnoses, because our estimates indicated that this diagnosis chapter is the most affected by changes in economic conditions.

According to official numbers summarized in Panel A of Table 1.14, the average cost per infant patient with perinatal health problems amounts to 6760 Euro in 2015 (9-day hospital stay X 751 Euro per day). In contrast, the hospital costs related to a healthy liveborn infant, i.e., a clinically healthy newborn (Z30-Z39) are only about 953 Euro (3.1 days X 307.5 Euro per day). This difference is due to shorter hospital stays and lower daily costs associated with healthy newborns than for newborns requiring treatment due to perinatal health problems. Hence, each newborn with a P-diagnosis instead of no health-related diagnosis (i.e., a healthy liveborn infant) increases hospital costs by about 5807 Euro.

Based on this figure on extra costs per each additional case with a P-chapter diagnosis, we now calculate by how much overall hospital costs increase due to a one percentage point rise in the local unemployment rate. Panel B of Table 1.14 reports the corresponding numbers based on our estimates on the impact of unemployment on perinatal health problems (P-diagnoses) in our preferred specification (top row; A. Individual hospital cases with hospital fixed effects and regional trends) and the alternative estimates based on regional diagnosis rates (bottom row; B. Diagnosis rate).

[35] We approximate the month of conception as the month 9 months before the hospital case and assign the annual unemployment rate in the respective year accordingly.

Our preferred estimates imply that a one percentage point rise in the unemployment rate increases the number of newborns with perinatal health problems by about 5.5 cases or 31,940 Euro per 1,000 newborns. Aggregated to the national level, this translates into about 4,060 additional cases and about 23.9 Million Euro. This corresponds to 0.03 percent of the total hospital costs of illness in 2015 and to 1.8 percent of all hospital costs related to P-diagnoses in 2015, respectively. The respective numbers using the lower estimate based on regional diagnosis rates are 2,655 additional cases on the aggregated level or 15.7 Million Euro (about 0.02% of total hospital cost of 1.2% of hospital costs related to P-diagnoses). When relating these estimates to the decline in the national unemployment rate between 2005 and 2013 (from 12.85% to 7.72%), Germany benefits by saving 80.54 to 122.61 Million Euro annually.

Our estimated costs most likely represent a lower bound of the total health-related problems and overall costs related to worsening economic conditions. They only account for increased health care costs at the time of birth, and hence short-run effects. Our framework neglects potential future negative health consequences of these affected children. Furthermore, the estimates are restricted to direct and immediate health costs in hospitals; hence, any additional treatment in ambulatory health care or negative spill-over effects of poor health at birth on child development, education, labor market, and future wages are not accounted for.

Table 1.14: Estimated Hospital Costs per 1pp Rise in Unemployment

Panel A: Hospital costs per case with P-diagnosis versus cost per healthy liveborn case

	Hospital stay (avg. days)	Costs per day (avg.)	Total cost per case (avg.)
(1) Newborn with P-diagnosis	9	751.1	6759.9
(2) Healthy liveborn infant (Z30-Z39)	3.1	307.5	953.25
Cost difference (1) - (2)			*5806.65*

Panel B: Estimated number of cases and associated hospital costs of additional P-diagnoses per 1 pp rise in local UR

	Per 1,000 live births		Among all liveborn infants (live births in 2015: 737,575)			
Estimation specification	Cases	Costs	Cases	Costs	in % of total hospital costs (88,562 Mill.)	in % of total hospital costs, P-diagnoses (1,298 Mill.)
A. Hospital cases	5.5	31939	4,057	23,922,139	0.03	1.84
B. Diagnosis rate	3.6	20905	2,655	15,658,127	0.02	1.21

Notes: Own calculations based on estimated rise in number of cases per 1pp rise in local unemployment. Figures on average cost per hospital day based on official statistics of main diagnoses and cost of illness by ICD-10 code and type of provider, German Statistical Office (Bonn). For an overview see Table A.9 in the Appendix.

1.6 Conclusion

In this study, we assess the relationship between changes in regional economic conditions and the health of newborn children. Seminal studies for the US and other European countries report a positive impact of recessions and increasing annual unemployment rates on the health of newborns. In contrast, our results based on hospital register data of main diagnoses of the universe of German patients imply a deterioration of newborn health due to in-utero exposure to worsening economic conditions. Exploiting regional and temporal variation across 245 local labor market regions in Germany during the period 2005 to 2013, we find that a one percentage point rise in the average unemployment rate during the pregnancy period increases the prevalence of health-related diagnoses of newborns in the range of 3.8 to 6.2 cases per 1,000 live births, i.e. about 1 in 200 live births. The vast majority of these additional cases are related to health conditions originating in the perinatal period (P-diagnoses), namely 3.6 to 5.5 cases per 1,000 live births, out of which neonatal jaundice stands out (about 2 additional cases pre 1,000 live births). Moreover, using monthly unemployment rates to test for heterogeneous effects during the pregnancy period, we find that the adverse health outcomes at birth are mainly driven by exposure to higher unemployment rates during the third trimester. These findings hold for newborn girls and newborn boys alike and they are remarkably consistent across regions with very distinct labor market characteristics, e.g., female labor market attachment. They are also not sensitive to the level of regional aggregation or to using yearly instead of monthly unemployment rates. Using birth registry data, we also find additional evidence that rising local unemployment leads to a small, but significant reduction in birth weight and the Ponderal index. These results corroborate our main findings that local unemployment during pregnancy is related to worsened health outcomes at birth.

To put our estimated effects in context, we quantify the additional total costs accruing to hospitals (and, hence, the health care system) per one percentage point rise in local unemployment rates. The estimated aggregated costs range between 15.6 to 23.9 Million Euro (in 2015) which corresponds to 1.2% to 1.8% of annual hospital costs accruing to health conditions originating in the perinatal period. Considering potential spill-over effects on the future health or development of children, these estimated short-run hospital costs are likely a lower bound estimate of the overall economic costs.

We explore potential mechanisms and alternative explanations for our findings. We address the role of endogenous selection into parenthood by analyzing maternal characteristics of newborns and by estimating the impact of unemployment rates prevailing before conception on later newborn health. Overall, our results do not provide any strong or robust indication of a statistically or economically relevant selection effect. While a few specifications or subsample regressions point to a potential adverse selection effect, the large bulk of our results and in particular our main specification and results do not seem to be driven (exclusively) by endogenous selection into pregnancy (and parenthood). Similarly, we cannot detect any significant unemployment-related impact on hospital visits during pregnancy or on miscarriages or stillbirths. However, as regards maternal health, we find significant effects of exposure to rising unemployment rates on complications during labor and delivery. This effect is again particularly driven by the impact during the third trimester. Hence, it could be that the pattern of the maternal outcomes at the time of birth is related to the worsened health conditions of newborns at birth. Finally, we investigate the potential roles played by air pollution on the one hand, and by hospital capacity-induced diagnoses on the other hand. None of the estimates indicate a significant impact of these potential (alternative) mechanisms on newborn health.

In the German institutional context - with its dismissal protection of pregnant women and a relatively strong and generous social security system, including public health and unemployment insurance - it is not very likely that our findings are driven by (extreme) income losses. Alternative potential explanations might be related to maternal stress or nutrition (Aizer et al., 2016). Indeed, our findings of adverse health impacts of rising unemployment (in the third trimester) are in line with papers assessing the effect of "stressful news about" rather than "actual physical exposure to" events like hurricanes (Currie and Rossin-Slater, 2013), regional plant closures (Carlson, 2015) or number of foreclosures in the neighborhood (Currie and Tekin, 2015) on the health of newborns. From this perspective, a worsening of local economic conditions might induce stress and worries among pregnant women with potentially detrimental effects on the development of their fetus on the risk of complications during delivery. To what extent such negative spill-over effects of local labor market conditions on maternal stress can be quantitatively substantiated and whether psychological counseling of pregnant women during economic downturns can help cushioning detrimental effects on maternal stress and newborn health are both important and policy-relevant questions for future research.

Chapter 2

Out of the Dark, into the Light? The Impact of Seasonal Time Changes on Work-Related Accidents

2.1 Introduction

2021 might be the last year in which European citizens experience seasonal time changes. The European Parliament voted on March 26, 2019, for a resolution to end the practice of adjusting clocks twice a year in the European Union (European Parliament, 2019a). This act reflects the population's repugnance of switching between Standard Time (ST) and Daylight Saving Time (DST). The clocks are currently moved forward by one hour on the last Sunday in March to enter DST and set back on the last Sunday in October to return to ST. In a public consultation in summer 2018, a majority of 84% of the 4.6 million respondents across Europe favored the abolition of biannual clock changes. Consequently, the European Commission proposed to discontinue seasonal time changes, although their analysis of whether the benefits outweigh the costs remained inconclusive (European Parliament, 2019b,a).[1] While the initial aim of introducing DST transitions was to reduce energy consumption, the general debate often centers on potentially detrimental effects in other areas, such as health, crime, and traffic collisions.

This paper contributes to the public discussion by investigating the impact of seasonal time changes on work-related accidents. Despite the occurrence of around 374 million annual occupational accidents worldwide, the toll of time adjustments is unclear.[2] To the best of my knowledge, this is the first study on the impact of DST transitions on work-related accidents considering commuting accidents as an explicit outcome variable. The discrete nature of the time change

[1] The phase-out of the existing time regime does not necessarily mean the abolishment of DST. Instead, it will fall to the individual member states whether to adopt Standard Time or Daylight Saving Time all year round.

[2] Estimations of the International Labour Organization (ILO) provide this number. Work-related accidents result together with work-related injuries in 2.78 million deaths annually, of which 2.4 million are disease-related (Wadsworth and Walters, 2019).

allows me to obtain causal estimates of the transitions into as well as out of DST. My results indicate no systematic influence of the DST regime on workplace or commuting accidents for the period 2013-2017. I identify insignificant positive estimates (for an increase in accidents) in spring when switching to DST and coefficients close to zero when reverting to ST in fall. The estimates of the transition to DST on accidents during commuting stand out but lack precision. They provide suggestive evidence for an increase in commuting accidents by around 9.1 percent, in the days following the time change in spring.

To identify the impact of DST transitions on work-related accidents, I use data from the German Social Accident Insurance for the period 2013-2017. It contains daily information at the state level and distinguishes between accidents that occur at the workplace and such during commuting. Additionally, I merge employment data, weather measures, and information on public holidays and school holidays. I exploit the quasi-experimental nature of DST and estimate a regression discontinuity (RD) design. My estimation strategy comprises two steps. First, I demean work-related accidents by systematic calendar (and weather) effects. Specifically, I account for variation in the casualty rate between states, over years, days of the week, and with respect to public and school holidays. Second, I take the demeaned residuals to estimate the causal effect of the DST transition in an RD design. In the estimation, I rely on state-of-the-art data-driven bandwidth selection procedures (Calonico et al., 2014b,a, 2017, 2018, 2019, 2020).

The change from ST to DST means shifting one hour of ambient light from the morning to the evening. When introducing DST transitions, policymakers expected that the "additional" hour of evening daylight would yield reductions in electricity consumption. Prominent studies, however, cast doubt that DST succeeds in achieving the prime goal of saving energy. For example, they indicate null effects on electricity consumption for the United States and Australia (Sexton and Beatty, 2014; Kellogg and Wolff, 2008). Kotchen and Grant (2011) even show an increase in residential electricity consumption for the US state Indiana. As Aries and Newsham (2008) outline in a review of the literature, the majority of studies find either small savings in lighting energy use or no effect. Havranek et al. (2018) provide a recent meta-analysis collecting 162 estimates from 44 studies. While these studies report, on average, a slight reduction in energy consumption of -0.34%, the estimated effect approaches zero when accounting for measures as data, method, and publication characteristics. A country's location seems to be an essential factor in explaining cross-country heterogeneity, with countries further away from the equator exhibiting more substantial electricity savings.

Sleep seems to be the major transmission channel of time changes causing detrimental outcomes.[3] A (mostly) non-economic literature indicates disruptions of the human biorhythm (circadian rhythm) caused by changes in the local clock time and the alteration of the length of the transition day (to 23 hours in spring and 25 hours in fall). It points towards asymmetric effects on how the loss of an hour due to the time change, compared to the gain of an hour, translates into sleep duration. While several studies discover reductions in the time slept on the days following the transition to DST (Barnes and Wagner, 2009; Lahti et al., 2006; Sexton and

[3]The light relocation also has general effects in other areas (see Section 2.2.2).

Beatty, 2014; Michelson, 2011), fewer identify a significant positive impact on the sleep length when returning to ST (Jin and Ziebarth, 2020; Michelson, 2011). Barnes and Wagner (2009) for example, who analyze American Time Use Survey data detect respondents to sleep on average 40 minutes less on the Monday following the spring transition. After the fall transition, there is no significant effect of the additional hour on sleep duration. Reviewing the sleep literature, Harrison (2013) finds a reduction in sleep continuity and sleep efficiency for up to one week following the change in the social clock. Furthermore, the step-wise adjustment of bed and rise times result in "a cumulative effect of sleep loss at both transitions." Kantermann et al. (2007) show that the human circadian clock[4] adjusts more easily to the transition in fall, out of DST, than the transition in spring, into DST.[5]

This paper contributes to the literature on the impact of DST transitions on health, occupational, and traffic accidents. In line with the disruption of the sleep routine, several papers report negative health consequences in the days following the transition from ST to DST. These entail a reduction in life satisfaction (Kountouris and Remoundou, 2014; Kuehnle and Wunder, 2016), a rise in unipolar depressive episodes (Hansen et al., 2017), suicides (Berk et al., 2008), ischemic strokes (Sipilä et al., 2016), and acute myocardial infarction (heart attacks) (Janszky and Ljung, 2008; Janszky et al., 2012; Jiddou et al., 2013; Toro et al., 2015; Manfredini et al., 2018). Contrary to these previous studies, Jin and Ziebarth (2020) find no effect at the time shift in spring but sharp reductions in German hospital admissions due to heart attacks and injuries in the four days following the fall transition.

The literature on the causal impact of seasonal time changes on work-related accidents is scarce, and existing studies focus on incidents at the workplace. Robb and Barnes (2018) estimate the effect of DST transitions on multiple accident categories in New Zealand, using data on the universe of injury claims by the Accident Compensation Corporation for the years 2005-2016. Their results indicate no significant impact on work accidents during any of the first seven days following a transition, both into as out of DST. Barnes and Wagner (2009) show that the transition to DST gave rise to mining injuries in the United States between 1983 and 2006. The spring transition leads to an increase in injuries by 5.7% on the following days. Other studies suggest no significant relationship between seasonal time changes and work-related accidents in Finland (Lahti et al., 2011), work injuries in Ontario, Canada (Morassaei and Smith, 2010), or construction accidents in the state of Washington (Holland and Hinze, 2000). However, those studies either fall short of accounting for calendar effects or seasonal trends.

I also advance the strand of literature on traffic accidents, given that commuting accidents are a subset of vehicle collisions. Smith (2016) made a seminal contribution to this literature by estimating the causal effect of DST transitions on fatal vehicle crashes. His results imply 5.6% additional fatal accidents in the six days following the transition to DST, which led to over 30

[4]The human circadian clock uses daylight to regulate the sleep-wake cycle and synchronize it to the environment.

[5]More specifically, they observe a seasonal alignment of mid-sleep and wake-up times to dawn during ST but not during DST. The adjustment to dawn starts and stops when leaving respectively when entering DST, despite the differential dates of the transition with respect to the season.

deaths annually between 2002 and 2011. By estimating both an RD design and a day-of-year fixed effects model, he can attribute the harmful effects to sleep deprivation (instead of to the allocation of ambient light). Earlier work, which typically compares the number of crashes in the days following DST to the week before (and after), finds partly an increase in traffic accidents (Coren, 1996; Varughese and Allen, 2001) and partly no effect of the spring transition (Vincent, 1998; Sood and Ghosh, 2007; Lahti et al., 2010). Robb and Barnes (2018) identify significantly higher rates of road accidents at the onset of DST.

Given the existing literature, one would expect either increased work-related accidents or no effect when switching from ST to DST and no effect or even a potential decrease when returning to ST in fall. Moreover, the literature on vehicle collisions hints towards a stronger impact on accidents while commuting than at the workplace. My results are in line with the previous studies on work and traffic accidents. They suggest that DST transitions have no impact on workplace accidents but include sizable estimates during commuting in spring. The sizable effects on commuting accidents are stable in magnitude across a rich set of alternative specifications but are imprecisely estimated. I argue that sampling error impedes the statistical significance of my results. This calls for future research on the impact of seasonal time changes on the incidence and severity of commuting accidents.

The paper proceeds as follows. Section 2.2 provides background information on DST, Section 2.3 describes the data, and Section 2.4 illustrates the variation of work-related accidents. Section 2.5 explains the estimation strategy, Section 2.6 discusses the results, and Section 2.7 concludes.

2.2 Background

2.2.1 History of Daylight Saving Time

The policy of seasonal time changes was first introduced, on a national level, in 1916 as a wartime measure by Germany and Austria-Hungary.[6] Other countries followed with the introduction of DST, also known as summer time, to save energy during the years of World War I and World War II. While widely abolished in peace times, the DST regime regained popularity with the 1973 oil crisis. For some countries like Germany, however, the alignment of their time with their neighbors' played a major role in (re)introducing DST (Bundesregierung, 1977). In the following, the European Union (EU) harmonized the beginning and the end of DST among all member states (European Parliament and Council, 1994). Since 1996, the clocks are moved forward by one hour on the last Sunday in March to enter DST and put back on the last Sunday in October to return to ST or winter time. On the days of the transition, the time adjustment takes place at 1 AM Greenwich Mean Time (GMT) (European Parliament and Council, 1997, 2001). For Germany accordingly, clocks are put forward from 2 AM to 3 AM on the last Sunday in March and reversed on the last Sunday in October from 2 AM to 1 AM.

[6] See Reichsgesetzblatt Nr.67/1916 in Grimm et al. (1994) for Germany and Reichsgesetzblatt 111/1916 (Österreichische Nationalbibliothek, 1916) for Austria-Hungary.

Currently, over 60 countries worldwide practice seasonal time changes, and almost as many countries experimented with DST regimes in the past.[7] In recent years, this biannual procedure has been subject to public debate both in Europe and in the United States. Following a proposal by the European Commission, the European Parliament voted on March 26, 2019, to end the practice of adjusting clocks twice a year in the EU (European Parliament, 2019a,b). The phase-out of the existing time regime does not necessarily mean the abolishment of Daylight Saving Time. Instead, it will fall to the individual member states whether to adopt Standard Time or Daylight Saving Time from 2021 onward. At the same time, a bi-partisan group of senators introduced the Sunshine Protection Act to the US Congress. This bill aims at establishing year-round DST in most states of the US (United States Congress, 2019).

2.2.2 Permanent Daylight Saving Time

There are studies that provide insight into whether a permanent installment of DST might be advantageous over the known standard time. Instead of focusing on the short-time consequences in the days following the transition to DST, they aim at studying the permanent effect of DST. The change to DST means shifting one hour of ambient light from the morning to the evening. This light effect prevails over the short-run impact on sleep duration.

Additionally to the discontinuous nature of DST, Doleac and Sanders (2015) exploit its 2007 extension in the United States to investigate the effect of light on criminal activity.[8] They show that robberies decrease by 7% due to DST, which is driven by the hours directly affected by the shift in daylight. Evaluating American Time Use Survey data around the same reform, Wolff and Makino (2012) find that people increase their time spent on recreational outdoor activities by 30 minutes a day, while they cut back on watching TV by 9 minutes. Bünnings and Schiele (forthcoming) estimate that permanent DST could save 8% of the costs caused by road accidents in Great Britain. They exploit variation in light conditions both across regions resulting from geographical differences and within regions resulting from changes in darkness over the year.[9]

Later sunset times are, however, associated with shorter sleep duration both within a location over the year and across locations within the same time zone. Gibson and Shrader (2018) use the exogenous variation in sunset times to identify the impact of sleep on labor productivity. They find positive earnings effects driven by wage changes, which they interpret as intensive margin effect of sleep on labor productivity. Giuntella et al. (2017) also use sunset time as an instrument and find that sleep duration raises cognitive skills and reduces depression symptoms of Chinese workers aged 45 years and older. These effects are driven by employees living in urban areas, plausibly due to rigid work schedules.

[7] `www.timeanddate.com` provides a historical and statistical overview. Depending on the definition of what constitutes a country, 62-75 countries practice DST in 2020.

[8] As part of the Energy Policy Act of 2005, both transitions dates were changed. As of 2007, DST starts on the second Sunday in March and ends on the first Sunday in November. Compared to earlier years, this meant an extension by four to five weeks.

[9] For the United States, there is evidence that the relocation of ambient light leads to a reduction in crashes in earlier years (1976-2001) but that this channel vanished more recently (Sood and Ghosh, 2007; Smith, 2016).

2.3 Data

2.3.1 Work-Related Accidents

To measure work-related accidents, I use data provided by the German Social Accident Insurance (Deutsche Gesetzliche Unfallversicherung). The German Social Accident Insurance (hereafter DGUV) is the umbrella association of the statutory accident insurance institutions for the industrial and public sectors.[10] In 2016, it covered more than 40 million full-term equivalent employees in around 3.9 million establishments. For that year, the DGUV recorded a total of 1,063,141 reportable work-related accidents, of which 877,071 occurred at the workplace and 186,070 during commuting to or from the place of work (DGUV, 2017b; 2017c). A work-related accident is reportable if it results in an incapacity to work for more than three days (or in death). My data set includes work-related accidents for the group of dependent employees, which caused 92.2% of the reportable accidents in 2016. It contains daily information for the 16 German states (Bundesländer) in the years 2013 to 2017.[11] A major drawback of the data is that the DGUV does not record the universe of accident spells, but projects them based on a sample of casualties (DGUV, 2017a). Accidents are sampled for those accident victims whose birthday is on the 10th or 11th day of a month in the private sector and on the 10th, 11th, or 12th day of a month in the public sector and extrapolated to the entire population. This introduces measurement error in the dependent variable.

2.3.2 Control Variables

Holidays. As sketched above, school and public holidays serve as main control variables for my analysis. Data on school holidays is provided by the *standing conference of the ministers of education and cultural affairs of the Länder in the Federal Republic of Germany*.[12] This conference also determines the summer school holidays of the states in a rotating system to alleviate the burden on transport infrastructure and to please the tourism industry. The state governments schedule the remaining school holidays themselves, as they do with public holidays apart from *Day of German unity (3rd of October)*, which is set by the federal government.[13] Figure 2.1 illustrates the school and public holidays for the 16 German states in 2016. It captures the differences between states and the incidence of school and public holidays around the changes from ST to DST and back. Since holidays around the change to DST vary with Easter and the

[10] The German Social Accident Insurance does not only insure employees and apprentices but extends to other groups of people. These include, for example, pupils and students in schools and universities as well as people providing home care (§2-§6 SGB VII (1996)).

[11] 1.008% of the accidents are not assigned to a state, either because they happened abroad or due to missing information. I drop them from the estimation sample. Similarly, I exclude 0.001% of accidents which have missing information on the exact day the accident took place (while being assigned to a state in a given month and year).

[12] Data on school holidays was retrieved from `https://www.kmk.org/service/ferien/archiv-der-ferientermine.html` on July 19, 2018.

[13] Data on public holidays was retrieved between July 19, 2018 and August 1, 2018 from `https://www.bmi.bund.de/DE/themen/verfassung/staatliche-symbole/nationale-feiertage/nationale-feiertage-node.html`, `schulferien.org` and cross checked with data provided by the corresponding state ministries.

fall holidays around the change back to ST vary with the summer holidays, this pattern varies over the years.

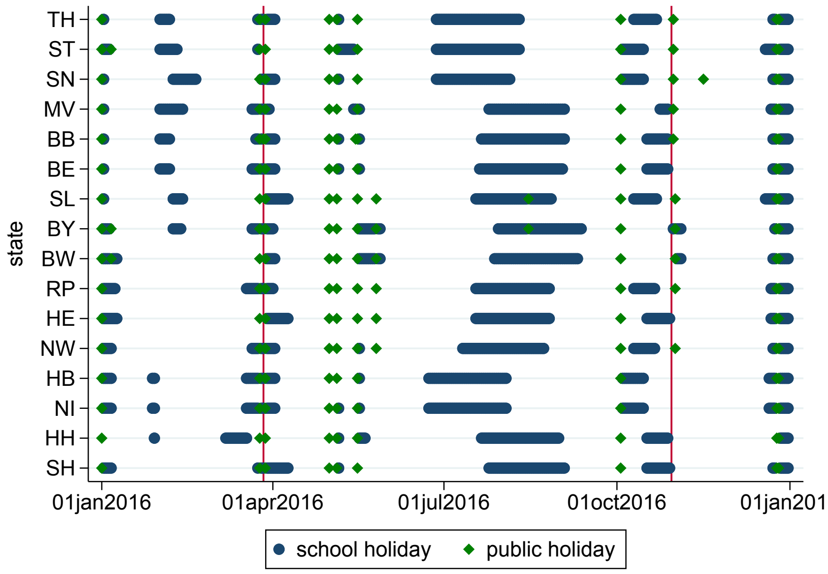


Figure 2.1: Holidays across States in 2016

Notes: The figure illustrates the variation in school and public holidays over states in the year 2016. The first red lines represent the change from ST to DST; the second red line the change back from DST to ST.

Weather. The state level aggregation of the DGUV data set limits the possibility to accurately control for extreme weather, since it varies at a more local level. Nevertheless, I try to explain some of the variation by including daily weather measures provided by the German Meteorological Service (Deutscher Wetterdienst) from various stations across the country.[14] These include data on the daily mean, maximum and minimum of air temperature, the daily minimum air temperature at 5 cm above ground, the daily precipitation, snow depth, and sunshine duration, the daily mean of wind speed and maximum of wind gust, and the mean of pressure, of cloud cover, of vapor pressure, and of relative humidity. The majority of weather stations stops to report snow depth during the months May to September. I impute snow depth by 0 cm if these closures seem to be systematic due the absence of snow in summer.[15] I interpolate weather to the municipality centroids and weigh by socially insured employees per municipality to receive the state level means.[16]

[14] DWD Climate Data Center (CDC): Historical daily station observations (temperature, pressure, precipitation, sunshine duration, etc.) for Germany, version v006, 2018

[15] I impute snow depth by 0 cm for the months May to September if the station has non-missing values for at least 95% of the days for the remainder of the year and one of the following conditions is met: (i) The station does not report snow depth at all from May to September and records a snow depth of 0 cm on April 30 and October 1. (ii) The station does not report snow depth for at least an entire calendar month between May and September and reports a snow depth of 0 cm on all days with non-missing values during this period. (iii) The station does not report snow depth for at most 28 days between May and September and reports a snow depth of 0 cm on all days with non-missing values during this period.

[16] Weather measures are interpolated by inverse distance weighting ($weight = 1/distance^2$), taking up to four nearest weather stations with non-missing values into account. The set of neighbors is restricted in the following

2.3.3 Summary Statistics

Using daily data for the 16 German states in the years 2013 to 2017, the estimation sample comprises 29,216 observations. Table 2.1 captures the corresponding employment weighted summary statistics.[17] The (weighted) mean number of 302.82 accidents per state-day observation is caused by on average 3.62 million employees (subject to social insurance). This translates into 8.44 accidents per 100,000 employees, consisting of 6.94 workplace accidents and 1.5 commuting accidents. 3% of the days are public holidays, while further 22% are school holidays.[18] Furthermore the table displays the mean, minimum, and maximum of all weather variables.

Table 2.1: Summary Statistics

	Mean	Std Dev	min	max	N
Work-related accidents (absolute numbers)					
(all) work-related accidents	302.82	271.12	0.00	2,106.43	29,216
workplace accidents	250.83	224.09	0.00	1,324.03	29,216
commuting accidents	52.00	60.54	0.00	1,257.60	29,216
Labor market variables					
Number of employees (in Mill.)	3.62	2.12	0.30	6.86	29,216
Work-related accidents (per 100,000 employees)					
(all) work-related accidents	8.44	5.40	0.00	152.35	29,216
workplace accidents	6.94	4.39	0.00	51.37	29,216
work-related travel accidents	1.50	1.73	0.00	100.98	29,216
Holiday indicators					
public holiday	0.03	0.17	0.00	1.00	29,216
school holiday	0.22	0.42	0.00	1.00	29,216
christmas holiday	0.03	0.17	0.00	1.00	29,216
day before public holiday	0.03	0.16	0.00	1.00	29,216
day after public holiday	0.03	0.16	0.00	1.00	29,216
Weather variables					
mean of air temperature (in °C)	10.19	7.01	-12.04	29.82	29,216
maximum of air temperature (in °C)	14.56	8.35	-9.39	37.84	29,216
minimum of air temperature (in °C)	5.89	6.10	-17.82	21.51	29,216
min. air ground (at 5 cm) temp. (in °C)	3.95	6.20	-19.85	19.71	29,216
precipitation height (in mm)	2.04	3.69	0.00	117.90	29,216
snow depth (in cm)	0.47	1.59	0.00	23.44	29,216
sunshine duration (in h)	4.43	4.05	0.00	16.01	29,216
maximum of wind gust(in m/s)	10.03	3.52	2.13	32.38	29,216
mean of wind speed (m/s)	3.33	1.40	0.74	13.14	29,216
mean of cloud cover	5.65	1.87	0.00	8.00	29,216
mean of vapor pressure (in hPa)	10.23	4.00	1.98	23.54	29,216
mean of pressure (in hPa)	986.03	20.11	923.50	1,041.28	29,216
mean of relative humidity (in %)	78.76	10.77	34.58	99.89	29,216

Notes: Sample descriptives for the years 2013-2017, weighted by the number of employees subject to social insurance. Data sources: German Social Accident Insurance (DGUV), German Federal Employment Agency, standing conference of the ministers of education and cultural affairs of the Länder in the Federal Republic of Germany, schulferien.org, and German Meteorological Service (DWD) Climate Data Center.

way: Stations must be based within 100 km of the centroid, be observed throughout all days of a year, and have non-missing values for at least 95% of the days for a given weather measure and year.

[17] Table B.1 provides the respective unweighted summary statistics. The mean number of accidents is 8.67 per 100,000 employees (7.06 workplace accidents and 1.6 work-related travel accidents).

[18] Public holidays are not treated as school holidays here.

2.4 Variation in Work-Related Accidents

To estimate causal effects of seasonal time changes in an RD design, I need to account for systematic variation in the number of work-related accidents. In this section, I provide descriptive evidence on how calendar effects and long-run trends affect this number. Figure 2.2 gives a first idea of the underlying variation of work-related accidents. It plots daily accidents at the national level over the years, weeks within a year, and days of the week, and reports both the absolute number and the rate (per 100,000 employees). Compared to the other categories, the variation over the years seems rather small. In line with previous literature (e.g., Robb and Barnes (2018); Poland et al. (2019)), the accident count decreases continuously over the course of a week, with a particularly strong reduction on the weekend. The crash profile over a given year is particularly low at the beginning and end of the year and exhibits considerable seasonal variation. To get a better idea from what this fluctuation stems from, Figure 2.3 shows the prevalence of work-related accidents over the year 2016.

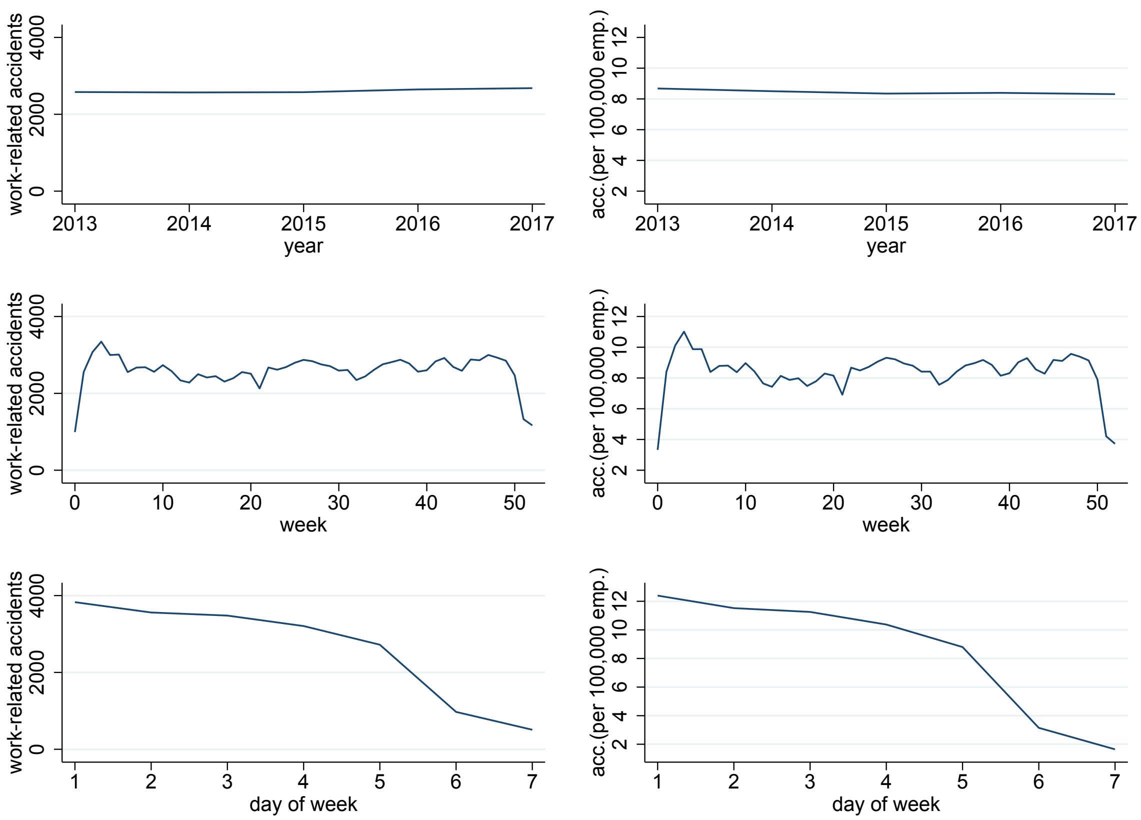


Figure 2.2: Work-Related Accidents by Year, Week, and Day of Week

Notes: This figure illustrates the incidence of daily work-related accidents in Germany by year, week of year, and day of week. The figures on the left side use the absolute number of work-related accidents, while the ones on the right side present them per 100,000 employees (subject to social insurance.) Week 1 starts on the first Monday of the year (middle panel), and Monday is the first day of the week (bottom panel).

While the mean number of accidents is 2647.70 per day, their occurrence seems to vary a lot over the different days of the year. Besides capturing the day-of-week variation, the figure suggests that holidays matter. On national public holidays the frequency is comparable to the one on Sundays. This can be observed for Good Friday and Easter Monday at the end of March. Likely the incidence of accidents is also lower during school holidays than in normal working weeks, in line with a reduced propensity to work. This pattern is, however, less obvious due to the heterogeneity of state-dependent school holidays (see Section 2.3.2). For that reason, I have a closer look at how work-related accidents vary over different days of the week and with respect to public holidays and school holidays in Figure 2.4. Moreover, there is sizable dispersion in the data which cannot exclusively be explained by the factors mentioned so far. It gets even more evident when looking at work-related travel accidents and workplace accidents separately (see figures B.1a and B.1b in the Appendix). The winter months are more prone to outliers, especially in the case of work-related travel accidents. Plausibly, extreme weather contributes to this variation. Therefore, I combine the accident data with information on school and public holidays, as well as weather measures. Given the different population sizes and thus accident counts over states, I normalize by dividing by 100,000 employees (subject to social insurance).[19]

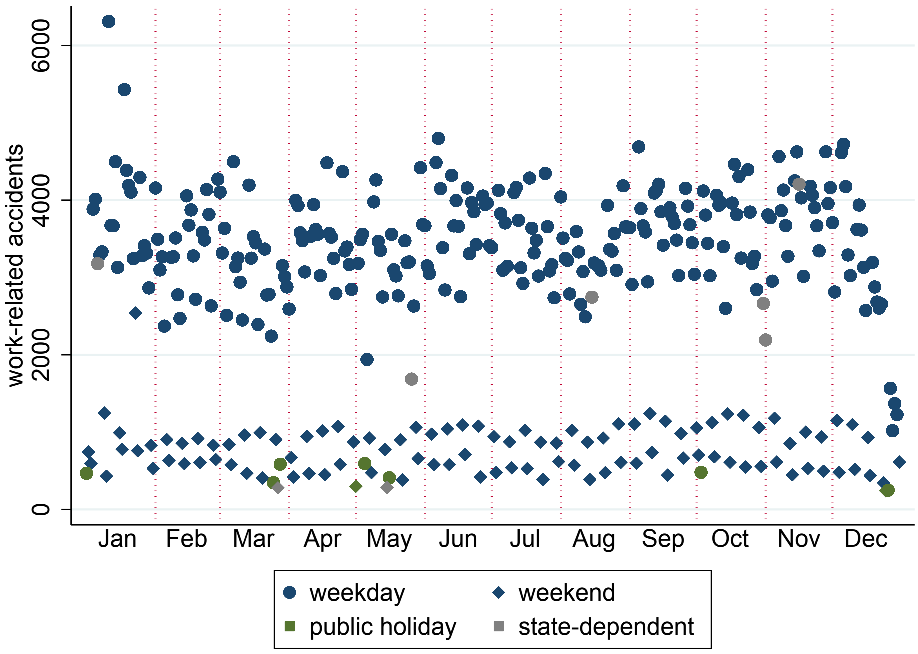


Figure 2.3: Work-Related Accidents in 2016

Notes: The figure illustrates the absolute number of work-related accidents per day over the year 2016. The red dashed lines represent the first day of a month.

Figure 2.4 shows the mean number of accidents by day of week and type of day, distinguishing between public holidays, school holidays and the remaining "normal days". First, the figure

[19]The data were provided by the German Federal Employment Agency on October 5, 2018. Since I do not have daily information on insured employees or working hours, I use the monthly data on employees subject to social insurance by state.

confirms the pattern of a falling incidence over the week, with a strong reduction towards the weekend. Second, the number of accidents is fundamentally lower on public holidays than on the corresponding working days, irrespective of the day of the week. Third, fewer accidents seem to occur during school holidays than on normal working days. Thus when estimating the impact of the DST regime on work-related accidents, it is important to account for heterogeneity coming from type of the day and day of the week. Figure B.2 further shows the mean accident counts during different school holidays and the surrounding weeks.[20] Besides mirroring the seasonal variation in work-related accidents, the figures support generally lower accident numbers during school holidays than during the surrounding 'normal' weeks. This gap is however much bigger in the case of the Christmas holidays. For that reason, I separately account for Christmas holidays in my estimation strategy. In a robustness check, I alternatively exclude December and January from my estimation sample.

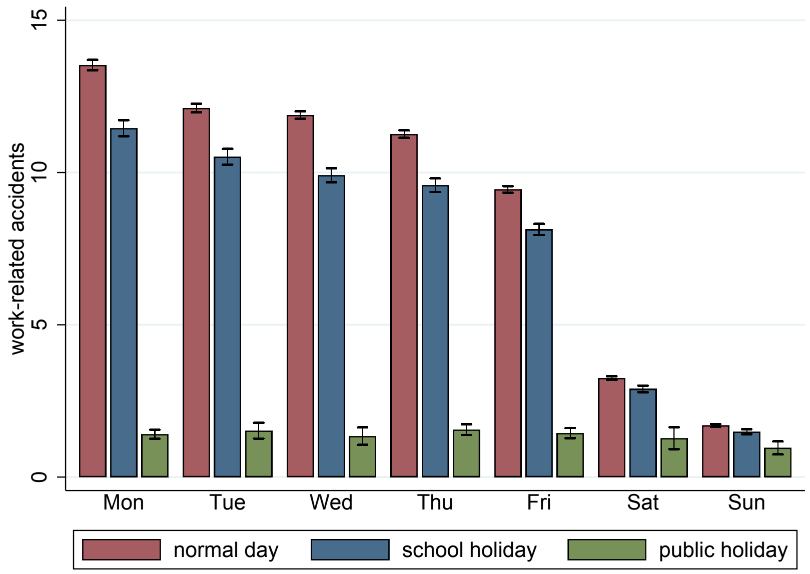


Figure 2.4: Work-Related Accidents by Type of Day and Day of Week

Notes: The figure illustrates the mean number of work-related accidents (per 100,000 employees) by type of day and day of week. The whiskers indicate 95% confidence intervals. *normal day* refers to days which are neither school holidays nor public holidays. Numbers are weighted by the number of employees subject to social insurance. The figure is inspired by Cox (2009).

Naturally, the number of accidents varies within type of day-day of week combinations. Figure B.3a illustrates the variation within these categories on the state-day level. There is a high variation in accidents over the different working days and a non-neglectable number of zero accidents, especially on public holidays and Sundays. Part of this variation and zero-accident observations may be attributable to the sampling error.[21] As Figure B.4 captures, a considerable

[20] School holidays are compared with non-school holidays which occurred one week before and one week after a school holiday in the respective state. Public holidays are excluded form both groups.

[21] When looking at workplace and commuting accidents separately, there is naturally a higher share of zero-accident observations. In the case of commuting accidents even the 5th percentile of a working day is zero (see Figure B.3c).

amount of variation remains when looking at 5-year state means instead of daily observations. These differences between states may occur for various reasons, for example due to differences in sectors, working hours, or population density. Thus, I variably control for state-day interactions. Moreover, given the presence of zero-accident observations, I use a non-logarithmized outcome variable.

Public holidays may not only affect work-related accidents directly but also indirectly. Employees might reduce their labor supply by reducing their working hours or by taking up vacation on days preceding or following a public holiday. This may especially but not exclusively be the case on bridge days, days between a public holiday and the weekend. In this spirit, Figure 2.5 further splits up the non-school and non-public holiday days into the day preceding a public holiday, the day following a public holiday and the remainder. The lowest accident prevalence occurs on bridge days, i.e., Mondays preceding a public holiday or Fridays following one. But the figure further suggests fewer accidents on Tuesday-Thursday and Sunday before a public holiday and on Monday, Thursday, and Saturday after one. For that reason, my estimation strategy also accounts for whether a day follows and/or precedes a public holiday.

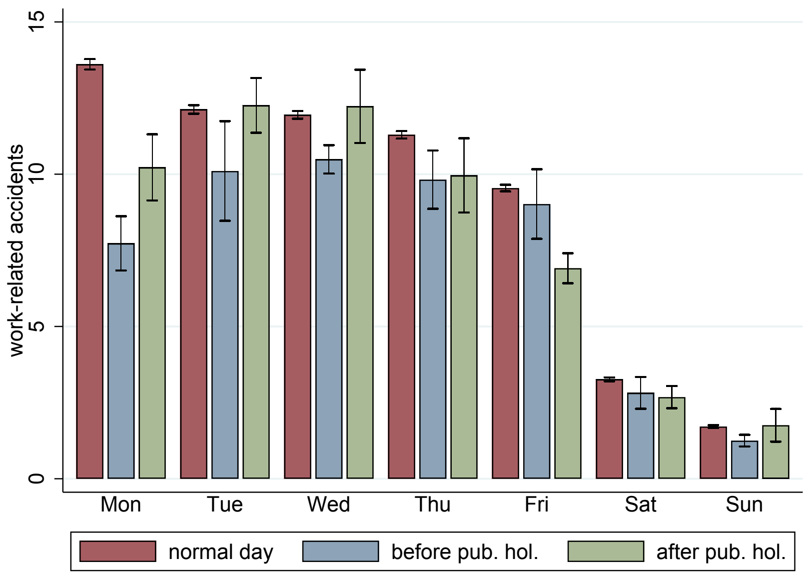


Figure 2.5: Work-Related Accidents before/after a Public Holiday

Notes: The figure illustrates the mean number of work-related accidents (per 100,000 employees) by type of day and day of week. The whiskers indicate 95% confidence intervals. *normal day* refers to days which are neither school holidays nor public holidays. Numbers are weighted by the number of employees subject to social insurance. The figure is inspired by Cox (2009).

2.5 Empirical Strategy

My empirical analysis aims to identify the causal effects of the DST regime on work-related accidents. I do so in exploiting the discrete change from ST to DST in spring and back from ST to DST in fall. If the transition has a significant effect on work-related accidents, there should be a jump in the number of accidents from just before to just after the clock change.

In the spirit of Smith (2016), I demean my outcome measure first by persistent calendar effects and long-term time trends before estimating a standard RD specification with the demeaned accident data. In my setting, however, it is not sufficient to just sort out day-of-week and year effects. The different public and school holiday regulations of the different states require both a disaggregated analysis and the following demeaning procedure:

$$Acc_{s,d,y} = \alpha + \eta_{s,k} + Holiday'_{d,y}\delta_k + \theta_y + X'_{s,d,y,k}\,\beta + \upsilon_{s,d,y} \tag{2.1}$$

$Acc_{s,d,y}$ is the number of accidents (per 100,000 employees) in state s on day d of year y. Additionally, k denotes day of the week. $Acc_{s,d,y}$ is regressed on state-specific day-of-the-week (k) effect $\eta_{s,k}$, year fixed effect θ_y and on holiday-specific day-of-the-week effects captured in δ_k. $Holiday_{d,y}$ is a vector including 5 indicators whether day d in year y is a public holiday, the day before a public holiday, the day after a public holiday, a school holiday, and a day in the Christmas holidays. Due to the additivity of the model, public holidays are set to take the value 0 for the other four indicators. I phase in weather controls in $X_{s,d,y,k}$. Given the strong variation in base accident levels over different types of days the effect of weather plausibly also varies with these. Thus, I include interactions of weather with state-day-of-week and holiday-day-of-week.[22] I proceed with taking the residuals $\upsilon_{s,d,y}$ from equation 2.1 as dependent variable in the following regression discontinuity specification:

$$\upsilon_{s,d,y} = \beta_0 + \tau DST_{d,y} + f(DaysToTran_{d,y}) + f(DST_{d,y} * DaysToTran_{d,y}) + \epsilon_{s,d,y} \tag{2.2}$$

$DST_{d,y}$ is a binary variable that equals 1 if DST is in place on day d in year y and 0 otherwise. $DaysToTran$ is the running variable with respect to the day of the transition and allows me to control for seasonal trends. It takes the value 0 on the last Sunday in March and tracks otherwise the number of days d is before/after the transition date. Thus τ is the coefficient of interest, capturing the effect of the transition to DST on work-related accidents.[23] My estimation relies on the data-driven local-polynomial-based inference procedures provided by Calonico et al. (2014b; 2014a; 2017; 2018; 2019; 2020, hereafter CCFT). In my baseline specification, I choose the bandwidth such that the asymptotic mean squared error (MSE) of τ is minimized. Moreover, I use a local-linear function with triangular kernel weighting, and mass point adjustment.The regressions are employee-weighted and nearest-neighbor based standard errors account for state-level clustering. To provide a consistent estimate of the effect of DST in the RD design, the conditional expectation functions of $\upsilon_{s,d,y}$ need to be continuous around the threshold.

[22] In Table 2.6, I further provide results estimated with less-interacted weather variables.

[23] A similar specification is used to assess the effect of switching from DST to ST. In this case the coefficient of interest is $ST_{d,y}$ and $DaysToTran_{d,y}$ is centered around the last Sunday in October.

Time adjustment might have a pure mechanical effect on the number of accidents on the day of the transition. Since switching from to DST (to ST) shortens (extends) the day by one hour, the accidents might be simply affected by the duration of the day. To tackle this problem, Smith (2016), for example, multiplies the crash count between 1 AM and 2 AM on the transition day in spring by two (and divides accordingly the crash count between 2 AM and 3 AM by two for the day of the fall transition). Comparably, Janszky et al. (2012) multiply the number of acute myocardial infarctions by 24/23 and 24/25, respectively. I use the reported number of accidents on the day of the transition, but assess the robustness of the results with respect to scaling to 24 hours and to excluding the day of the transition. I do so, because I expect there to be a much lower incidence of work-related accidents in the night hours compared to the remaining day and thus a risk of over-estimation (under-estimation) of the number of accidents around the spring (fall) transition.[24]

2.6 Results

2.6.1 Main Results

My main results provide no systematic evidence for an impact of seasonal time changes on work-related accidents. Table 2.2 shows the results from a regression discontinuity design (RD) on residual accidents, as explained in Section 2.5. The estimates in the odd columns rely on the basic demeaning procedure of equation 2.1, without weather variables in the first step. In the even columns, also weather variables are netted out, before estimating the discontinuity caused by the respective time change. The size of the discontinuity is estimated by means of a data driven bandwidth-selection and a local linear function with triangular kernel weighting (CCFT). While columns 1-2 show the impact on the joint accident measure, columns 3-6 provide the coefficients for workplace accidents and commuting accidents separately.

The coefficients on the impact of the transition to DST are provided in the upper part of the table. Although the baseline estimates are sizable in absolute terms, they are insignificant. Column 1 captures a coefficient of 0.2868 on general work-related accidents, which is smaller than the corresponding standard error. The estimate for workplace accidents amounts to half of the size of the overall estimate and is insignificant, too. In relative terms, it constitutes 2.2% of the sample mean. It is worth noticing, that the coefficient on commuting accidents in column 5 just falls short of being significant at the 10% level. However, its magnitude is economically important: it indicates 0.14 additional commuting accidents per 100,000 employees, which corresponds to a 9.1 percent effect compared to the sample mean. This lack of precision is partially induced by measurement error in the accident numbers. The estimates change little with the inclusion of weather measures. While the coefficient on workplace accidents is estimated slightly larger, the estimated impact on commuting accidents stays rather constant. Generally,

[24] This is in line with the time-of-day variation of fatal work accidents in the catchment area of the Institute of Legal Medicine in Frankfurt am Main, as well as of occupational injuries of male workers in Queensland, Australia.(Dieterich et al., 2020; Wigglesworth, 2006)

the estimates for the sub-outcomes, workplace accidents and commuting accidents roughly add up to the coefficient for all work-related accidents. Figure 2.6 provides a visualization of the results.

Table 2.2: Impact of DST Transitions on Work-Related Accidents

	All accidents		Workplace accidents		Commuting accidents	
	(1)	(2)	(3)	(4)	(5)	(6)
Transition to DST						
DST	0.2868	0.3303	0.1560	0.1903	0.1376	0.1364
	(0.2952)	(0.2842)	(0.2172)	(0.2143)	(0.0894)	(0.1026)
Observations	3600	4240	4240	4400	4240	4080
Dep. var. (mean)	8.445	8.445	6.94	6.94	1.505	1.505
Relative effect	.034	.039	.022	.027	.091	.091
Transition to ST						
ST	0.1048	-0.0053	0.0943	0.0101	-0.0251	-0.0216
	(0.2440)	(0.2239)	(0.2257)	(0.2099)	(0.0722)	(0.0618)
Observations	3120	3440	3440	3760	3760	3440
Dep. var. (mean)	8.445	8.445	6.94	6.94	1.505	1.505
Relative effect	.012	-.001	.014	.001	-.017	-.014
Year FE	✓	✓	✓	✓	✓	✓
State-dow FE	✓	✓	✓	✓	✓	✓
Holiday-dow FE	✓	✓	✓	✓	✓	✓
Weather variables		✓		✓		✓

Notes: Dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week (and weather variables). All specifications use the common MSE-optimal bandwidth selector for the RD treatment effect by CCFT, a first-order polynomial, and a triangular kernel. DST (ST) is the estimate of the discontinuity in work-related accidents at the transition into DST (out of DST). Nearest neighbor based standard errors are clustered at the state level (in parentheses). *Relative effect* reports the estimate relative to the sample mean of the dependent variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The bottom part of Table 2.2 shows the estimates on the impact of the transition to ST. They are generally smaller in absolute terms than their spring counterparts and highly insignificant. Interestingly, the baseline estimation provides small positive coefficients for the general measure and accidents in the workplace, in columns 1 and 3. Both approach zero when including weather controls, with the estimate for all accidents even turning marginally negative. There is a small, insignificant, negative coefficient on commuting accidents in column 5, which changes little when accounting for weather. Overall, there is little indication of effects around the fall transition. For that reason, I will focus on the consequences of the spring transition in the remainder of this section. However, I run similar estimations for the transition back to ST. These, corroborate

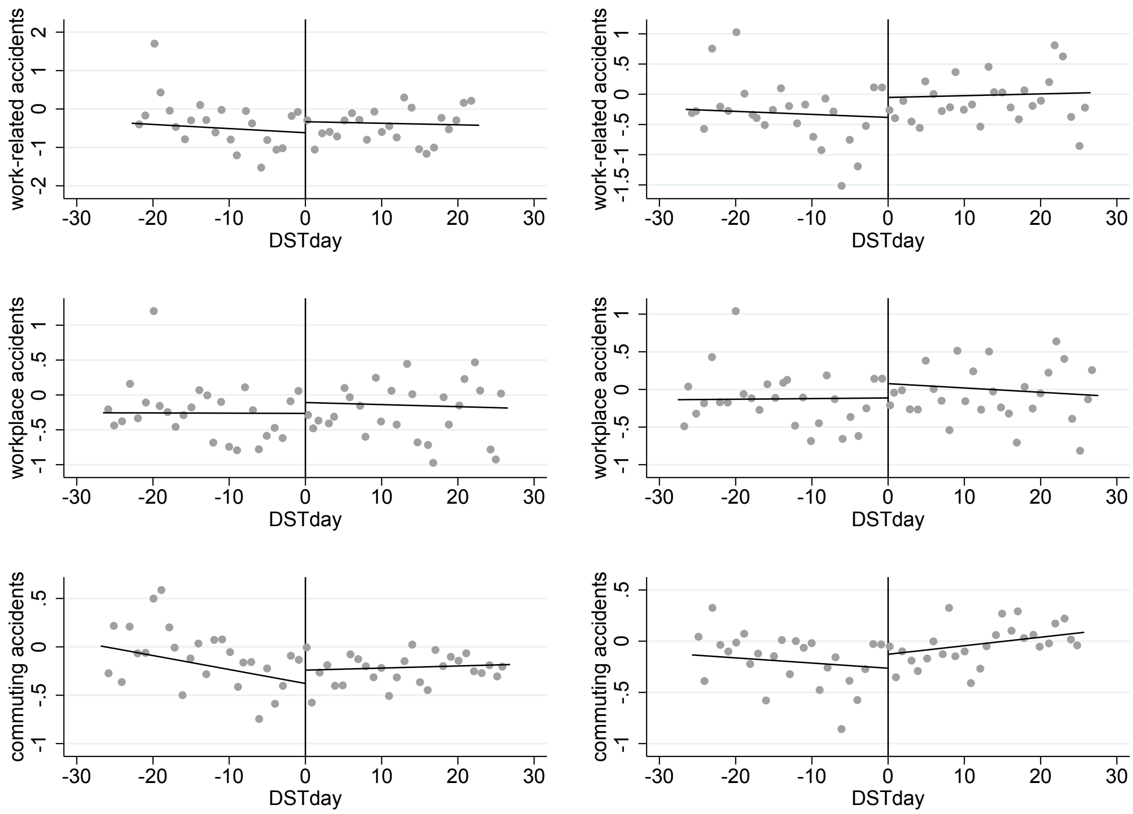


Figure 2.6: Impact of the Transition to DST on Work-Related Accidents

Notes: The figure provides RD plots, corresponding to the upper part of Table 2.2. In the plots on the left side, the dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week. In the plots on the right side, the dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week, and weather variables. The points represent the weighted average of residuals per day relative to the transition to DST. All specifications use the common MSE-optimal bandwidth selector for the RD treatment effect by CCFT, a first-order polynomial, and a triangular kernel.

my general findings of no effects of switching the clocks back on work-related accidents. The corresponding figures and tables can be found in the Appendix.

Selecting the appropriate bandwidth and assessing the sensitivity with respect to other bandwidth choices is essential when employing an RD design. Although the results in Table 2.2 use data-driven bandwidths, there are alternatives to relying on a common mse-optimal bandwidth selector (mserd). For that reason, I also estimate the coefficients with four alternative bandwidth selectors, provided by CCFT.[25]

Table 2.3 presents the corresponding coefficients. Results are quite similar to the main estimations, with the exception of employing a separate two-sided mse-optimal bandwidth selection (msetwo), in the columns 1 and 2. While it leads to smaller estimates than the three other selectors for workplace accidents (second panel), it provides the largest coefficients for commuting accidents (third panel). One has to note that these results are based on much larger bandwidths, taking up to 61 days on both sides into account. The estimates for commuting accidents range between 0.1371 and 0.15 in the base regression (odd columns). For the msetwo selector, this represents a significant increase by 0.15 commuting accidents per 100,000 employees. Adding weather in the demeaning procedure gives further rise to the two-sided estimate to a highly significant 0.202. Contrary, the coefficients remain rather unchanged for the other bandwidth selectors. Moreover, none of the estimates in the columns 3-8 is significant. This also applies to all estimates for the joint accident measure as well as for workplace accidents. Summing up, Table 2.3 provides evidence in support for my earlier findings, both in terms of the effect size and the non-significance of the effects.

Sampling only a subset of employees and projecting their accident numbers to the overall population, creates measurement error in the dependent variable. While this, in expectation, does not affect the mean number of accidents, it increases the underlying variance. Thus, the sampling procedure does not bias my estimates, but adds noise to the estimation and reduces the chance of rejecting the null hypothesis. In this light, the stable coefficients and sizable relative effects suggest a potential increase of commuting accidents by 8.9% to 10% following the transition from ST to DST.

My results are in line with the existing literature. Only one of the previous studies, provides significant results of the DST regimes on work-related accidents. Barnes and Wagner (2009) however, do so for the occupational group of miners. Given the plausibly more strenuous work conditions and more rigid time schedules, their larger and significant estimates of workplace accidents seem not to be surprising. While Robb and Barnes (2018) do not identify an effect on work accidents either, commuting accidents are included in road accidents in their study. They

[25]The first alternative selector employs two different MSE-optimal bandwidth selectors to the two sides of the cutoff (msetwo), which likely results in an asymmetric bandwidth. The second one, is a common selector which minimizes the asymptotic MSE of the sum (instead of the difference) of the two conditional expectation functions (msesum). Selector number 3 (msec2) takes two separate bandwidths based on the median of the previous selectors(mserd, msetwo, msesum) on each side of the threshold. The fourth alternative (cerrd), is a common coverage error rate (CER) optimal bandwidth selector for the RD estimator τ. It chooses the common bandwidth to both sides of the threshold based on minimizing the CER instead of the MSE (Calonico et al., 2017).

Table 2.3: Impact of the Transition to DST with Alternative Bandwidth Selectors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All accidents								
DST	0.1346	0.3751	0.2859	0.3297	0.2843	0.3300	0.2218	0.3284
	(0.2826)	(0.2554)	(0.2961)	(0.2821)	(0.2958)	(0.2831)	(0.3226)	(0.3058)
Observations	6960	7600	3600	4400	3600	4320	3120	3600
Dep. var. (mean)	8.445	8.445	8.445	8.445	8.445	8.445	8.445	8.445
Relative effect	.016	.044	.034	.039	.034	.039	.026	.039
Workplace accidents								
DST	0.0792	0.1627	0.1537	0.1873	0.1565	0.1898	0.1375	0.1935
	(0.2319)	(0.2158)	(0.2149)	(0.2137)	(0.2170)	(0.2142)	(0.2384)	(0.2306)
Observations	7200	7360	4400	4560	4320	4480	3600	3760
Dep. var. (mean)	6.94	6.94	6.94	6.94	6.94	6.94	6.94	6.94
Relative effect	.011	.023	.022	.027	.023	.027	.02	.028
Commuting accidents								
DST	0.1500**	0.2019***	0.1431	0.1361	0.1371	0.1361	0.1465	0.1339
	(0.0765)	(0.0740)	(0.0936)	(0.1019)	(0.0905)	(0.1019)	(0.0998)	(0.1111)
Observations	9760	9040	4080	4240	4160	4240	3600	3440
Dep. var. (mean)	1.505	1.505	1.505	1.505	1.505	1.505	1.505	1.505
Relative effect	.1	.134	.095	.09	.091	.09	.097	.089
Bandwidth selector	msetwo	msetwo	msesum	msesum	msec2	msec2	cerrd	cerrd
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State-dow FE	✓	✓	✓	✓	✓	✓	✓	✓
Holiday-dow FE	✓	✓	✓	✓	✓	✓	✓	✓
Weather variables		✓		✓		✓		✓

Notes: Dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week (and weather variables). All specifications use the common MSE-optimal bandwidth selectors by CCFT, a first-order polynomial and a triangular kernel. *msetwo*: different MSE-optimal bandwidth selectors to the two sides of the cutoff. *msesum* one common bandwidth selector which minimizes the asymptotic MSE of the sum of the conditional expectation functions. *msec2:* two separate bandwidths based on median(mserd, msetwo, msesum). *cerrd:* one common coverage error rate (CER) optimal bandwidth selector for the RD estimator DST is the estimate of the discontinuity in work-related accidents at the transition into DST. Nearest neighbor based standard errors are clustered at the state level (in parentheses). *Relative effect* reports the estimate relative to the sample mean of the dependent variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

find a rise in road accidents on the first two days of DST by 16% respectively 12%. When one further includes their non-significant coefficients for the following 4 days (of on average 6%), their effect averages to 8.67% which is in line with my insignificant estimates of around 9.1% on commuting accidents. The smaller increase of 5.6% observed by Smith (2016) in fatal road accidents could be due to a stronger increase in less severe accidents than fatal ones or a smaller death toll on recreational driving. Bünnings and Schiele (forthcoming) find (marginal) support for an increase in fatal accidents but not accidents in general in Great Britain, which speaks in favor of the latter. In this light, further research on commuting accidents and their severity would be beneficial but calls for larger sample sizes to increase the precision of the estimates.

2.6.2 Robustness

Kernel choice and transition date. General concerns often include the choice of the kernel and the treatment of the cutoff date. In my baseline specification, I take the number of accidents as sampled and employ triangular kernel weighting. Table 2.4 assesses the robustness with respect to the day of the transition and the kernel used in the local linear estimation. It provides estimates for alternatives as scaling the day of the transition, dropping it from the estimation, uniform kernel weights, and epanechnikov kernel weights. As discussed earlier, there might be a pure mechanical effect on the number of accidents on the day of the transition due to the clock change. Since the change to DST is associated with moving the clock forward, the day of the transition has 23 instead of 24 hours. This might lead to an underestimation of the true effect. Given the daily nature of my data and literature indicating smaller accident counts during the night hours (Dieterich et al., 2020; Wigglesworth, 2006), I do not to scale the number of cases on the day of the transition in my main specification. An overestimation of the incidence of work-related accidents on the day following the transition to DST might likely lead to an upward bias in the estimate. While columns 1 and 2 provide the estimates from my main specification, the coefficients in columns 3 and 4 stem from multiplying the accidents on the day of the transition by 24/23, as Janszky et al. (2012) do. Unsurprisingly, the estimates marginally increase but remain insignificant. Excluding the day of the transition results in larger coefficients for workplace accidents and smaller estimates for commuting accidents, as captured in columns 5 and 6.

Alternating the kernel leads to a more nuanced picture. For the joint accident measure and workplace accidents all estimates are elevated a bit, without affecting their insignificance. Contrary, a uniform kernel entails smaller coefficients for commuting accidents, especially when accounting for weather with an estimate of 0.0986 in column 8. Using epanechnikov weighting on the other side, leads to a slight increase in the estimated effects. The baseline regression in column 9 captures a (marginally) significant rise in travel-to-work accidents by 9.6%, which is rendered insignificant once accounting for weather (column 10). Overall, my findings are qualitatively not impacted by changes in the kernel weighting or the transition date procedure.

Table 2.4: Impact of the Transition to DST on Work-Related Accidents with respect to Kernel and Transition Date

	baseline		scaling day of transition		exclude day of transition		uniform kernel		epanechnikov kernel	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All accidents										
DST	0.2868	0.3303	0.2984	0.3472	0.2465	0.3766	0.3314	0.4044	0.3287	0.3505
	(0.2952)	(0.2842)	(0.2957)	(0.2846)	(0.3243)	(0.3214)	(0.2635)	(0.2795)	(0.2936)	(0.2836)
Observations	3600	4240	3600	4240	3520	3840	3120	3440	3440	4080
Dep. var. (mean)										
Relative effect	.034	.039	.035	.041	.029	.044	.039	.048	.039	.042
Workplace accidents										
DST	0.1560	0.1903	0.1652	0.2066	0.1652	0.2460	0.1855	0.2101	0.1795	0.2126
	(0.2172)	(0.2143)	(0.2174)	(0.2144)	(0.2330)	(0.2330)	(0.2226)	(0.2091)	(0.2198)	(0.2169)
Observations	4240	4400	4240	4400	4160	4160	2960	3120	3920	3920
Dep. var. (mean)										
Relative effect	.022	.027	.024	.03	.024	.035	.027	.03	.026	.031
Commuting accidents										
DST	0.1376	0.1364	0.1389	0.1379	0.0986	0.1280	0.1254	0.0890	0.1438*	0.1418
	(0.0894)	(0.1026)	(0.0890)	(0.1020)	(0.1220)	(0.1397)	(0.0941)	(0.1025)	(0.0871)	(0.1025)
Observations	4240	4080	4240	4080	3680	3360	2640	2960	4080	3760
Dep. var. (mean)										
Relative effect	.091	.091	.092	.092	.065	.085	.083	.059	.096	.094
Transition day	Yes	Yes	24/23	24/23	No	No	Yes	Yes	Yes	Yes
Kernel	tri	tri	tri	tri	tri	tri	uni	uni	epa	epa
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
State-dow FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Holiday-dow FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Weather variables		✓		✓		✓		✓		✓

Notes: Dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week (and weather variables). All specifications use the common MSE-optimal bandwidth selector for the RD treatment effect by CCFT and a first-order polynomial. *tri* refers to triangular kernel weights, *uni* to uniform kernel weights, and *epa* to epanechnikov kernel weights. 24/23 multiplies the accidents on the day of the transition by 24/23. *No* excludes the day of the transition from the estimation. DST is the estimate of the discontinuity in work-related accidents at the transition into DST. Nearest neighbor based standard errors are clustered at the state level (in parentheses). *Relative effect* reports the estimate relative to the sample mean of the dependent variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Bias correction. Bandwidths selected by means of MSE-minimizing bandwidth selectors might be relatively large and pose a threat to the validity of the distributional approximations. These bandwidth selectors then can lead to a bias in the distributional approximation of the RD treatment effect. Calonico et al. (2014b) propose robust bias-corrected confidence intervals to account for this problem. First, the estimator is bias-corrected by recentering the t-statistic with an estimate of the leading bias. Second, the bias-corrected t-statistic is rescaled to account for the variability added by the bias estimate. Table 2.5 presents my main specification with such bias-corrected estimates and robust confidence intervals. Regarding all accidents and workplace accidents this adjustment does not alter the coefficients on the DST transition much. The estimated effect on commuting accidents increases marginally in the base specification. When accounting for weather the coefficient experiences a drop to 0.1108, at a t-value below 1. Naturally, this process leads to increased standard errors.

Table 2.5: Bias-Corrected RD Estimates with Robust Confidence Intervals

	All accidents		Workplace accidents		Commuting accidents	
	(1)	(2)	(3)	(4)	(5)	(6)
Transition to DST						
DST	0.2838	0.3219	0.1636	0.1989	0.1419	0.1108
	(0.3613)	(0.3361)	(0.2647)	(0.2585)	(0.1175)	(0.1243)
Observations	3600	4240	4240	4400	4240	4080
Dep. var. (mean)	8.445	8.445	6.94	6.94	1.505	1.505
Relative effect	.034	.038	.024	.029	.094	.074
Year FE	✓	✓	✓	✓	✓	✓
State-dow FE	✓	✓	✓	✓	✓	✓
Holiday-dow FE	✓	✓	✓	✓	✓	✓
Weather variables		✓		✓		✓

Notes: Dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week (and weather variables). All specifications use the common MSE-optimal bandwidth selector for the RD treatment effect by CCFT, a first-order polynomial, a triangular kernel, and bias-correction (Calonico et al., 2014b). DST is the estimate of the discontinuity in work-related accidents at the transition into DST. Robust nearest neighbor based standard errors are clustered at the state level (in parentheses). *Relative effect* reports the estimate relative to the sample mean of the dependent variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Variance estimation. The fact that my analysis is on the state-day level and Germany only has 16 states, naturally raises questions on the appropriateness of the provided standard errors. Table B.6 provides the p-values of alternative variance estimations. As Calonico et al. (2014b) point out, nearest neighbor based standard errors are more robust in finite samples than such using plugin residuals, because the latter rely on the same bandwidth choices as the underlying regression functions. This is supported by the more conservative p-values of column 1 compared to column 2. Moreover, the p-values compare well against alternatives as bootstrapping, two-way-clustering and a permutation test for all accidents and workplace accidents. In the case of commuting accidents, however, the p-values of these alternatives are larger and further support the insignificance of my estimation.

Demeaning procedure. So far, all robustness checks were related to features of the RD design. Given the two steps which I employ, however, when estimating the effects of interest, the demeaning procedure is also an essential part of my estimation strategy. For that reason, I assess different variations of equation 2.1, before estimating the discontinuity on the residuals. Table 2.6 presents the corresponding results. While the first two columns provide the baseline estimates from Table 2.2, columns 3-6 vary the explaining variables, 7-8 the functional form, and column 9 restricts the sample.

First, phasing in less interacted weather variables, adds further imprecision to the estimates. On the contrary, a reduction in the standard errors of our estimates of interest can be observed when the holiday-day-of-week fixed effects are interacted with the state identifiers. As portrayed by Figure 2.4, a lower accident count can be observed on a day preceding or following a public holiday. Failing to account for this relationship in the days around Easter might lead to an overestimation of the causal effect. Consequently, columns 5 and 6 indicate the relevance of accounting for the days before and after a public holiday. Relaxing the control variables from before/after public holiday identifiers to just bridge days, elevates the estimated casualty rates in particular for workplace accidents, while extending it to the 2nd lag/lead bears coefficients quite similar to the base model.

In order to assess the risk of an overspecification of the demeaning procedure, I employ an elastic net regularization in column 7. The resulting RD estimates are much in line with the baseline results of column 1, and thus provide support for the set of interactions I use. Column 8 approaches the issue of a potential misspecification due to estimating a linear model on non-normally distributed variables due to frequent occurrence of non-accident observations. In fact, estimating a zero-inflated negative binomial model leads to estimates which are 10-20% smaller than in my main specification. An equal reduction in the coefficient is observed for commuting accidents when December and January are removed from the estimation sample, in column 9. This is done to account in a different way for the strong reduction in work-related accidents around the change of the year. The estimates of other two outcome measures are unaffected by this exercise. All in all, Table 2.6 provides estimates which are within a reasonable range of my main specification, supporting my estimation strategy. None of the coefficients indicates a significant effect of seasonal time changes on occupational accidents, neither at the workplace, nor during commuting. Yet, the estimates for commuting accidents amount to 7.8%-10% of the sample mean.

Table 2.6: Additional Robustness: Transition to DST

	baseline		explaining variables				functional form		sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
All accidents									
DST	0.2868	0.3303	0.3103	0.2869	0.3791	0.2708	0.2890	0.2507	0.2896
	(0.2952)	(0.2842)	(0.3104)	(0.2822)	(0.2974)	(0.2957)	(0.2950)	(0.2831)	(0.2658)
Observations	3600	4240	4400	3760	3600	3600	3600	3760	4080
Dep. var. (mean)	8.445	8.445	8.445	8.445	8.445	8.445	8.445	8.445	8.512
Relative effect	.034	.039	.037	.034	.045	.032	.034	.03	.034
Workplace accidents									
DST	0.1560	0.1903	0.1515	0.1518	0.2303	0.1522	0.1584	0.1246	0.1527
	(0.2172)	(0.2143)	(0.2203)	(0.2114)	(0.2199)	(0.2158)	(0.2172)	(0.2137)	(0.2079)
Observations	4240	4400	4880	4400	4240	4240	4240	4400	4560
Dep. var. (mean)	6.94	6.94	6.94	6.94	6.94	6.94	6.94	6.94	7.09
Relative effect	.022	.027	.022	.022	.033	.022	.023	.018	.022
Commuting accidents									
DST	0.1376	0.1364	0.1504	0.1289	0.1438	0.1329	0.1358	0.1167	0.1162
	(0.0894)	(0.1026)	(0.1165)	(0.0835)	(0.0897)	(0.0890)	(0.0882)	(0.0803)	(0.0813)
Observations	4240	4080	4400	4400	4240	4240	4240	4560	4400
Dep. var. (mean)	1.505	1.505	1.505	1.505	1.505	1.505	1.505	1.505	1.422
Relative effect	.091	.091	.1	.086	.096	.088	.09	.078	.082
b/a Pub. hol.-dow FE	✓	✓	✓	✓		✓	✓	✓	✓
Christmas Hol.-dow FE	✓	✓	✓	✓	✓	✓	✓	✓	
Weather variables		✓							
Fewer weather var.			✓						
State-Hol.-dow FE				✓					
Bridge days					✓				
2b/2a Pub. hol.-dow FE						✓			
Elastic net							✓		
Zero-inflated neg. bin.								✓	
w/o Jan. & Dec.									✓

Notes: Dependent variable is the residual number of work-related accidents per 100,000 employees. All specifications are demeaned by year, state-day-of-week, public holiday-day-of-week, and school holiday-day-of-week. Further *b/a Pub. Hol.-dow FE* includes an interaction between indicators for the day before and the day after a public holiday interacted with day-of-week, *Christmas Hol.-dow FE* separate Christmas holiday-day-of-week indicators, *Bridge days* includes indicators for a Monday before a public holiday and a Friday after a public holiday, *2b/2a Pub. hol.-dow FE* includes an interaction between indicators for the second day before and the second day after a public holiday interacted with day-of-week, *weather variables* includes weather measures interacted with state-day-of-week and holiday-day-of-week, *Fewer weather var.* includes weather measures interacted with day-of-week, state, and the five holiday indicators. *Elastic net* uses an elastic net regularization instead of a simple OLS regression and *Zero-inflated neg. bin.* estimates a zero-inflated negative binomial model. *w/o Jan & Dec.* excludes the months January from the estimation sample. All specifications use the common MSE-optimal bandwidth selector for the RD treatment effect by CCFT, a first-order polynomial, and a triangular kernel. DST is the estimate of the discontinuity in work-related accidents at the transition into DST. Nearest neighbor based standard errors are clustered at the state level (in parentheses). *Relative effect* reports the estimate relative to the sample mean of the dependent variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.7 Conclusion

This paper provides evidence on the impact of seasonal time changes on work-related accidents in Germany. It is the first paper in this literature considering commuting accidents as an explicit outcome variable. The discrete nature of DST transitions allows me to obtain causal estimates of this relationship and the DGUV data to distinguish between accidents that occurred at the workplace and such during commuting. I employ a thorough estimation procedure based on the data variation I uncover in descriptive statistics. After demeaning by persistent year, state, day-of-the-week, and holiday variation, I estimate a regression discontinuity design.

My results indicate no systematic effects of seasonal time changes on work-related accidents for the years 2013-2017: neither at the transition to Daylight Saving Time nor on the days following the transition back to Standard Time. Nevertheless, the estimates of the spring transition on commuting accidents stand out and provide suggestive evidence for an increase in accidents during commuting. The sizable effects are stable in magnitude across a rich set of alternative specifications, but are imprecisely estimated. Plausibly, sampling error impedes the precision of my estimation. The magnitude of my estimates and the previous findings on road accidents call for future research on the impact of seasonal time changes on the incidence of commuting accidents and their severity. Governments need to understand the potentially detrimental effects of seasonal time changes when designing future policies.

Chapter 3

All Geared Towards Success? Cultural Origins of Gender Gaps in Student Achievement*

3.1 Introduction

Over the past decades, reportage of girls "outgunning" boys in end-of-school examinations has repeatedly made headline news in many parts of the world. Yet there is more to it than that. Gender gaps in student achievement continue to exhibit large geographical and temporal variations. They differ across countries, across districts within countries, and across communities within districts (Pope and Sydnor, 2010; Reardon et al., 2019). As for changes over the past decades, while many countries have witnessed a closure or even reversal of male-favoring gender gaps in education, some countries have seen them widening (Evans et al., 2019). Given society's interest in tackling inequality of opportunity, it is important to understand the causes of these variations. A broad literature examines how gender gaps in student achievement are shaped by early childhood influences, characteristics of schools, and educational systems (Dee, 2005; Machin and McNally, 2008; Bertrand and Pan, 2013; Deming et al., 2014). Yet, much less is known about the extent to which boys and girls are differentially affected by entrenched cultural values transmitted from generation to generation.

Using a large Swedish administrative data set, we present new evidence on the cultural origins of gender gaps in student achievement. Our analysis complements existing work in two important dimensions. First, building on the epidemiological approach to culture (Carroll et al., 1994; Fernández and Fogli, 2006; Giuliano, 2007; Fernández and Fogli, 2009), which exploits cultural diversity among second-generation immigrants who share the same economic and institutional environment, we provide a powerful and tightly controlled test of the effect of culture on gender gaps in student achievement. Our test relies on within-family, cross-gender sibling comparisons,

*This chapter is based on joint work with Helena Holmlund and Helmut Rainer.

identifying the differential effect of culture on girls relative to boys after controlling, *inter alia*, for unobserved family heterogeneity and gender-specific neighborhood effects. Second, differently from existing studies that equate culture to a single aggregate economic variable or index from the country-of-ancestry, we operationalize culture in a multidimensional way.

In order to describe culture, social scientists have used the analogy of an onion with basic beliefs, values, and attitudes forming the core of culture, and actual behavior and manifestations thereof representing the outer layers (Taras et al., 2009). In applying the epidemiological approach, economists have often focused on outer-layer measures of culture. For example, in a paper examining the role of gender norms in explaining the gender gap in math, Nollenberger et al. (2016) used the World Economic Forum's Gender Gap Index to measure gender equality in an immigrant's country of ancestry, and examined whether this measure is able to explain variations in the gender math gap across second-generation immigrant youth living in 9 host countries. In this paper, we zoom in on beliefs, norms and attitudes that plausibly underlie manifestations of gender (in)equality in society, and analyze their effect on gender gaps in student achievement. To that end, we draw upon the work of Dutch sociologist Hofstede (2001), who was among the first to develop a coherent theoretical model of culture and corresponding cross-country indices describing cultural values along several dimensions. Importantly, if one connects Hofstede's multidimensional notion of culture with the economics literature on gender differences and gender convergence (see, e.g., Goldin, 2006; Bertrand, 2018; Niederle and Vesterlund, 2010), then one arrives at several potentially important cultural channels through which gender gaps in education may be affected.

Consider as a first cultural dimension the extent to which a society emphasizes ambition, competition, and achievement (labeled *masculinity vs. femininity* by Hofstede). This cultural trait might affect gender gaps in student achievement for various reasons. For example, it is conceivable that parents from achievement-oriented cultures choose higher quality schools for their offspring, exposure to which has been linked to an educational advantage of boys relative to girls (Autor et al., 2016). Moreover, a series of studies in behavioral economics have shown that males and females differ in their response to competition, with the effects in mixed-sex settings ranging from women failing to perform well in competitions (Gneezy et al., 2003) to women shying away from environments in which they have to compete (Niederle and Vesterlund, 2007). Thus, one might hypothesize that if parents transmit to their children achievement-oriented attitudes, this raises competitive pressures associated with test-taking, which may cause boys to outperform girls in class.

The second cultural dimension we single out is the extent to which a society accepts an unequal distribution of power (labeled *power distance* by Hofstede). Although women have made significant progress in the labor market around the world, they remain under-represented in high-earnings, high-power occupations (Bertrand, 2018; Bertrand et al., 2019). This phenomenon is commonly referred to as the glass ceiling. If parents and their children are accustomed to expect and accept unequal distribution of power, it may reinforce perceptions of the glass ceiling. This, in turn, may lead to differential parental investments in human capital of sons and daughters and act as an impediment to girls' education. Thus, one may hypothesize that among children from cultures with a high degree of inequality acceptance, we are less likely to observe girls having caught or overtaken boys in their academic attainment.

The third potentially relevant cultural dimension is a society's tolerance for uncertainty and ambiguity (labeled *uncertainty avoidance* by Hofstede). When women in the US started to increase their investments in formal schooling, they altered their identity in a way that placed a career on equal footing, or even ahead, of marriage (Goldin, 2006). Although this change was important for women's progress, it did not come without risks and uncertainty. For example, when women started to move away from "safe", traditionally female-dominated jobs in the public sector (e.g., teachers, nurses) to male-dominated fields, it involved the risks of breaking gender norms, of social rejection, and of professional failure. Hence, if children are socialized to avoid choices that involve risks and uncertainty, this may be an obstacle to females increasing career-oriented human capital investments. As a consequence, the likelihood of girls catching up with or overtaking boys educationally may be smaller.

The final cultural dimension we draw attention to is the extent to which members of society are willing to delay short-term material or social rewards in order to prepare for the future (labeled *long-term vs. short-term orientation* by Hofstede). Figlio et al. (2019) have shown that immigrant students in the US from countries with long-term oriented attitudes perform better than students from cultures with less emphasis on delayed gratification. Beyond this intriguing finding, there are also reasons to expect a link between long-term orientation and gender gaps in education. Specifically, Goldin (2006) has argued that a change from static decision-making with limited horizons to dynamic decision-making with long-term horizons was a key factor behind the "quiet revolution" that transformed women's education and employment in the US. Based on this, we hypothesize that we are more likely to observe girls having caught up with or overtaken boys educationally if parents transmit long-term oriented attitudes to their offspring.

We empirically analyze these hypotheses by relating gender gaps in student achievement among children of contemporary immigrants to the cultural characteristics of their parents' birth countries. Any such epidemiological approach faces a key identification challenge: to avoid conflating the effect of culture with the effect of unobserved family characteristics. Take, as an example, selection into neighborhoods: Since immigrants are not randomly assigned to neighborhoods within host countries, immigrant parents from a, say, achievement-oriented culture might select into better neighborhoods. If, in turn, girls and boys are differentially affected by neighborhood characteristics, as some studies suggest they are (see, e.g., Deming et al., 2014; Hastings et al., 2006), estimates of the effect of achievement orientation on gender gaps will suffer from selection bias. To overcome this identification challenge, we combine several registries from Statistics Sweden to construct a large administrative dataset which contains educational outcomes and background characteristics of almost 80,000 opposite-sex siblings, all born in Sweden but with parents who immigrated to the country from other nations. The main outcome variable in our investigation is students' grade-point average (GPA) at age 16. Using Hofstede's cross-country data, we assign to each student the cultural dimensions characterizing their countries of ancestry.

The resulting dataset allows us to run specifications that exploit within-family, cross-gender sibling comparisons, allowing us to not only separate out the impact of unmeasured family variables that are constant across siblings, but also to control, *inter alia*, for gender-specific neighborhood effects. To check for external validity, we re-examine our results for Sweden using data from five waves of the Program for International Student Assessment (PISA), which provides us with a sample of roughly 35,000 second-generation immigrant students residing in 29 host countries.

Our empirical analysis yields several interesting results. First, in our analysis of Swedish administrative data, we find that the central cultural dimension that matters for gender gaps in student achievement is the extent to which a society emphasizes ambition, competition, and achievement. In our population of interest—i.e., the universe of second-generation immigrant students graduating from 9^{th} grade in the period 1988-2017—girls reach, on average, GPAs that are 31 percent of a standard deviation higher than those of boys. Our main result shows that among children from countries with achievement-oriented attitudes, girls' comparative GPA advantage significantly decreases. For example, if immigrants from Denmark, a society that puts little emphasis on ambition, competition, and achievement (*Masculinity Score*=0.16), had the same degree of achievement orientation as those from Germany (*Masculinity Score*=0.66), our findings suggest that the mean GPA of girls relative to boys would decrease by roughly 38%. Another cultural dimension that matters for gender gaps in student achievement, but with an effect size only roughly half as large, is long-term orientation. In particular, and as hypothesized at the outset, a culture of long-term orientation is associated with an educational advantage of girls relative to boys. Hofstede's other two cultural dimensions have no, or at best small, effects on gender gaps in student achievement. These results remain qualitatively the same when examining student grades in mathematics and Swedish, and are not picking up source countries' characteristics that may affect girls and boys differentially.

Second, we explore mechanisms that may explain why cultural background has implications for the achievement gap between girls and boys. We consider four possibilities: (i) parents with different cultural beliefs might gender-discriminate when choosing schools for their children, i.e., enroll sons in higher quality schools than daughters; (ii) irrespective of their children's gender, parents from achievement-oriented cultures might enroll their offspring in higher quality schools, and boys might benefit more from exposure to higher quality schools than girls do (see, e.g., Autor et al., 2016); (iii) in parallel and not mutually exclusive to (ii), parents from achievement-oriented cultures might be positively selected in terms of SES or host country experience compared to immigrants from other cultures, and this could disproportionately promote the educational outcomes of boys (see, e.g., Autor et al., 2019); (iv) parents from achievement-oriented cultures might adopt more traditional role models than those from other cultures, and this in turn could explain an educational advantage of boys relative to girls. In examining these four possibilities, we find no evidence in favor of differential treatment of girls versus boys. Instead, we obtain results that are supportive of mechanisms (ii) and (iii), whereby parents from achievement-oriented cultures send their children to higher quality schools and have a stronger socioeconomic position, which in turn is more beneficial for boys. However, school quality and parental SES are far from fully explaining the impact of culture on the gender gap in education. The mechanisms through which cultural values are passed on and affect offspring are therefore remaining, in part, unobserved. When it comes to achievement orientation, the strongest predictor of gender gaps in student achievement, an important explanation might lie, as discussed in the outset, in the different reactions of girls and boys to competitive pressure.

Third, we find qualitatively and quantitatively remarkably similar results when we replicate our results for Sweden using PISA data. In our PISA sample of second-generation immigrants, girls have, on average, higher reading scores than boys, but they are outperformed by boys in math and science. Among children from achievement-oriented cultures, girls' comparative advantage in reading vanishes, while the math and science gap in favor of boys significantly increases. In other words, a culture of achievement orientation is associated with boys performing as well as girls in reading and much better than them in math and science. As with Swedish administrative data, Hofstede's other cultural dimensions are less prominent in explaining gender gaps in student achievement, a result that holds irrespective of whether we analyze each variable in isolation or run "horse-race" regressions between them. The fact that we obtain remarkably similar results in two very different samples of second-generation immigrants suggests that a cultural heritage that emphasizes ambition, competition and achievement plays a central role for the existence of educational disadvantages of girls relative to boys.

Our study contributes a unified set of sights on how cultural values along different dimensions shape gender gaps in student achievement. Our results broadly add to findings on culture's impact on various economic outcomes such as female work and fertility (Fernández and Fogli, 2006; Fernandez, 2007; Fernández and Fogli, 2009), education (Figlio et al., 2019), family living arrangements (Giuliano, 2007), or preferences for redistribution (Luttmer and Singhal, 2011). Fernández (2011) provides an insightful review of this strand of literature. More narrowly, our study has two important antecedents in the works of Guiso et al. (2008) and Nollenberger et al. (2016). Both studies, the former using an analysis across countries and the latter employing the epidemiological approach, provide evidence that more gender equality in society is associated with a higher math performance of girls relative to boys. Our analysis adds to these works by shifting focus to cultural values, beliefs, and attitudes that plausibly underlie a society's gender (in)equality and by adding a tightly controlled test using sibling contrasts to identify culture's impact on gender gaps in education. Finally, our study adds a cultural and gendered aspect to a large body of research on the economic outcomes of second-generation immigrants in host countries (e.g., Chiswick, 1977; Card et al., 2000; Bleakley and Chin, 2008; Aydemir et al., 2009; Algan et al., 2010; Dustmann et al., 2012).

The remainder of the paper proceeds as follows. Sections 2 through 4 describe data, stylized facts, and the empirical strategy. Sections 5 through 8 present the empirical evidence from Swedish administrative data, an in-depth analysis of the robustness of the results, and investigation of potential mechanisms, and the findings using PISA. The final section concludes. All supplementary material is in the Appendix.

3.2 Data

3.2.1 Cultural Dimensions Data

The seminal contribution in the field of culture measurement is the work of Hofstede (1980, 2001), who has developed a concise set of quantitative indices describing cultural values, beliefs, and attitudes along several dimensions. Hofstede's original measures of national culture were based on survey data from IBM employees across the world, and were later expanded using data from the Chinese Values survey and from the World Values Survey 1995-2004. Although alternative measures of culture gained recognition over the years (e.g., Schwartz, 2000; House et al., 2004), they have all been shown to be fairly consistent in their approach and closely resemble the methodology used by Hofstede (Taras et al., 2009). Thus, we apply the current version of Hofstede's cultural dimensions data (Hofstede et al., 2010).[1] Motivated by the hypotheses we sketched out in the introduction, we focus in particular on the following four dimensions of national culture.

Masculinity versus Femininity (MAS). A high MAS country score reflects that individuals in society put strong emphasis on ambition, competition and achievement. By contrast, a society that scores low on MAS is defined as being consensus-oriented and exhibiting a prefence for cooperation, modesty, and caring for the weak. Hofstede's measure of MAS was created using a factor analysis model that loads on answers to eight work goal questions administered to samples of respondents across the world. The questions were designed to tap into individuals' views of the importance of, *inter alia*, high earnings, opportunities for advancement, having challenging work to do, working in an cooperative environment, or having a good working relationship with superiors. The MAS variable ranges between 5 and 110, which we rescaled to lie between 0.05 and 1.1. Over the years, there has been some controversy surrounding the labeling of this cultural dimension, within experts in the field and Hofstede et al. (2010) themselves suggesting it should be relabeled as performance or achievement orientation.

Power Distance (PDI). Individuals in societies showing a high PDI score expect and accept that power is distributed unequally. By contrast, in countries scoring low on PDI, individuals strive to equalize the distribution of power and demand justification for inequalities of power. Based on factor analysis, the three survey items used to compose the measure of PDI tapped into individuals' acceptance of power and inequality at the workplace (i.e., frequency of disagreements with superiors, perceptions of leadership-styles, preferences for leadership styles). The PDI variable ranges between 11 and 104, which we rescaled to lie between 0.11 and 1.04.

Uncertainty Avoidance (UAI). Individuals in societies that score high on UAI are more likely to feel threatened by ambiguous or unknown situations and to show intolerance of unorthodox behavior and ideas. By contrast, in countries with a low UAI score, individuals maintain a more relaxed attitude towards situations that are novel, unknown, surprising, different from usual. The UAI measure is constructed by combining three survey items that tap into individu-

[1] The data is available at https://geerthofstede.com/research-and-vsm/dimension-data-matrix.

als' feelings of stress at work, their perceptions of the importance of rule orientation, and their openness towards new work experiences. The UAI variable ranges between 8 and 112, which we rescaled to lie between 0.08 and 1.12.

Long-Term Orientation versus Short-Term Orientation (LTO). A high LTO country score reflects that individuals in society foster virtues oriented towards future rewards, perseverance, and thrift. By contrast, societies with a low LTO score are characterized by norms towards immediate need gratification. The LTO was constructed based on a factor analysis model that loads on three survey questions presented to respondents across the world, tapping into whether thrift is considered a desirable personality trait, national pride, and importance of service to others. The LTO variable ranges between 0 and 100, which we rescaled to lie between 0 and 1.

The four dimensions of national culture are largely independent, as is evident in Figure 3.1. It shows that the pairwise cross-country correlations between each of the four cultural dimensions are virtually zero ($r \leq 0.05$) in four out of six cases and small ($r \leq 0.2$) in the remaining two. Appendix Figures A1-A4 map the distributions of the four cultural variables around the globe. For all four variables, we observe considerable variation not only across but also within supranational geographic regions. For example, while some Latin countries in Europe (e.g., France) score high on PDI, others (e.g., Spain) are characterized by a much lower PDI score. In a similar vein in Eastern European and ex-Soviet countries, some show, for example, high MAS values (e.g., Slovakia, Hungary, Poland), while others (e.g., Russia, Latvia, Slovenia) belong to the lowest part of the MAS distribution.

Hofstede's database contains cross-country measures for two additional cultural dimensions: (i) the extent to which members of society are supposed to take care of themselves as opposed to being strongly integrated and loyal to a cohesive group (labeled *individualism vs. collectivism*); and (ii) the degree to which a society allows relatively free gratification of basic and natural human drives related to enjoying life and having fun (labeled *indulgence vs. restraint*). Based on the economics literature on gender differences and gender convergence, we did not arrive at clear-cut predictions within respect to these cultural dimensions, and hence exclude them from our main analysis. However, we carry out sensitivity checks to probe whether our results are robust to the inclusions of these two dimensions of national culture.

3.2.2 Registry-Based Student Data from Sweden

The student data are based on several registers compiled by Statistics Sweden. Our population of interest consists of the universe of second-generation immigrants observed in the 9th grade graduation registers between 1988 and 2017, at the end of compulsory education at age 16. Second-generation immigrants are defined as individuals born in Sweden to two foreign-born parents. In the 1988 graduating cohort, 4.6 percent were second-generation immigrants, and with increasing immigration to Sweden, this fraction has risen and constitutes 10 percent of the graduating cohort in 2017. We merge this population to their parents and siblings through the

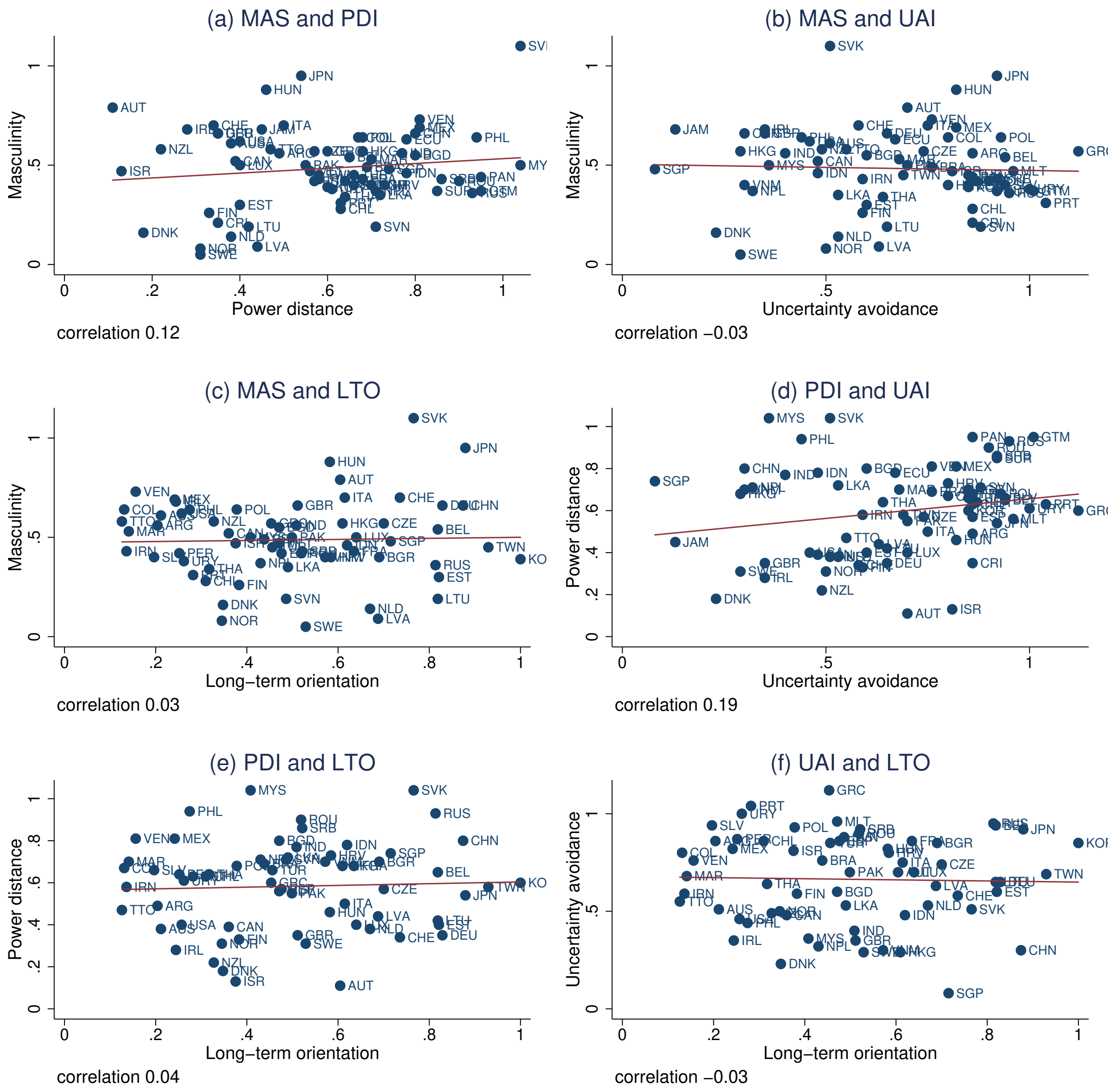


Figure 3.1: Correlations between Hofstede's Cultural Dimensions

Notes: This figure plots the pairwise cross-country correlations between four dimensions of national culture (Hofstede et al., 2010) : Masculinity versus Femininity (MAS), Power Distance (PDI), Uncertainty Avoidance (UAI), Long-Term Orientation versus Short-Term Orientation (LTO).

multi-generation register and obtain additional family background information through education and tax registers. Our main variables are detailed below.

Outcome Variables. We study student achievement using grades at the end of compulsory school. Our main outcome is the GPA, an average over grades in all subjects at the end of compulsory education (age 16). The GPA is standardized by graduation year to mean zero and standard deviation one, within the population of second-generation immigrants. Since the GPA is a teacher assessment, we complement our analysis by separately studying grades in mathematics and Swedish, subjects in which students take standardized tests. In Table 3.1, we provide summary statistics for our estimation sample. Rows 1-3 show our outcome measures used for estimation. Rows 4-6 show our outcome variables when standardized within a sample of all students. From the latter it can be seen that, relative to the whole population, second-generation immigrant children show lower educational achievements, in the order of roughly 15 percent of a standard deviation.

Table 3.1: Summary Statistics

	N	Mean	Std Dev	min	max
Outcome variables					
standardized GPA (2nd gen imm.)	78040	-0.01	1.00	-3.405508	2.60006
stand. Math grade (2nd gen imm.)	78040	-0.01	1.00	-2.027084	2.470137
stand. Swedish grade (2nd gen imm.)	77740	-0.02	0.99	-2.415584	2.227661
Student outcomes (overall)					
standardized GPA (whole pop.)	78040	-0.14	1.03	-3.426538	2.563698
stand. Math grade (whole pop.)	78040	-0.18	0.97	-2.273564	2.117368
stand. Swedish grade (whole pop.)	77740	-0.15	0.95	-2.485414	2.054244
Hofstede's cultural dimensions					
Masculinity vs. Feminity (MAS)	78040	0.43	0.12	.08	.864171
Power Distance (PDI)	78040	0.59	0.16	.18	.8982282
Uncertainty Avoidance (UAI)	78040	0.72	0.15	.23	1.021508
Long-Term Orientation (LTO)	78040	0.39	0.15	.1289433	.9369735
Covariates					
Female	78040	0.50	0.50	0	1
Graduation year	78040	2,004.29	7.75	1988	2017
Age	78040	16.00	0.28	14	19
Birth month	78040	6.47	3.41	1	12
Birthorder	78040	2.24	1.27	1	7
Individualism	78040	0.44	0.14	.19	.90425
Indulgence	78040	0.42	0.14	.1509012	.9088017
log ppp GDP p.c. (2000)	78040	9.24	0.62	7.247311	10.51245
Socioeconomic status					
Years of schooling mother	75887	10.53	2.63	7	19
Years of schooling father	72958	10.68	2.62	7	19
Income mother (in SEK)	77336	156,784.49	134,608.78	0	2287446
Income father (in SEK)	75328	209,817.36	203,960.95	0	1.55e+07

Notes: The table shows summary statistics for our estimation sample of opposite-gender second generation immigrants graduating between 1988 and 2017. Age is captured as the difference between the year of graduation and the year of birth. Parental income captures the average income of the mother/father at the age 13-16 of the child.

Family Background Variables. Our model includes family-fixed effects, which make background variables such as parents' education and earnings redundant. When exploring mechanisms we will, however, consider the role of socio-economic background and exploit variation in parents' education and earnings. We use information on parents' highest achieved education level observed in the education register when the child is aged 15, and a measure of average parental earnings when the child is aged 13–16. Additionally, we adopt an earnings-based measure of labor force participation previously used in the immigration literature for Sweden, which assigns participation to individuals with annual earnings of at least 50 percent of the median of a 45-year old worker (Erikson et al., 2007; Forslund et al., 2011). Summary statistics for some of the family background variables are shown in Table 3.1. Immigrants parents have roughly 10.5 years of education (mothers: 10.54; fathers: 10.68), which is more than a year less than the schooling of their native counterparts (non-reported). Average annual earnings of immigrant mothers amounts to roughly 157,000 SEK, while those of immigrants fathers amount to roughly 210,000 SEK. When compared to the average earnings of their native counterparts (non-reported), the earnings of immigrants are between 28 percent (mothers) and 35 percent (fathers) lower.

Source Country Cultural Variables. The data include information on parents' birth country or birth region. Specifically, for the source countries from which immigration to Sweden is rare, birth countries have been combined into birth regions to protect anonymity in the data. Table 3.2 lists the distribution of birth countries/regions of parents present in our sample. We merge Hofstede's contemporaneous cultural indicators to parents' source countries, regardless of parents' birth and immigration year, making the assumption that the cultural traits represented by these indicators are fixed over time. For source regions, we use weighted averages of the culture variables across countries belonging to each region, where weights are based on age-, graduation year, and gender-specific immigrant shares from different source countries provided by Statistics Sweden. Hofstede data are missing for some countries (see Figures A1–A4) and students with both parents originating from these countries are dropped from our analysis. This concerns primarily individuals from the horn of Africa and Iraq, which are large immigrant groups in Sweden.[2] 81.5 percent of the second-generation immigrants in our sample have parents from the same source country and as such there is no ambiguity in terms of their cultural heritage. For the remaining 18.5 percent, we define cultural origin as the average across parents. The estimation sample consists of 78,040 opposite-sex biological siblings.

[2]In the full population of second-generation immigrant students, 5 percent have parents from the horn of Africa, and 5–6 percent from Iraq, respectively. As seen in Table 3.2, these country groups are underrepresented in our data set and students are only included if their other parent is from a different country where culture is non-missing.

Table 3.2: Distribution of Birth Countries/Regions of Parents

	Fathers		Mothers	
Birth Region	Freq.	Percent	Freq.	Percent
Finland	13,513	17.32	15,816	20.27
Denmark	919	1.18	924	1.18
Norway and Iceland	720	0.92	828	1.06
Bosnia	335	0.43	353	0.45
Former Yugoslavia	10,517	13.48	10,137	12.99
Poland	1,892	2.42	2,689	3.45
Great Britain and Ireland	655	0.84	470	0.6
Germany	558	0.72	561	0.72
Southern Europe	2,431	3.12	1,970	2.52
The Baltic states	114	0.15	100	0.13
Former USSR, Rumania, Bulgaria, Albania	822	1.05	1,108	1.42
Slovakia, Check republic, Hungary	1,098	1.41	1,038	1.33
France, Benelux, Swizerland, Austria	396	0.51	344	0.44
Canada and USA	204	0.26	164	0.21
Mexico and Central America	361	0.46	347	0.44
Chile	2,531	3.24	2,504	3.21
Rest of South America	1,079	1.38	1,038	1.33
African horn (Ethiopia, Eritrea, Somalia)	145	0.19	78	0.1
North Africa and Middle east	15,240	19.53	15,472	19.83
Other Africa	358	0.46	132	0.17
Iran	3,280	4.2	2,675	3.43
Iraq	1,439	1.84	250	0.32
Turkey	13,067	16.74	12,468	15.98
China	754	0.97	906	1.16
South east Asia	2,553	3.27	2,776	3.56
Other Asia	3,017	3.87	2,858	3.66
Oceania	23	0.03	25	0.03
Unknown	19	0.02	9	0.01

Notes: The table captures the distribution of birth countries/regions of parents present in our sample of second-generation immigrant students with opposite-sex siblings. Source countries from which immigration to Sweden is rare have been combined into birth regions to protect anonymity in the data.

3.3 Stylized Facts

We begin our empirical analysis by providing descriptive evidence on the hypotheses we have formulated at the outset. Figure 3.2 presents binned scatter plots of the mean gender GPA gap versus the mean level of cultural dimension $C \in \{MAS, PDI, UAI, LTO\}$. To construct this figure, we first collapse the gender GPA gap at the level of second-generation immigrant groups. Then, we divide the horizontal axis into 40 equal-sized bins and plot the mean gender GPA gap versus the mean value of C in each bin. The binned scatter plots in Panels A through D provide representations of the correlations between the gender GPA gap and Hofstede's cultural dimensions. As an alternative to it, Appendix Figure C.5 shows the same correlations in scatter plots where the gender GPA gap is averaged by second-generation immigrant groups and cultural dimension C.

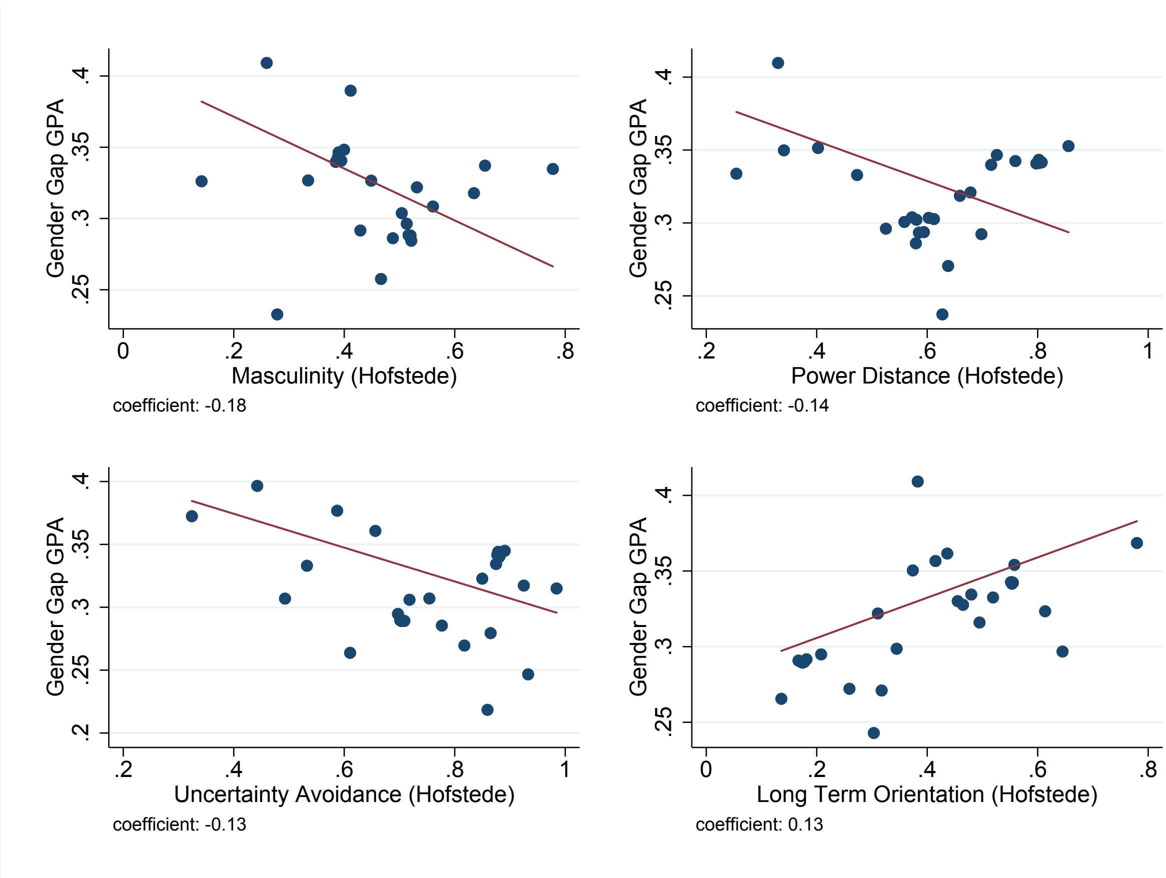


Figure 3.2: Hofstede's Cultural Dimensions and Gender GPA Gap

Notes: This figure presents binned scatter plots of the relationship between the gender gap in student achievement and cultural dimension $C \in \{MAS, PDI, UAI, LTO\}$. To construct the plots, we divide the horizontal axis into 40 equal-sized bins and plot the mean level of the gender achievement gap against the mean level of cultural dimension C within each bin. Some bins a larger than others since some groups of second-generation immigrants account for more than 2.5% of our sample. The figure is inspired by Figlio et al. (2019).

Overall, the figure provides some first descriptive evidence in support of our hypotheses. To recap, we started out by arguing that a culture of achievement orientation may cause boys to outperform girls in class, for the reason that it can create parental pressure for children to be competitive and excel in school, to which girls tend to respond more adversely than boys do (Niederle and Vesterlund, 2010). Panel A shows, indeed, that the higher the country-of-ancestry MAS score of second-generation immigrants, the lower the GPA of girls relative to boys. The correlation is 0.18, suggesting that, as we, for example, compare immigrant youth from Denmark (with a low MAS score of 0.16) with those from Germany (with a high MAS score of 0.66), the mean GPA of girls relative to boys would decrease by roughly 29%.

Our second argument was that a culture of inequality acceptance may reinforce perceptions of the glass ceiling. This, in turn, may cause parents to differentially invest in the human capital of sons and daughters, which can act as an impediment to girls catching up or overtaking boys educationally. Consistent with this argument, Panel B shows a negative relationship between PDI and the mean GPA of girls relative to boys. The correlation is -0.14, and thus roughly three-quarters of the size of that for MAS.

Our third argument was that young women may be less likely to increase career-oriented human capital investments, and thus less likely to catch up or overtake boys educationally, if they are socialized to avoid choices that involve risks (e.g., of breaking gender norms) or uncertainty (e.g., career progression). Panel C shows, indeed, that the higher Hofstede's UAI score among second-generation immigrants, the lower the GPA of girls relative to boys. The correlation is -0.13, and thus of roughly the same magnitude as that for PDI.

Finally, a powerful factor in the transformation of women's education and work was a change from static decision-making with limited horizons to dynamic decision-making with long-term horizon (Goldin, 2006). Based on this, we have formulated the hypothesis that we are more likely to observe girls catching up with or overtaking boys if parents transmit long-term oriented attitudes to their offspring. In line with this hypothesis, Panel D shows a positive relationship between LTO and the mean GPA of second-generation immigrant girls relative to boys. The correlation is 0.13, and thus in terms of magnitude similar to those for PDI and UAI.

3.4 Using Siblings to Identify the Impact of Culture on Gender Achievement Gaps

We build on and expand the epidemiological approach used to estimate correlations between source country characteristics and immigrant outcomes in the host country (Carroll et al., 1994; Giuliano, 2007; Fernández and Fogli, 2009). The epidemiological approach isolates the impact of source country cultural norms by studying immigrants who face the same social and economic institutions in the host country. This empirical strategy is however limited by the fact that social beliefs, gender equality and cultural norms are correlated with other underlying factors that affect immigrant opportunities in the host country. For example, the level of economic development and education in the source country, which likely affect immigrants' outcomes in the host country, are also likely to be correlated with cultural norms, complicating a causal interpretation.

We are interested in how gender achievement gaps are affected by cultural traits passed on from immigrant parents to their offspring. By introducing a family-fixed effect and comparing sisters and brothers from the same family, we net out all unobserved family-level characteristics that affect sisters and brothers in the same way, while we are still able to identify the impact of culture on the gender gap in student achievement. By studying immigrant gender gaps instead of levels of immigrant outcomes, we can identify the impact of culture without conflating the estimate with the potentially underlying correlation between culture and other determinants of outcomes.

Consider the following regression model:[3]

$$y_{ift} = \beta_0 + \beta_1 Female_i + \beta_2 (Female_i \times Culture_f) + \beta_3 \mathbf{X}_i' + \beta_4 (Female_i \times \mathbf{X}_i') + \delta_t + \theta_f + \varepsilon_{ift} \quad (3.1)$$

where i denotes individual i from family f, graduating in year t. $Female_i$ is a dummy that takes the value one if the individual is a girl; $Culture_f$ is the cultural index and δ_t represents graduation cohort dummies. The vector $\mathbf{X}_i$ controls for a set of individual attributes, which in our basic specification only includes a student's age. The family-fixed effect θ_f nets out all (observed and unobserved) source-country specific effects on achievement levels, but still allows us to estimate the interaction between source country characteristics and the female dummy. β_2 is the coefficient of interest that informs us how the achievement gap varies with cultural background. By including the family-fixed effect, we essentially compare how achievement gaps between brothers and sisters who grow up in the same family and most often attend the same school, are related to the cultural norms and beliefs in their parents' birth countries.

In terms of identification, the remaining concern is that unobserved traits correlated with culture affect girls and boys differently, and therefore prevent us from interpreting β_2 as the impact of culture on the gender gap. Such 'confounders' could however also be considered as 'mediators' or mechanisms, depending on the causal pathways underlying the correlations.[4] We think of confounders/mediators both as originating in the economic and social conditions in the source country and as the result of, e.g., sorting of immigrants within the host country. For example, culture may be related to immigrants' education level, economic opportunities and selection into different neighborhoods within host countries, and to the school quality of their children. Previous literature has shown that girls and boys respond differently to family and

[3] See Finseraas and Kotsadam (2017) for a similar application of the sibling-fixed effects specification, linking source country female labor force participation to gender differences in labor supply among second-generation immigrants in Norway. See also Ericsson (2020) for an application similar to ours, but which proxies cultural norms using female-to-male labor force participation in the source country.

[4] If cultural norms causally affect the socioeconomic position of immigrants or parents' school choice (e.g., through aspirations and work ethics), such variables should be seen as mediators through which cultural norms are transmitted. Instead, if causality goes the other way, or correlations arise due to other reasons, we should think of these variables as confounders.

neighborhood disadvantage and school quality (Autor et al., 2016, 2019; Deming et al., 2014; Hastings et al., 2006) and the culture-gender interaction could pick up such effects.

Therefore, β_2 should be interpreted as the impact of culture, and factors correlated with culture, on the gender achievement gap. In comparison with the traditional epidemiological approach (see, e.g., Fernández and Fogli, 2006; Fernández and Fogli, 2009), we believe that this is a significant advancement, since the estimate will be biased only to the extent that unobserved factors have gendered implications. We test the sensitivity of our results to such bias by expanding our model and including interactions between the female dummy and additional source country characteristics such as GDP per capita. Importantly, our rich data allow us to include interactions between host country neighborhood and gender, which control for all time-constant neighborhood characteristics that have differential impacts on girls and boys.[5] This sensitivity test therefore directly controls for the gendered impact of sorting to neighborhoods with different degrees of disadvantage.[6] Our contribution compared to the previous literature therefore lies both in the focus on the inner layers of cultural beliefs, and in the empirical specification which allows us to compare siblings and at the same time control for the gendered impact of unobserved confounders at the neighborhood level. Furthermore, in Section 7 we explicitly discuss mechanisms that could explain our findings.

3.5 Main Results

We present our baseline results in Table 3.3. The "raw" within-family gender achievement gap, i.e., the coefficient of the female dummy in a specification including family-fixed effects, graduation-year-fixed effects and age as controls, amounts to a girl advantage of the magnitude 0.31 of a standard deviation.

In columns 1–4, we start out with regressions of second-generation immigrants' GPA on each cultural domain $C \in \{MAS, PDI, UAI, LTO\}$ separately, interacted with the female dummy (Equation 1). All columns show that the cultural beliefs, norms and attitudes in parents' source countries have implications for the gender achievement gap among second generation immigrants in Sweden today, even when comparing opposite-sex siblings. The signs of the interactions are all in line with our hypotheses, generated by research on gender gaps in the economics literature. First, column 1 shows that a society's emphasis on ambition, competition and achievement (MAS) is associated with a smaller girl advantage. Moving up one standard deviation in the distribution of achievement orientation is associated with a closing of the gender gap by 9 percent.[7] Comparing two of the most common source countries of refugee immigrants

[5]We use neighborhoods defined by Statistics Sweden's SAMS (small areas for market statistics) units. A SAMS area is a geographical neighborhood, developed to correspond to "real" physical neighborhoods. On average, a SAMS unit has 1000 inhabitants, and there are around 9,200 units in total.

[6]Nollenberger et al. (2016) study the gender math gap and how it is related to the gender gap index developed by the World Economic Forum, an example of a variable that captures the outer layers, or manifestations, of cultural beliefs. In addition, Nollenberger et al. (2016) have a more limited set of controls to account for the fact that other factors correlated with culture and gender gaps could drive the results.

[7](0.119*0.2411)/0.313.

to Sweden, Chile (MAS index 0.28) and Turkey (MAS index 0.45), we predict that the girl-favoring gender gap among children of Turkish immigrants should be 4 percent of a standard deviation smaller than that of children to Chilean immigrants. As an alternative comparison using non-refugee source countries, we can compare the performance among children originating in the neighboring Nordic countries, with weak norms regarding performance and ambition (e.g., Denmark with a MAS index 0.16), to Germany with a much stronger achievement culture (MAS index 0.66): the estimate predicts that the gender gap among children of German immigrants should be 12 percent of a standard deviation smaller than among second generation Danes, or put differently, would reduce the "raw" achievement gap by 38 percent. As such, cultural norms that emphasize ambition and competition seem to be associated with sisters performing worse relative to their brothers, which is in line with the hypothesis generated by insights from the behavioral economics literature on women's and men's performance in competitive situations.

Table 3.3: Gender GPA Gap and Cultural Dimensions, Baseline Results

	(1)	(2)	(3)	(4)	(5)
MAS * Female	-0.2411***				-0.1840***
	(0.0455)				(0.0503)
PDI * Female		-0.1195***			-0.0523
		(0.0346)			(0.0463)
UAI * Female			-0.1097***		-0.0560
			(0.0355)		(0.0427)
LTO * Female				0.0871**	0.1123***
				(0.0360)	(0.0376)
Observations	78040	78040	78040	78040	78040
R-squared	.674	.674	.674	.674	.674
Dependent var. (mean)	-.007	-.007	-.007	-.007	-.007
Dependent var. (sd)	.996	.996	.996	.996	.996
Cultural var. (mean)	.427	.587	.723	.387	
Cultural var. (sd)	.119	.158	.149	.148	
Cultural var. * Fem. (beta)	-.029	-.019	-.016	.013	
Number of clusters	30018	30018	30018	30018	30018
Gender Gap	.313	.313	.313	.313	.313
Family FE	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. The dependent variable is normalized to be mean 0 and standard deviation 1 relative to the universe of all second-generation immigrant students. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Second, consider column 2 which shows how the acceptance of inequality (PDI) is related to the gender achievement gap. We expect acceptance of unequal power distribution to increase boys' performance relative to girls, since parents who carry such beliefs might invest more in boys relative to girls as they perceive women's access to high positions in society as either undesirable or unattainable. Higher acceptance of power inequality is as expected associated with a reduction of the gender gap: a one standard deviation increase in the index is associated with a lowering of the gender gap by 6 percent.

Third, column 3 presents how norms that emphasize uncertainty avoidance (UAI) influence the gender gap. We hypothesize that tolerance for uncertainty and ambiguity makes it easier for women to step out of their traditional gender role, in order to put more emphasis on career and to try out male-dominated professions. Our findings are in line with this idea: higher uncertainty avoidance is associated with girls performing worse relative to boys. A one standard deviation shift in the uncertainty avoidance index corresponds to a 5.2 percent change of the gender achievement gap.

Column 4 focuses on the extent to which members of societies are willing to delay short-run rewards for long-term goals and returns (LTO). Long-term orientation can in this context increase girls' effort and parents' investments in their daughters, since it involves a perception of women having a life-long commitment to the labor force rather than a static short-term view that focuses on family formation. We find that long-term orientation is associated with sisters performing better relative to their brothers, which confirms our hypothesis.

Finally, column 5 shows the results from our preferred specification. It includes all four cultural dimensions simultaneously in order to test the robustness of the results to potential correlations between the domains and to shed light on their relative importance. As Figure 1 generally shows low correlations between the indices, we expect the results to be relatively insensitive to this test. We find that two out of four indicators remain highly significant: achievement orientation and long-term orientation have clear implications for the relative achievement of girls and boys. The coefficients on acceptance of power distance and uncertainty avoidance are roughly halved (and become insignificant), which is not surprising since these indices are the ones most highly correlated between each other. As for the relative importance of the four cultural dimensions, the strongest predictor of gender gaps in student achievement is a society's emphasis on ambition, competition and achievement (MAS). Long-term orientation (LTO) also has a statistically significant effect, but its magnitude is only about two-thirds that of achievement orientation.

To sum up, we have found compelling and intriguing evidence that differences in beliefs and attitudes across cultures have gendered consequences for the academic outcomes of second-generation immigrants, holding constant the host country and its institutions. The previous epidemiological literature has primarily focused on source country indicators that reflect actual behavior, such as female labor force participation and fertility, and that are one-dimensional manifestations of potentially many different cultural beliefs and attitudes. Our contribution is the first to demonstrate that multi-dimensional measures of norms and attitudes have implications for gender gaps in the host country, in ways that can be predicted by findings in the earlier economics literature.

3.6 Robustness

We next, in Table 3.4, provide a number of robustness tests to check the sensitivity of our baseline results to potential confounding factors. First, column 1 simply repeats our preferred specification from Table 3.3. Column 2 includes birth month and birth order controls interacted with gender. Column 3 includes controls for Hofstede's additional cultural domains, indulgence and individualism, interacted with gender. Column 4 controls for log GDP per capita in the source county, again interacted with gender to pick up source countries characteristics that may affect girls and boys differentially.

Table 3.4: Gender GPA Gap and Cultural Dimensions, Sensitivity Checks

	(1)	(2)	(3)	(4)	(5)	(6)
MAS * Female	-0.1840***	-0.1802***	-0.2262***	-0.1869***	-0.1865***	-0.2090**
	(0.0503)	(0.0502)	(0.0627)	(0.0521)	(0.0693)	(0.0816)
PDI * Female	-0.0523	-0.0538	-0.1385	-0.0700	-0.1495	0.0026
	(0.0463)	(0.0463)	(0.1338)	(0.1075)	(0.1480)	(0.1742)
UAI * Female	-0.0560	-0.0617	-0.0304	-0.0479	-0.0153	-0.0320
	(0.0427)	(0.0425)	(0.0476)	(0.0615)	(0.0634)	(0.0746)
LTO * Female	0.1123***	0.1094***	0.1334***	0.1159***	0.1060**	0.0943*
	(0.0376)	(0.0376)	(0.0398)	(0.0416)	(0.0430)	(0.0504)
Observations	78040	78040	78040	78040	77702	73448
R-squared	.674	.676	.674	.674	.681	.715
Dependent var. (mean)	-.007	-.007	-.007	-.007	-.007	-.007
Dependent var. (sd)	.996	.996	.996	.996	.996	.996
Number of clusters	30018	30018	30018	30018	29898	28201
Gender Gap	.313	.313	.313	.313	.314	.313
Family FE	✓	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓	✓
Birth variables		✓			✓	✓
Birth var. * Fem		✓			✓	✓
Individualism * Fem.			✓		✓	✓
Indulgence * Fem.			✓		✓	✓
LogGDPpc2000 * Fem.				✓	✓	✓
Municipality FE					✓	
Mun. FE * Fem.					✓	
Neighborhood FE						✓
Neighb. FE * Fem.						✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. The dependent variable is normalized to be mean 0 and standard deviation 1 relative to the universe of all second-generation immigrant students. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Birth variables include dummies for the month of birth and birth order. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Finally, columns 5 and 6 include all these controls simultaneously, while additionally phasing in municipality and neighborhood fixed effects, and their interactions with gender, respectively. The baseline estimates are virtually unchanged in these alternative specifications. We believe that this is a very powerful test of the potential impact of confounders, since we can control for all unobserved neighborhood-specific factors that have differential impacts on girls and boys. In other words, immigrant selection to neighborhoods and the different environments their children are exposed to, are not driving our results.

In addition, Table 3.5 and Table 3.6 present results separately for mathematics and Swedish, subjects in which students take standardized tests. We see that our main conclusions are confirmed, and that in these cases the impact of all four different cultural dimensions is even more visible in the 'horse-race' specification in the last column. Appendix Tables C.1 and C.2 present the corresponding sensitivity tests which show that including additional controls reduces the importance of power distance and uncertainty avoidance, but that the impacts of achievement orientation and long-term orientation are robust both in terms of magnitude and precision.

Table 3.5: Gender Math Gap and Cultural Dimensions

	(1)	(2)	(3)	(4)	(5)
MAS * Female	-0.2340***				-0.1775***
	(0.0490)				(0.0560)
PDI * Female		-0.1063***			-0.0072
		(0.0366)			(0.0521)
UAI * Female			-0.1452***		-0.1275**
			(0.0392)		(0.0504)
LTO * Female				0.1101***	0.1374***
				(0.0387)	(0.0414)
Observations	78040	78040	78040	78040	78040
R-squared	.636	.636	.636	.636	.636
Dependent var. (mean)	-.008	-.008	-.008	-.008	-.008
Dependent var. (sd)	.997	.997	.997	.997	.997
Cultural var. (mean)	.427	.587	.723	.387	
Cultural var. (sd)	.119	.158	.149	.148	
Cultural var. * Fem. (beta)	-.028	-.017	-.022	.016	
Number of clusters	30018	30018	30018	30018	30018
Gender Gap	.015	.015	.015	.015	.015
Family FE	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. The dependent variable is normalized to be mean 0 and standard deviation 1 relative to the universe of all second-generation immigrant students. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table 3.6: Gender Gap in Swedish and Cultural Dimensions

	(1)	(2)	(3)	(4)	(5)
MAS * Female	-0.4156***				-0.1833***
	(0.0501)				(0.0569)
PDI * Female		-0.3359***			-0.2984***
		(0.0378)			(0.0528)
UAI * Female			-0.2655***		-0.0970*
			(0.0399)		(0.0501)
LTO * Female				0.1998***	0.3091***
				(0.0397)	(0.0421)
Observations	77601	77601	77601	77601	77601
R-squared	.615	.615	.614	.614	.615
Dependent var. (mean)	-.015	-.015	-.015	-.015	-.015
Dependent var. (sd)	.993	.993	.993	.993	.993
Cultural var. (mean)	.427	.588	.723	.387	
Cultural var. (sd)	.118	.158	.149	.148	
Cultural var. * Fem. (beta)	-.049	-.053	-.04	.03	
Number of clusters	29865	29865	29865	29865	29865
Gender Gap	.471	.471	.471	.471	.471
Family FE	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. The dependent variable is normalized to be mean 0 and standard deviation 1 relative to the universe of all second-generation immigrant students. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

3.7 Mechanisms and Potential Confounders

In this section, we open up the discussion and consider that confounders, such as parental education and socioeconomic status, or choice of neighborhood and school, could also be considered as mechanisms or mediators if culture per se is causally linked to such outcomes. For example, if parents from achievement-oriented cultures are more successful in the labor market because of higher effort, or send their children to higher quality schools because of their high valuation of performance, parents' SES and children's school environment should not necessarily be seen as confounders. The purpose of this section is therefore to highlight potential pathways through which culture could be mediated, while acknowledging that we are not able to identify them in a causal sense. As will become apparent, the pathways we explore are not necessarily mutually exclusive, nor can we rule out the possibility that alternative mechanisms are at work. Our results are summarized in Tables 3.7 and 3.8. In discussing the results, we focus mainly on factors that may explain why a culture of achievement orientation leads to an educational advantage of boys relative to girls, as this is the cultural dimension that matters most strongly for gender gaps in student achievement. Where appropriate, we also discuss how the mechanisms play out for the other cultural dimensions.

Intentional Differential Treatment of Sons vs. Daughters through Parents. A natural starting point for thinking about plausible mechanisms driving our main results is to ask whether parents with different cultural backgrounds treat girls and boys differently. Gendered treatment can involve passing on different aspirations, ambitions, values and gender roles in the education and upbringing of children. Parents can also invest differentially in boys' and girls' skill formation either through differential time investments or through choosing schools of different qualities. While mechanisms reflecting values and time investments are hard to observe in most data sets, with our data we can partly address whether sons and daughters are treated differently by observing whether parents gender-discriminate when choosing schools for their offspring.

Table 3.7: Gender Gap in Quality and Type of School Attendance, Baseline Results

	(1) Res. school quality	(2) Private school
MAS * Female	-0.4015	0.0164
	(0.3787)	(0.0159)
PDI * Female	-0.2352	-0.0016
	(0.3389)	(0.0132)
UAI * Female	0.0769	-0.0210
	(0.3418)	(0.0136)
LTO * Female	-0.3440	-0.0203
	(0.3043)	(0.0130)
Observations	57692	57348
R-squared	.614	.757
Dependent var. (mean)	.205	.096
Dependent var. (sd)	6.194	.295
Number of clusters	22817	22697
Gender Gap	.103	.005
Family FE	✓	✓
Grad. year FE	✓	✓
Age	✓	✓
Age * Female	✓	✓
Nonmover sample	✓	✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. *Nonmover sample* additionally restricts the sample to families who lived in the same neighborhood at graduation of all their children. Residual school quality measures the average peer achievement by school and graduation year, after netting out variation across schools that is explained by children's gender, age, and birth country as well as mothers' and fathers' education, earnings, birth country and immigration age. Private school is a binary variable which indicates whether the student attends a private school in the year of graduation. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

In column 1 of Table 3.7, we explore this possibility by investigating whether parents from different cultures systematically place sons in higher quality schools than daughters. To that end, we construct a measure of school quality, which represents the average peer achievement by school and graduation year, after netting out variation across schools that is explained by students' family background.[8] The specification is analogous to our main analysis: we identify whether there is a within-family difference in the school quality at schools attended by sisters and brothers by regressing school quality on culture interacted with gender in a family-fixed effects model. We restrict the sample to non-moving families to avoid picking up differences in school quality that are due to moving to a new neighborhood, rather than due to choice of school. We find that none of the cultural indicators predicts a within-family difference in school quality between brothers and sisters. As school quality can be difficult for parents to observe and act on, we additionally examine differential treatment in the probability to send children to private schools.[9] If private schools are perceived as more selective and of higher quality, differential treatment could manifest itself through this type of school choice. In column 2 of Table 3.7, we again find no evidence that parents' cultural background is related to differences in the educational investments of sons and daughters. Table 6 thus gives us no reason to believe that parents from different cultural origins intentionally treat their sons and daughters differently when investing in their skills.[10]

Non-Intentional Mechanisms. Next, in Table 3.8, we investigate a set of mechanisms that no longer build on the idea that parents from achievement-oriented cultures intentionally treat sons and daughters differently. First, we conjecture that, irrespective of their children's gender, parents from achievement-oriented cultures might place their children in higher quality schools, and that boys might benefit more from this than girls do (see, e.g., Autor et al., 2016). Thus, we first regress the quality of schools attended by immigrant children on the cultural variables to explore whether there is a correlation (Panel A, column 1). In a similar fashion, we also regress our binary dependent variable for a child attending private school on the cultural variables (Panel A, column 2). Panel A shows that children from achievement-oriented cultures (MAS) attend higher quality schools, and they are also more likely to go to private schools. We also find that the other cultural indicators are related to school characteristics. As an example, long-term orientation (LTO) is positively associated with school quality, and acceptance of power (PDI)

[8]We construct the measure of school quality by regressing percentile ranked GPA (by graduation year) on children's gender, age, and birth country, and mothers' and fathers' education, earnings, birth country and immigration age, in the full population of graduating students. We use the residuals from this regression and create leave-out means at the school-graduation cohort level, leaving out the index individual from the average. This is our measure of school quality, which informs how well the school performs relative to other schools after taking into account student background. Unfortunately, we do not have data on prior test scores to construct a value-added quality measure.

[9]Private or "independent" schools were uncommon among the early cohorts in our sample, but after a reform in the 1990s, the share of students attending private schools has risen. Private schools are tuition-free but operated by independent foundations, small companies or large for-profit school corporations.

[10]Tables A3 and A4 in the Appendix show that these findings generally are robust to including a wider set of controls. However, when including controls for individualism and indulgence, we find some significant interactions between gender and culture in the choice of private schooling. However, the interaction between MAS and female is positive, which would predict a larger girl-favoring gender gap and as such cannot explain our baseline result.

and uncertainty avoidance (UAI) show negative correlations. Based on these correlations and the previous literature on gender gaps in education, we hypothesize that school quality may be a possible explanation to the link between achievement orientation and gender gaps in education. The positive correlation between LTO and school quality does however not yield a prediction consistent with our baseline findings in Table 3.3, since a higher LTO is to the benefit of girls' school performance.

In the next step, we regress student GPA on school quality interacted with gender in a family-fixed effects model (Panel B, column 1). Similarly, we regress student GPA on the private school dummy interacted with gender in a family-fixed effects specification (Panel B, column 2). These specifications adopt the identification strategy previously used by Autor et al. (2016) and essentially identify whether school quality has differential impacts on girls and boys by comparing sisters and brothers. Our results confirm those in Autor et al. (2016): column 1 shows that girls benefit less relative to boys from higher school quality. Similarly, in column 2 we observe that the gender gap is smaller in private compared to public schools. These results therefore support the explanation that differences in school quality and school characteristics across children with different cultural origins unintentionally could affect gender gaps in education, as girls and boys are differentially affected by school quality.

An alternative, and partly overlapping explanation, is that parents from achievement-oriented cultures are positively selected in terms of SES compared to immigrants from other cultures, and that this disproportionately promotes the educational outcomes of boys (see, e.g., Autor et al., 2019; Figlio et al., 2019). The correlation between culture and SES can be considered a mechanism if culture *per se* is causing differences in socio-economic status across immigrants from different source countries. This could be the case either if differences in SES originate from selective migration, or if cultural origin affects the integration and socioeconomic position of migrants in the host country. For example, achievement orientation might induce the well-educated to emigrate, e.g., to secure well-paying jobs or the best possible educational opportunities for their offspring. Similarly, even without selective migration, achievement orientation might induce immigrants to work harder to integrate in the host country and consequently reach higher socio-economic positions.

Column 3 (Panel A) shows that parental SES (measured with an index incorporating both parental education and earnings[11]) is correlated with the cultural variables in a similar way as school quality: parents from achievement-oriented cultures appear to have higher SES in terms of education and earnings. In Panel B, similar to the results in Autor et al. (2016), we show that the female GPA advantage is reduced with higher SES. Socio-economic background and school quality/private school are positively correlated and likely to pick up similar mechanisms—that in comparison to girls, boys' relative behavioral and academic outcomes are particularly sensitive to disadvantage, both in terms of school and family environment. The magnitudes

[11]The parental SES index is based on a regression of GPA on parents' education and earnings, while controlling for age, gender, and graduation year dummies. We use the prediction—i.e., the "expected GPA"—as an index of students' SES.

Table 3.8: Mechanisms, Baseline Results

	(1) Res. school quality	(2) Private school	(3) Predicted GPA	(4) Par. time in Sweden	(5) Traditional LFP
MAS	2.3661***	0.1271***	0.0960***	-2.8025***	0.0226
	(0.3084)	(0.0166)	(0.0246)	(0.2959)	(0.0340)
PDI	-2.0610***	-0.1213***	-0.3394***	-12.2919***	0.0347
	(0.2947)	(0.0151)	(0.0215)	(0.2751)	(0.0334)
UAI	-0.6001**	0.0020	-0.0851***	3.3107***	-0.0556*
	(0.2880)	(0.0146)	(0.0202)	(0.1969)	(0.0314)
LTO	1.5369***	0.0235	0.3248***	8.7004***	-0.0350
	(0.2585)	(0.0143)	(0.0192)	(0.2098)	(0.0268)
Observations	57697	57472	78040	66454	78040
R-squared	.012	.06	.313	.165	.001
Dependent var. (mean)	.205	.096	-.231	5.618	.058
Dependent var. (sd)	6.195	.295	.468	4.671	.616
Number of clusters	22822	22821	30018	25530	30018
Gender Gap	.081	.004	.352	-.052	.004
	(1) Std. GPA	(2) Std. GPA	(3) Std. GPA	(4) Std. GPA	(5) Std. GPA
Res. sch. qual. * Female	-0.0019*				
	(0.0011)				
Res. school quality	0.0202***				
	(0.0010)				
Priv. school * Female		-0.0798***			
		(0.0216)			
Private school		0.2129***			
		(0.0244)			
Pred. GPA * Female			-0.0236*		
			(0.0134)		
Par. time Swe. * Female				0.0007	
				(0.0012)	
Trad. LFP * Female					-0.0113
					(0.0090)
Observations	57692	57348	78040	66454	78040
R-squared	.687	.682	.674	.672	.674
Dependent var. (mean)	.038	.038	-.007	.023	-.007
Dependent var. (sd)	.982	.983	.996	.986	.996
Mechanism (mean)	.205	.096	-.231	5.618	.058
Mechanism (sd)	6.194	.295	.468	4.671	.616
Mechanism * Fem. (beta)	-.012	-.024	-.011	.003	-.007
Number of clusters	22817	22697	30018	25530	30018
Gender Gap	.312	.314	.313	.309	.313
Family FE	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓
Nonmover sample	✓	✓			

Notes: Panel A reports estimates of regressing each mechanism on the cultural dimensions (w/ control variables but w/o family FE). Panel B reports estimates of regressing normalized GPA on the mechanisms (w/ control variables and family FE). The sample is restricted to second-generation immigrant students with opposite-sex siblings. *Nonmover sample* additionally restricts the sample to families who lived in the same neighborhood at graduation of all their children. Residual school quality measures the average peer achievement by school and graduation year, after netting out variation across schools that is explained by children's gender, age, and birth country as well as mothers' and fathers' education, earnings, birth country and immigration age. Private school is a binary variable which indicates whether the student attends a private school in the year of graduation. Predicted GPA is obtained by regressing GPA on parents' education and earnings, age, a female indicator and graduation year dummies. Parental time in Sweden captures the host country experience prior to the birth of the oldest sibling. Traditional LFP takes 1 if only the father is working, 0 if both/none work, and -1 if only the mother is working. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

of the estimates in Table 3.8 are however very small, implying that they are far from fully explaining the culture-gender interactions in Table 3.3.[12]

Finally, in columns 4 and 5, we explore two additional mechanisms-—host country experience in terms of parental time in the country, and traditional gender roles in terms of parents' labor force participation. We measure parental time in the country and labor force participation in the year the oldest sibling turns 15. Recent evidence suggests that boys benefit more than girls from integration interventions targeting immigrants (Dahl et al., 2020), leading us to hypothesize that it may also be that boys disproportionately benefit from parents' host country experience. Panel A shows correlations between parents' time in the country and the cultural indices. We observe that high MAS immigrant groups have shorter time in the country, and high LTO immigrants have longer time in the country. However, as shown in Panel B, the gender GPA gap is unaffected by how long parents have lived in Sweden, and we can rule out gendered integration processes as a likely explanation.

In column 5, we explore parental role models as a potential mechanism. The outcome variable captures traditional gender roles in terms of parental labor force participation, and takes the value 1 if only the father is working, 0 if both parents are working or none of them are working, and -1 if only the mother is working. The motivation for this analysis is that besides directly influencing values and investments in children, cultural origins – if they have gendered consequences – are likely to also manifest themselves in the division of labor among parents, which in turn can affect girls' and boys' perceptions about their future and their effort in school. Column 5 (Panel A) however shows limited evidence that the cultural indicators are correlated with traditional labor division among immigrant parents in Sweden. In Panel B we find no evidence that girls' relative advantage to boys is smaller in more traditional families.

To sum up, the results in Table 3.7 show that although achievement-oriented cultural origin substantially reduces the girl GPA advantage among second-generation immigrants in Sweden, there is little evidence that points in favor of intentional differential treatment of sisters and brothers in Sweden.

Alternative explanations investigated in Table 3.8 show that the indices reflecting cultural origins are correlated with children's disadvantage, both in terms of school quality and family SES.[13] Disadvantage in turn disproportionally affects boys, with gender gaps, even within the same family, that are larger in low SES environments. As such, the relative advantage of boys with achievement-oriented origins could be explained by lower disadvantage. However, our estimates suggest that this mechanism can by no means fully explain the culture/gender-gap

[12] The correlations in Panel A and the gender interactions in Panel B would imply that school quality and predicted GPA can explain less than 1 percent of the MAS*gender gap in Table 3.3. The private schooling mechanism can explain somewhat more—up to 6 percent. As will become apparent in the analysis that follows (see Appendix Table C.7, discussed below), jointly the mechanisms explored here explain roughly 25 percent of the MAS*gender gap in Table 3.3.

[13] Appendix Tables C.5 and C.6 show that the results in Table 3.8 are largely robust to alternative specifications. One exception is the positive correlation between predicted GPA and MAS in Panel A of Table 3.8, which disappears (respectively, turns negative) once we include municipality fixed effects (respectively, neighborhood fixed effects).

interaction. Returning to our theoretical predictions based on findings in the previous literature, we should maybe not be surprised that our results are not fully explained by the mechanisms discussed here. When it comes to achievement orientation, we should perhaps instead seek explanations in that girls and boys react differently to competitive pressure, and not only in the gendered implications of growing up in disadvantage.

Above, we have emphasized that if culture is the driving force behind disadvantage, we should see this as a mechanism rather than as a confounder. It is, however, also possible that culture and SES are correlated without a causal link, and in that case the impact of achievement orientation on gender gaps should be attributed to disadvantage *per se*, not to culture. In order to understand whether cultural origin, in particular the MAS index, survives as an independent explanation, we return to our baseline specification and include controls for the gender interactions with school quality, private school and parental SES. The results in Appendix Table C.7 show that the gender interaction with achievement orientation (MAS) decreases in magnitude by roughly 25 percent but remains statistically significant if we simultaneously control for the mechanisms explored in this section. Thus, the impact of achievement orientation on gender gaps in education exists beyond the potentially confounding role of disadvantage.[14] The results also show that the gender interaction with long-term orientation (LTO) is virtually unaffected by controlling for the mechanisms explored here.

3.8 Findings Based on Data from PISA

To explore whether the associations between culture and gender achievements gaps observed among second-generation youth in Sweden also exist in different populations and circumstances, we use data from the OECD's Programme for International Student Assessment (PISA). The idea behind PISA is to test the knowledge and skills of students through a metric that is internationally agreed upon, and to link test scores with data from students, parents, teachers, schools and systems to understand performance differences.

Drawing on the PISA studies from 2003, 2006, 2009, 2012, and 2015, we obtain standardized test scores in mathematics, science, and reading.[15] Following the previous literature (Fernández and Fogli, 2009; Rodríguez-Planas and Nollenberger, 2018), we drop second-generation immigrants whose countries of ancestry have fewer than 15 observations in a given host country.[16] Our sample contains 35,512 second-generation immigrant students residing in 29 host countries. We combine 41 mother source countries and 40 father source countries to 74 source country groups.[17]

[14] In Appendix Tables C.8 and C.9, we show that this conclusion also holds when phasing in additional control variables and municipality and neighborhood fixed effects, respectively.

[15] We build on resources provided by Figlio et al. (2019) in the preparation of the data set.

[16] As with the Swedish data, we assign only other parent's culture and source country in the case of missing cultural dimensions.

[17] We distinguish between mother's and father's ancestry when combining them. For example, students with a mother from Italy and a father from Spain are assigned a different ancestry than students with a mother from Spain and a father from Italy.

Importantly, this dataset contains the country of origin of the mother and the father of each second-generation immigrant student. Based on this, we assign two values of cultural trait $C \in \{MAS, PDI, UAI, LTO\}$ to each student. The first, which we think of as our "broad" measure of culture, accounts for the possibility that a child's parents originate from different cultures, and is the average value of cultural trait C of mother and father, defined at the level of their respective birth countries. For the second, which we think of as our "narrow" measure of culture, we define a student's cultural background based on the value of C in the mother's country of origin.

The main dependent variable used in the analysis is a student's PISA grade-point average, computed as the average normalized test scores of mathemathics, science, and reading. We carry out sensitivity checks that use three standardized subject scores as dependent variables.

Although the PISA dataset does not allow for empirical specifications that rely on within-family, cross-gender sibling comparisons, we fit a reasonably tightly specified model to the data:

$$\begin{aligned} y_{ihgtf} =& \beta_0 + \beta_1 Female_i + \beta_2(Female_i \times Culture_f) + \beta_3 \mathbf{X}'_i + \beta_4(Female_i \times \mathbf{X}'_i) \\ &+ \gamma_f + \gamma_h + \gamma_g + \gamma_t + \beta_5(Female_i \times \gamma_h) + \varepsilon_{ihgtf} \end{aligned} \tag{3.2}$$

where index i denotes a second-generation immigrant student, h her country of residence, g the grade she attends, t the year she partakes in PISA, and f her mother's and father's combined ancestry.[18] $Female_i$ is an indicator for whether a student is a girl, and $Culture_f$ measures a cultural dimension (or a set of cultural dimensions) based on Hofstede's data for the individual's country of ancestry. To coefficient of interest is β_2, which identifies culture's differential impact on girls relative to boys. The vector $\mathbf{X}_i$ controls for a set of individual attributes, namely a student's age in our basic specification and parental characteristics including age and education in extended specifications. We include ancestry fixed effects (γ_f) in all regressions to net out the effects of unmeasured country-of-ancestry factors which are common to girls and boys. In extended specifications, we probe whether our results are robust to allowing potential confounding characteristics of the country of ancestry to affect girls and boys differentially. Finally, we control for unmeasured confounders common to girls and boys partaking in PISA in a given year (through a set of year dummies, γ_t), attending a given grade (through a set of grade dummies, γ_g), and living in a given host country (through a set of host country dummies, γ_h). The interaction between $Female$ and host-country dummies (γ_h) accounts for differential gender achievement gaps that may arise from economic, cultural and institutional differences across host countries.

As shown in Table 3.9, the replication of our findings for Sweden with a very different sample of second-generation immigrants drawn from PISA yields results that are qualitatively and quantitatively remarkably similar. Specifically, the most important and robustly significant effect of culture on gender achievement gaps turns out to be again the extent to which a society emphasizes ambition, competition, and achievement, measured by Hofstede's MAS dimension.

[18] We combine mothers' and fathers' countries of ancestry to 74 groups.

Table 3.9: Gender GPA Gap and Cultural Dimensions, PISA Data

Dependent Variable:	*Standardized PISA Grade-Point Average*				
	(1)	(2)	(3)	(4)	(5)
MAS * Female	-0.2048**				-0.2921**
	(0.0895)				(0.1267)
PDI * Female		-0.1091			-0.1717*
		(0.0913)			(0.0893)
UAI * Female			0.0326		-0.0819
			(0.0758)		(0.1033)
LTO * Female				-0.0842	-0.0046
				(0.0671)	(0.0934)
Observations	35512	35512	35512	35347	35347
R-squared	.398	.398	.398	.399	.399
Dependent var. (mean)	0	0	0	.001	.001
Dependent var. (sd)	.964	.964	.964	.965	.965
Cultural var. (mean)	.563	.700	.554	.664	
Cultural var. (sd)	.137	.154	.292	.23	
Cultural var. * Fem. (beta)	-.029	-.017	.01	-.02	
Number of Clusters	74	74	74	73	73
Gender Gap	-.011	-.011	-.011	-.011	-.011
Year FE	✓	✓	✓	✓	✓
Grade FE	✓	✓	✓	✓	✓
Anc. Country FE	✓	✓	✓	✓	✓
Host Country FE	✓	✓	✓	✓	✓
Host Country FE * Fem.	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Age * Fem.	✓	✓	✓	✓	✓

Notes: The table reports estimates of equation (3.2) on a sample of second-generation immigrant students tested in PISA studies 2003, 2006, 2009, 2012, and 2015. The dependent variable is a student's PISA grade-point average, computed as the average normalized test score of mathematics, science, and reading. Each subject score is normalized to be mean 0 and standard deviation 1 in our estimation sample. All regressions include the female dummy (non-reported). Standard errors are adjusted for clustering at parents' country-of-origin level (combining mother's and father's origin and distinguishing between the two). ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

This result holds irrespective of whether we analyze each of the four cultural dimensions in isolation (Columns 1 through 4) or include them jointly in a regression (Column 5). In terms of effect sizes, suppose once more that immigrants from Denmark (MAS=0.16) had the same degree of achievement orientation as those from Germany (MAS=0.66): our estimates in Column 5 suggest that this would cause a relative GPA disadvantage of girls compared with boys of almost one-sixth of a standard deviation, i.e., we would observe a change from a negligibly small male-favorable GPA gap of 1% of a standard deviation to a substantial male-favorable GPA gap of 14% of a standard deviation. These findings pass several sensitivity checks, specified to resemble those we have conducted for Sweden (see Appendix Table C.10).

The results are also confirmed when, instead of using students' PISA grade-point average, we analyze their subject scores in math, science and reading separately. The findings, reported in Appendix Tables C.11 through C.13, can be summarized as follows. In our PISA sample of second-generation immigrants, girls have, on average, higher reading scores than boys, but they are outperformed by boys in math and science. Among children from achievement-oriented cultures, girls' comparative advantage in reading vanishes, while the math and science gap in favor of boys significantly increases.

3.9 Conclusion

We have studied the cultural origins of gender gaps in student achievement, departing from the existing literature in two important ways. The first point of departure concerns how we operationalize culture. In the social sciences, culture is often described by the analogy of an onion, with basic beliefs, values, and attitudes forming the core of culture and actual behavior and manifestations thereoff representing the outer layers. From this perspective, the majority of related studies to date has focused not on the role of core cultural values and beliefs *per se*, but on one important manifestation of culture in society, namely whether more gender equality is associated with an educational advantage of girls relative to boys Guiso et al., 2008; Nollenberger et al., 2016. Our analysis adds to this literature by shifting focus to cultural values, beliefs an attitudes that plausibly underlie manifestations of gender (in)equality in society. In particular, based on the multi-dimensional measures of culture developed by Dutch sociologist Geert Hofstede, and motivated by hypotheses derived from the economics literature on gender differences and gender convergence, we have explored whether and how cultural dimensions such as achievement orientation, acceptance of inequality, risk avoidance, and long-term orientation relate to gender gaps in student achievement.

Second, on research methods and depth of analysis, we have used administrative data linking children, parents and schools to study the cultural origins of gender gaps in student achievement. The first key advantage over student survey data used in related studies is that its detail and scale allows for a tightly-controlled, well-powered test of culture's impact on gender gaps in student achievement. Building on the epidemiological approach, our test relies on within-family, cross-gender sibling comparisons, controlling for unobserved family heterogeneity while identifying the differential effect of culture on girls relative to boys. The second advantage of our data is that it allows for an in-depth analysis of potential mechanisms linking culture to gender gaps in student achievement. Indeed, our analysis of mechanisms builds a bridge from the economics of culture to recent advances in the economics of education quantifying the contribution of school quality or family disadvantage to the gender gap in academic outcomes.

We conclude from our exercise that the central cultural dimension that matters for gender gaps in education is the extent to which a society emphasizes ambition, competition and achievement, which is strongly predictive of a relative achievement disadvantage of girls relative to boys. Cultural dimensions such as long-term orientation, inequality acceptance, and uncertainty avoidance matter too, but they do not as strongly and robustly influence the gender gap in academic outcomes. An important mechanism driving our main result appears to be parental school choice: parents from achievement-oriented cultures place their children in higher quality schools compared to those from other cultures, which is more consequential for boys than it is for girls. It is, however, not the case that parents with different cultural origins intentionally treat sons and daughters differently when choosing schools for their offspring, i.e., we find no evidence that they enroll sons in higher quality schools than daughters. These findings underscore the value of an augmented epidemiological approach combining detailed administrative data with multi-dimensional measures of cultural beliefs and attitudes. It offers the opportunity to open-up the black box of cultural transmission and provide a nuanced account of possible pathways from specific cultural traits to gendered economic outcomes. Determining whether the gendered effects of culture highlighted here persist once youth enter the labor market and form families on their own is a promising and important area for future work.

Appendix A

Local Labor Markets and Health at Birth

A.1 Additional Figures

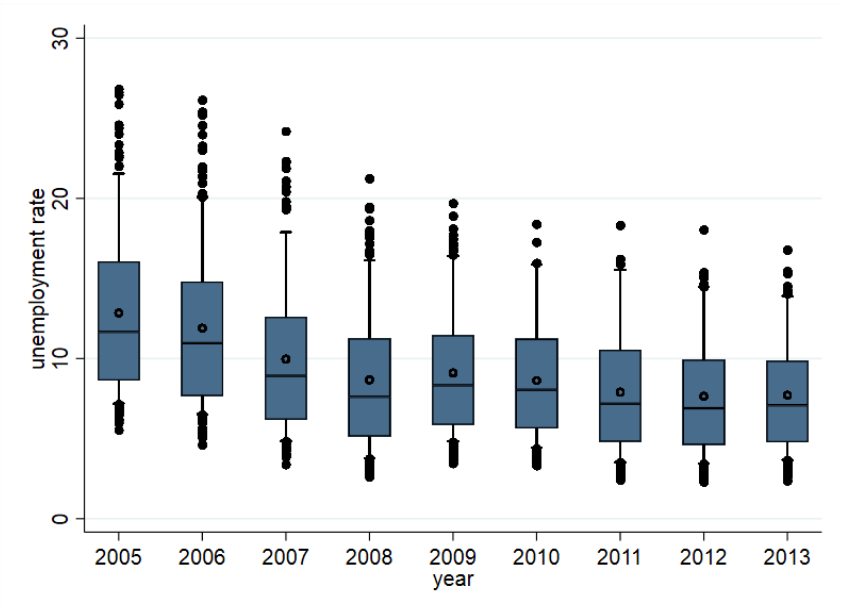


Figure A.1: Distribution of the Annual Unemployment Rate (in %) between 2005 and 2013

Notes: The figure illustrates the LLM level distribution of the annual unemployment rate between 2005 and 2013. Colored boxes indicate the lower and upper quartiles of the LLM observations. The median is represented by the line subdividing the box. The mean is added by the hollow point symbol. Whiskers represent the 5th and 95th percentile of the distribution. Numbers are weighted by the average number of live births in a LLM. The figure is inspired by Cox (2009).

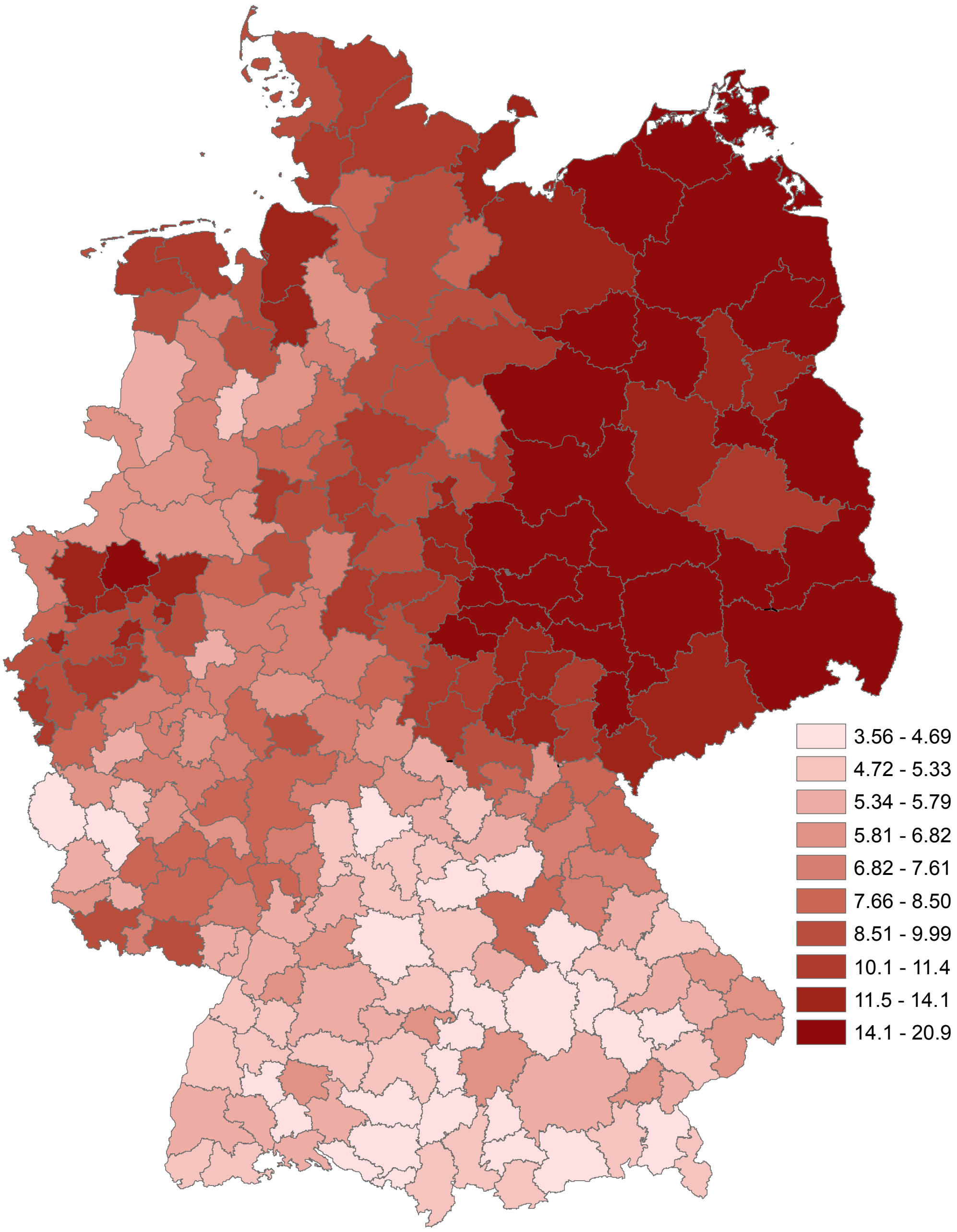


Figure A.2: Mean Unemployment Rate across Regions (in %)

Notes: This map pictures the variation in the mean unemployment rate across LLMs. It refers to the mean of $UR_{\tilde{i}}$ per LLM over the sample period. The different colors refer to the different deciles of the distribution of the unemployment rate.

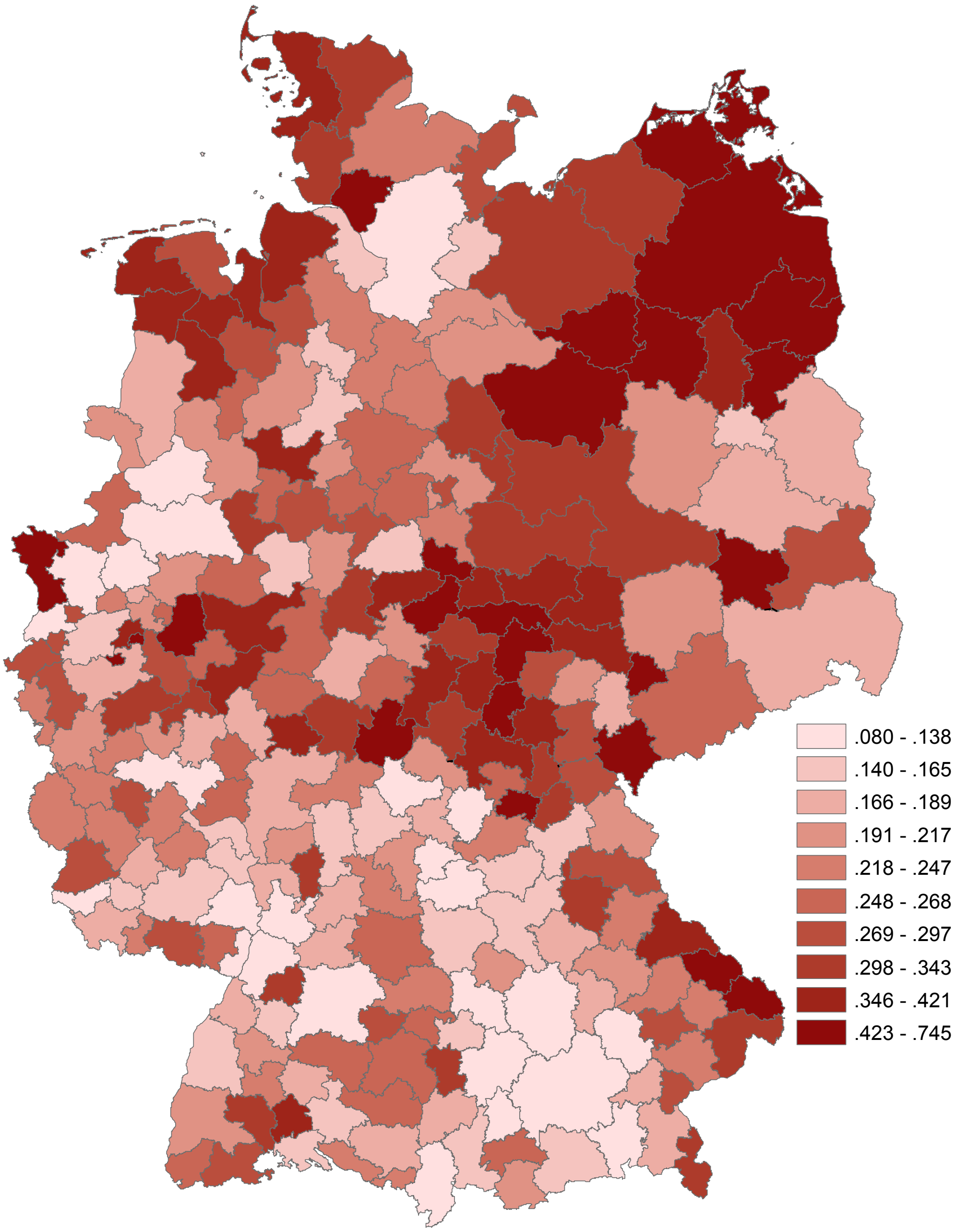


Figure A.3: Residual Unemployment Rate across Regions (in pp)

Notes: This map pictures the variation in the residual unemployment rate across LLMs. It refers to the mean absolute residual per LLM of regressing $UR_{\tilde{t}}$ on population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, month-by-year fixed effects, and a LLM-specific time trend. The different colors refer to the different deciles of the distribution of the residual unemployment rate.

A.2 Additional Tables

Table A.1: Summary Statistics on Regional Characteristics

	Mean	Std Dev	min	max
Fertility Outcomes				
Fertility rate (per 1000 women, 15-44 years)	45.13	5.42	21.31	78.33
Fertility rate (per 1000 women, 15-24 years)	24.23	5.29	6.76	49.75
Fertility rate (per 1000 women, 25-29 years)	85.18	13.83	47.26	158.60
Fertility rate (per 1000 women, 30-34 years)	84.25	13.38	25.29	139.60
Fertility rate (per 1000 women, 35-44 years)	21.76	5.95	4.95	44.33
Average age mother	30.39	0.93	26.07	33.65
Share of married mothers (per 1000 births)	672.94	136.46	239.44	893.00
Share of first births (per 1000 births)	450.10	41.84	179.10	631.07
Labor market variables				
Unemployment Rate	8.99	4.04	1.90	28.48
Demographic variables				
Population size (in Mill.)	0.96	0.96	0.06	3.46
Share of the population, foreign	0.09	0.04	0.01	0.19
Share of pop., aged 15-64 years	0.66	0.01	0.61	0.71
Share of pop., female aged 15-44 years	0.19	0.02	0.13	0.22
Influx of intra-county movers, share	0.04	0.01	0.01	0.18
Influx of intra-county movers, share	0.04	0.01	0.02	0.19
Share school-leavers without degree	0.07	0.02	0.00	0.16
Share with university-entrance degree	0.31	0.09	0.00	0.56
Environmental pollution				
Sulfur dioxide (in $\mu g/m^3$)	3.54	2.03	0.41	19.92
Nitrogen dioxide (in $\mu g/m^3$)	28.57	10.68	5.34	69.57
Particulate matter 10 (in $\mu g/m^3$)	22.82	7.37	5.70	79.97
Carbon monoxide (in $\mu g/m^3$)	0.41	0.14	0.09	1.15

Notes: Sources: Own calculations based on German birth registry (2005-2013); Regional statistics (2005-2013); Pollution data.

Table A.2: Impact of Unemployment on "Liveborn infant" (Z38) per 1,000 Live Births

				Trimester specification					
ICD-10 Chapter and Chapter Title		$UR_{\tilde{t}}$ (1)	R^2 (2)	UR_{t-10} (3)	Tri_1 (4)	Tri_2 (5)	Tri_3 (6)	R^2 (7)	Mean (SD) (8)
Z38	Liveborn infant,	-4.0058	.53	0.4154	-1.0373	0.7371	-4.5309**	.53	704.2490
	no trends	(4.0284)		(1.9258)	(1.5589)	(1.5181)	(2.0187)		94.8378
Z38	Liveborn infant,	-7.9342**	.64	-1.2841	-1.7042	-0.7661	-5.6206***	.64	704.2490
	with regional trends	(3.6399)		(1.7700)	(1.6038)	(1.4912)	(2.0927)		94.8378

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. Dependent variable is the regional diagnosis rate of "Liveborn infants" (aggregated number of newborns with diagnosis code Z38 per 1,000 live births)

Table A.3: Impact of Unemployment on Newborn Health by All Diagnosis Chapters

				Trimester specification					
ICD-10	*Chapter and Chapter Title*	$UR_{\tilde{t}}$ (1)	R^2 (2)	UR_{t-10} (3)	Tri_1 (4)	Tri_2 (5)	Tri_3 (6)	R^2 (7)	Mean (SD) (8)
A-T	Any health problems	6.1591***	.16	1.7960	1.5121	0.3061	4.1478***	.16	299.46
		(2.2848)		(1.2076)	(1.0834)	(1.1709)	(1.2280)		*458.02*
AB	Infectious, Parasitic	-0.4058	.02	-0.0007	0.1296	-0.4467***	0.0186	.02	3.12
		(0.2814)		(0.0787)	(0.1666)	(0.1217)	(0.0794)		*55.81*
C-Da	Neoplasms	0.0019	.01	-0.0346	0.0391	-0.0650	0.0708*	.01	0.67
		(0.0427)		(0.0333)	(0.0377)	(0.0407)	(0.0361)		*25.87*
Db	Blood, Immune system	0.0147	.00	-0.0077	0.0119	0.0114	-0.0077	.00	0.24
		(0.0254)		(0.0237)	(0.0311)	(0.0215)	(0.0175)		*15.38*
E	Endocrine, Nutritional, Metabolic	0.1273**	.01	-0.0193	0.0981*	0.0328	0.0059	.01	1.06
		(0.0580)		(0.0437)	(0.0591)	(0.0420)	(0.0466)		*32.59*
F	Mental, Behavioral, Neurodevelopm.	-0.0222	.03	-0.0289	-0.0121	-0.0273	0.0417	.03	0.15
		(0.0263)		(0.0235)	(0.0208)	(0.0324)	(0.0492)		*12.32*
G	Nervous system	-0.0229	.00	0.0570**	-0.0611	0.0113	-0.0072	.00	0.64
		(0.0401)		(0.0278)	(0.0411)	(0.0319)	(0.0314)		*25.2*
Ha	Eye and adnexa	0.0133	.00	0.0355*	-0.0555**	0.0497**	-0.0137	.00	0.29
		(0.0299)		(0.0186)	(0.0261)	(0.0226)	(0.0215)		*17.10*
Hb	Ear and mastoid process	0.0554*	.01	0.0556**	-0.0221	0.0610***	-0.0278	.01	0.32
		(0.0312)		(0.0270)	(0.0263)	(0.0192)	(0.0278)		*17.78*
I	Circulatory system	-0.0173	.01	0.0213	-0.0208	-0.0151	0.0124	.01	0.75
		(0.0437)		(0.0365)	(0.0532)	(0.0424)	(0.0367)		*27.46*
J	Respiratory system	0.2544**	.01	-0.2843***	0.1299	0.2428**	-0.0343	.01	4.08
		(0.1140)		(0.0990)	(0.0963)	(0.0941)	(0.0929)		*63.77*
K	Digestive system	-0.1176*	.01	0.0234	-0.1324	0.0377	-0.0595	.01	2.26
		(0.0666)		(0.0597)	(0.0804)	(0.0631)	(0.0548)		*47.51*
L	Skin, Subcutaneous tissue	-0.0064	.01	0.0498	-0.0510	-0.0943**	0.1549***	.01	1.24
		(0.0793)		(0.0480)	(0.0495)	(0.0441)	(0.0499)		*35.14*
M	Musculoskel., Connective tissue	0.0131	.04	-0.0178	0.0082	0.0157	-0.0061	.04	0.10
		(0.0198)		(0.0137)	(0.0169)	(0.0149)	(0.0147)		*10.1*
N	Genitourinary system	-0.0479	.00	-0.0236	0.0484	-0.0510	-0.0237	.00	0.87
		(0.0446)		(0.0415)	(0.0583)	(0.0477)	(0.0337)		*29.46*
P	Conditions origin. in perinatal period	5.4967***	.11	1.6791	1.0435	0.4484	3.7451***	.11	254.65
		(2.0236)		(1.0820)	(0.9922)	(1.0587)	(1.0801)		*435.66*
Q	Congenital malform., Chromos. abnorm.	0.5852	.06	0.2813	0.2900	-0.0712	0.3168	.06	23.60
		(0.4667)		(0.2682)	(0.3032)	(0.2523)	(0.2845)		*151.80*
R	Other symptoms, abnormal findings	0.1594	.01	-0.0335	0.0987	0.1397	-0.1012	.01	3.77
		(0.1175)		(0.0821)	(0.1130)	(0.1031)	(0.0916)		*61.28*
ST	Injury, Poisoning	0.0775	.00	0.0433	-0.0301	0.0262	0.0627	.00	1.65
		(0.0559)		(0.0463)	(0.0665)	(0.0548)	(0.0504)		*40.55*
	Health-related diagnoses other than P, Q	0.0772	.03	-0.1645	0.1787	-0.0712	0.0859	.03	21.21
		(0.4030)		(0.1856)	(0.2490)	(0.2654)	(0.2287)		*144.1*
Hospital stay (uncond'l)		0.0152	.06	0.0212*	-0.0008	-0.0055	0.0151	.06	5.28
		(0.0171)		(0.0125)	(0.0173)	(0.0143)	(0.0139)		*8.72*
Hospital stay (cond'l on health-related diagnosis)		-0.0824	.08	0.0273	-0.0319	-0.0215	-0.0459	.08	9.27
		(0.0512)		(0.0385)	(0.0514)	(0.0497)	(0.0464)		*14.95*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, hospital fixed effects, LLM fixed effects, month-by-year fixed effects, and the gender of the newborn. Notes: The number of observations is 5,531,003 in all regressions except for "Hospital stay (cond'l on health-related diagnosis)" (1,656,314). Estimates stem from regressions controlling for linear local labor market trends. The outcomes are measured as ratio of number of certain diagnoses per 1000 hospital cases.

Table A.4: Hospital Diagnoses of Newborns Controlling for Fertility Composition

	Any		Perinatal health		Congenital defects		Neonatal mortality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Hospital cases								
$UR_{\tilde{t}}$	3.2059	6.1096***	3.2043	5.4168***	0.1132	0.5672	-0.0699	-0.0809
	(2.5539)	(2.2908)	(2.2767)	(2.0262)	(0.4636)	(0.4686)	(0.0585)	(0.0660)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
UR_{t-10}	0.7409	1.8028	1.0083	1.7146	0.0199	0.2897	-0.0260	-0.0336
	(1.3486)	(1.1913)	(1.2302)	(1.0682)	(0.2720)	(0.2709)	(0.0654)	(0.0702)
Tri_1	0.7853	1.5484	0.4021	1.1588	0.2538	0.2616	-0.0220	-0.0202
	(1.0222)	(1.0514)	(0.9374)	(0.9660)	(0.2914)	(0.3019)	(0.0759)	(0.0775)
Tri_2	-0.7253	0.1673	-0.4964	0.2101	-0.2384	-0.0728	-0.0818	-0.0815
	(1.1606)	(1.1550)	(1.0595)	(1.0478)	(0.2541)	(0.2568)	(0.0623)	(0.0622)
Tri_3	2.9915**	4.1969***	2.9057**	3.8024***	0.1300	0.3180	0.0680	0.0608
	(1.2615)	(1.2324)	(1.1196)	(1.0868)	(0.2694)	(0.2830)	(0.0625)	(0.0651)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
Mean	299.46	299.46	254.65	254.65	23.6	23.6	2.42	2.42
Standard Deviation	458.02	458.02	435.66	435.66	151.8	151.8	49.1	49.1
Observations	5531003	5531003	5531003	5531003	5531003	5531003	5531003	5531003
Panel B: Diagnosis rates								
$UR_{\tilde{t}}$	3.7298	3.8500*	3.6548	3.5246*	0.2492	0.4755	-0.0678	-0.1040
	(2.5573)	(2.3297)	(2.2597)	(2.0613)	(0.4783)	(0.5004)	(0.0574)	(0.0687)
R^2	.53	.61	.5	.58	.4	.47	.03	.04
UR_{t-10}	1.4075	1.3608	1.5509	1.3410	0.0998	0.2831	-0.0203	-0.0343
	(1.3700)	(1.2805)	(1.2189)	(1.1261)	(0.2863)	(0.2884)	(0.0671)	(0.0712)
Tri_1	0.8800	1.0620	0.5149	0.7682	0.2945	0.2718	-0.0229	-0.0304
	(1.1708)	(1.2022)	(1.0301)	(1.0610)	(0.3110)	(0.3219)	(0.0794)	(0.0818)
Tri_2	-0.3285	-0.3170	-0.1355	-0.1392	-0.2040	-0.1359	-0.0810	-0.0906
	(1.3302)	(1.3341)	(1.1708)	(1.1670)	(0.2731)	(0.2838)	(0.0658)	(0.0664)
Tri_3	2.4296*	3.0075**	2.3871**	2.7105**	0.1279	0.2910	0.0656	0.0578
	(1.3727)	(1.3723)	(1.1944)	(1.1926)	(0.3002)	(0.3093)	(0.0629)	(0.0673)
R^2	.53	.61	.5	.58	.4	.47	.03	.04
Mean	301.67	301.67	256.35	256.35	23.79	23.79	2.44	2.44
Standard Deviation	70.87	70.87	62.17	62.17	17.32	17.32	3.49	3.49
Observations	24010	24010	24010	24010	24010	24010	24010	24010
Reg. cont.	X	X	X	X	X	X	X	X
Fertility Composition	X	X	X	X	X	X	X	X
LLM trend		X		X		X		X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. The regressions in Panel A additionally control for hospital fixed effects and gender. The regressions in Panel B are weighted by the average number of live births in a LLM. Fertility composition controls for the fertiltiy rates of different age groups, average age of the mother and the shares of married, male, and Christian births, as well as the share of different parities of the total number of births.

Table A.5: Hospital Diagnoses of Newborns with Air Pollution Controls

	Any		Perinatal health		Congenital defects		Neonatal mortality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Hospital cases								
$UR_{\tilde{t}}$	2.7431	6.0085***	2.8400	5.3883***	0.1200	0.5652	-0.0748	-0.0776
	(2.5180)	(2.2537)	(2.2384)	(1.9923)	(0.4659)	(0.4653)	(0.0601)	(0.0665)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
UR_{t-10}	0.9905	2.0619*	1.1859	1.9066*	0.0261	0.2903	-0.0289	-0.0372
	(1.3421)	(1.2133)	(1.2250)	(1.0902)	(0.2715)	(0.2709)	(0.0652)	(0.0700)
Tri_1	0.5759	1.5906	0.2320	1.1177	0.1999	0.2570	-0.0193	-0.0058
	(1.0313)	(1.0781)	(0.9345)	(0.9804)	(0.2971)	(0.3088)	(0.0756)	(0.0776)
Tri_2	-0.8218	0.1797	-0.5184	0.3442	-0.1828	-0.0626	-0.0874	-0.0880
	(1.1945)	(1.1754)	(1.0796)	(1.0613)	(0.2558)	(0.2571)	(0.0609)	(0.0609)
Tri_3	2.6226**	3.9108***	2.5716**	3.5536***	0.1187	0.3100	0.0703	0.0611
	(1.2602)	(1.2306)	(1.1098)	(1.0813)	(0.2748)	(0.2838)	(0.0608)	(0.0627)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
Mean	299.46	299.46	254.65	254.65	23.6	23.6	2.42	2.42
Standard Deviation	458.02	458.02	435.66	435.66	151.8	151.8	49.1	49.1
Observations	5531003	5531003	5531003	5531003	5531003	5531003	5531003	5531003
Panel B: Diagnosis rates								
$UR_{\tilde{t}}$	3.3981	3.8418*	3.4140	3.5997*	0.2731	0.4721	-0.0724	-0.0987
	(2.5239)	(2.3053)	(2.2189)	(2.0397)	(0.4830)	(0.4983)	(0.0584)	(0.0690)
R^2	.53	.61	.5	.58	.4	.47	.03	.04
UR_{t-10}	1.5726	1.5500	1.6514	1.4535	0.1025	0.2926	-0.0236	-0.0378
	(1.3985)	(1.3117)	(1.2446)	(1.1557)	(0.2886)	(0.2910)	(0.0668)	(0.0711)
Tri_1	0.6873	1.3715	0.3502	0.9540	0.2525	0.2862	-0.0210	-0.0125
	(1.1926)	(1.2173)	(1.0436)	(1.0742)	(0.3160)	(0.3299)	(0.0787)	(0.0820)
Tri_2	-0.3832	-0.4735	-0.1009	-0.1308	-0.1535	-0.1448	-0.0867	-0.0995
	(1.3555)	(1.3441)	(1.1894)	(1.1749)	(0.2740)	(0.2827)	(0.0645)	(0.0653)
Tri_3	2.1866	2.8067**	2.1612*	2.5482**	0.1300	0.2824	0.0694	0.0600
	(1.3714)	(1.3917)	(1.1838)	(1.2077)	(0.3034)	(0.3096)	(0.0616)	(0.0652)
R^2	.53	.61	.5	.58	.4	.47	.03	.04
Mean	301.67	301.67	256.35	256.35	23.79	23.79	2.44	2.44
Standard Deviation	70.87	70.87	62.17	62.17	17.32	17.32	3.49	3.49
Observations	24010	24010	24010	24010	24010	24010	24010	24010
Reg. cont.	X	X	X	X	X	X	X	X
Air pollution	X	X	X	X	X	X	X	X
LLM trend		X		X		X		X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. The regressions in Panel A additionally control for hospital fixed effects and gender. The regressions in Panel B are weighted by the average number of live births in a LLM. Air pollution control variables include carbon monoxide, sulfur dioxide, nitrogen dioxide and particulate matter.

Table A.6: The Effect of Fertility on Hospital Diagnoses of Newborns

	Any		Perinatal health		Congenital defects		Neonatal mortality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Hospital cases								
$fertility_{\tilde{t}}$	0.3805	-0.0032	0.1512	-0.2320	0.0234	0.0379	0.0057	0.0336
	(0.6950)	(0.6072)	(0.6205)	(0.5288)	(0.1170)	(0.1211)	(0.0233)	(0.0254)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
$fertility_{t-10}$	-0.0444	-0.0886	-0.0521	-0.0979	-0.0283	-0.0235	-0.0003	0.0027
	(0.1061)	(0.0930)	(0.1005)	(0.0888)	(0.0271)	(0.0271)	(0.0071)	(0.0073)
Tri_1	0.1341	0.0095	0.0306	-0.1017	0.0293	0.0396	0.0014	0.0104
	(0.2580)	(0.2323)	(0.2335)	(0.2069)	(0.0565)	(0.0579)	(0.0122)	(0.0127)
Tri_2	-0.0284	-0.1756	0.0439	-0.1012	0.0003	-0.0026	0.0040	0.0123
	(0.2697)	(0.2456)	(0.2385)	(0.2119)	(0.0542)	(0.0581)	(0.0132)	(0.0135)
Tri_3	0.2765	0.1461	0.0728	-0.0549	-0.0079	-0.0050	0.0002	0.0116
	(0.2494)	(0.2247)	(0.2300)	(0.2057)	(0.0507)	(0.0496)	(0.0135)	(0.0140)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
Mean	299.46	299.46	254.65	254.65	23.6	23.6	2.42	2.42
Standard Deviation	458.02	458.02	435.66	435.66	151.8	151.8	49.1	49.1
Observations	5531003	5531003	5531003	5531003	5531003	5531003	5531003	5531003
Panel B: Diagnosis rates								
$fertility_{\tilde{t}}$	-0.1615	-0.1999	-0.3147	-0.4186	0.0036	0.0614	0.0110	0.0399
	(0.6729)	(0.6006)	(0.6104)	(0.5485)	(0.1196)	(0.1224)	(0.0228)	(0.0257)
R^2	.53	.61	.5	.58	.39	.47	.03	.04
$fertility_{t-10}$	-0.0958	-0.0847	-0.0906	-0.0916	-0.0372	-0.0278	0.0000	0.0034
	(0.1212)	(0.1070)	(0.1106)	(0.0986)	(0.0285)	(0.0281)	(0.0073)	(0.0076)
Tri_1	-0.1804	-0.1501	-0.2189	-0.2216	0.0040	0.0281	0.0007	0.0104
	(0.2626)	(0.2448)	(0.2381)	(0.2231)	(0.0580)	(0.0602)	(0.0123)	(0.0130)
Tri_2	-0.1665	-0.2097	-0.0850	-0.1508	0.0001	0.0147	0.0101	0.0187
	(0.2659)	(0.2545)	(0.2387)	(0.2278)	(0.0582)	(0.0610)	(0.0132)	(0.0137)
Tri_3	0.1822	0.1447	-0.0183	-0.0694	-0.0027	0.0114	-0.0001	0.0114
	(0.2625)	(0.2323)	(0.2431)	(0.2176)	(0.0535)	(0.0517)	(0.0134)	(0.0142)
R^2	.53	.61	.5	.58	.39	.47	.03	.04
Mean	301.67	301.67	256.35	256.35	23.79	23.79	2.44	2.44
Standard Deviation	70.87	70.87	62.17	62.17	17.32	17.32	3.49	3.49
Observations	24010	24010	24010	24010	24010	24010	24010	24010
Reg. cont.	X	X	X	X	X	X	X	X
LLM trend		X		X		X		X

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. The regressions in Panel A additionally control for hospital fixed effects and gender. The regressions in Panel B are weighted by the average number of live births in a LLM.

Table A.7: Hospital Diagnoses of Newborns Controlling for Fertility

	Any		Perinatal health		Congenital defects		Neonatal mortality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Hospital cases								
$UR_{\tilde{t}}$	3.1678	6.1827***	3.2055	5.4876***	0.0982	0.5951	-0.0691	-0.0756
	(2.5289)	(2.2838)	(2.2634)	(2.0236)	(0.4618)	(0.4660)	(0.0590)	(0.0662)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
UR_{t-10}	0.7412	1.7688	1.0214	1.6914	-0.0059	0.2697	-0.0268	-0.0358
	(1.3671)	(1.2077)	(1.2511)	(1.0836)	(0.2721)	(0.2681)	(0.0647)	(0.0695)
Tri_1	0.6745	1.4494	0.3323	1.0718	0.2268	0.2631	-0.0278	-0.0196
	(1.0259)	(1.0784)	(0.9361)	(0.9920)	(0.2899)	(0.3021)	(0.0746)	(0.0766)
Tri_2	-0.6398	0.3832	-0.4250	0.4135	-0.1896	-0.0381	-0.0781	-0.0783
	(1.1665)	(1.1715)	(1.0632)	(1.0556)	(0.2532)	(0.2536)	(0.0606)	(0.0608)
Tri_3	2.9607**	4.1475***	2.8814**	3.7452***	0.1024	0.3167	0.0714	0.0637
	(1.2689)	(1.2276)	(1.1188)	(1.0805)	(0.2692)	(0.2840)	(0.0606)	(0.0628)
R^2	.15	.16	.11	.11	.06	.06	.01	.01
Mean	299.46	299.46	254.65	254.65	23.6	23.6	2.42	2.42
Standard Deviation	458.02	458.02	435.66	435.66	151.8	151.8	49.1	49.1
Observations	5531003	5531003	5531003	5531003	5531003	5531003	5531003	5531003
Panel B: Diagnosis rates								
$UR_{\tilde{t}}$	3.4525	3.9538*	3.4664	3.6299*	0.2151	0.5041	-0.0707	-0.0985
	(2.5515)	(2.3146)	(2.2572)	(2.0502)	(0.4799)	(0.4988)	(0.0575)	(0.0690)
R^2	.53	.61	.5	.58	.4	.47	.03	.04
UR_{t-10}	1.3815	1.3374	1.5362	1.3197	0.0754	0.2663	-0.0205	-0.0351
	(1.4102)	(1.3050)	(1.2581)	(1.1509)	(0.2874)	(0.2861)	(0.0664)	(0.0707)
Tri_1	0.6918	1.0088	0.3797	0.7180	0.2688	0.2825	-0.0302	-0.0302
	(1.1788)	(1.2170)	(1.0321)	(1.0800)	(0.3096)	(0.3221)	(0.0779)	(0.0808)
Tri_2	-0.2465	-0.1182	-0.0531	0.0614	-0.1636	-0.1090	-0.0785	-0.0881
	(1.3241)	(1.3398)	(1.1629)	(1.1621)	(0.2714)	(0.2809)	(0.0643)	(0.0651)
Tri_3	2.2418	2.9526**	2.2309*	2.6472**	0.0870	0.2894	0.0675	0.0615
	(1.3795)	(1.3751)	(1.1886)	(1.1906)	(0.2999)	(0.3095)	(0.0612)	(0.0651)
R^2	.53	.61	.5	.58	.4	.47	.03	.04
Mean	301.67	301.67	256.35	256.35	23.79	23.79	2.44	2.44
Standard Deviation	70.87	70.87	62.17	62.17	17.32	17.32	3.49	3.49
Observations	24010	24010	24010	24010	24010	24010	24010	24010
Reg. cont.	X	X	X	X	X	X	X	X
Fertility	X	X	X	X	X	X	X	X
LLM trend		X		X		X		X

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. The regressions in Panel A additionally control for hospital fixed effects and gender. The regressions in Panel B are weighted by the average number of live births in a LLM. Fertility controls for the number of births per 1000 women aged 15-44.

Table A.8: Impact of Approximated Female and Male Unemployment on Health at Birth

	Panel A: Hospital cases		Panel B: Diagnosis rates	
	(1)	(2)	(3)	(4)
Average unemployment				
$UR_{\tilde{t}} - male$	2.3748	4.0806	2.4161	2.4459
	(3.0863)	(2.8880)	(3.1507)	(2.6909)
$UR_{\tilde{t}} - female$	0.7631	2.0048	1.0120	1.3795
	(2.9618)	(3.2346)	(3.0485)	(3.0815)
R^2	.15	.16	.53	.61
P-value $UR_{\tilde{t}}$ $\beta_m = \beta_f$	.7695	.7153	.8038	.8404
Trimester specification				
$UR_{t-10} - male$	1.2035	1.7316	1.2634	1.3824
	(1.1210)	(1.0663)	(1.1885)	(1.1845)
$Tri_1 - male$	1.2458	1.8604*	2.2292*	2.4871**
	(1.0645)	(1.1179)	(1.2536)	(1.2196)
$Tri_2 - male$	0.1311	0.8579	0.1690	0.3430
	(1.2253)	(1.2182)	(1.3330)	(1.3213)
$Tri_3 - male$	0.4046	1.6322	-0.5876	-0.3685
	(1.4442)	(1.4098)	(1.5857)	(1.4937)
$UR_{t-10} - female$	-0.4686	0.1393	0.4359	0.2299
	(1.5876)	(1.4486)	(1.6772)	(1.5771)
$Tri_1 - female$	-1.2711	-1.0235	-2.8220	-2.8180
	(1.6321)	(1.6805)	(2.0234)	(1.9868)
$Tri_2 - female$	-0.8792	-0.5691	-0.6377	-0.7180
	(1.8476)	(1.8304)	(1.9415)	(1.9607)
$Tri_3 - female$	3.3804*	3.1820*	4.0629**	4.6809**
	(1.9649)	(1.9168)	(2.0211)	(1.9391)
R^2	.15	.16	.53	.61
P-value UR_{t-10} $\beta_m = \beta_f$	.4664	.4660	.7359	.6301
P-value Tri_1 $\beta_m = \beta_f$	.2937	.2489	.0878	.0650
P-value Tri_2 $\beta_m = \beta_f$	.7172	.6064	.7852	.7210
P-value Tri_3 $\beta_m = \beta_f$	.3405	.6124	.1609	.1083
Obs.	5,531,003	5,531,003	24,010	24,010
Reg. cont.	X	X	X	X
LLM trend		X		X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the LLM level (in parentheses). Control variables include population size, its composition with respect to age, gender, migration background and relocations, the share of school leavers with university-entrance degree and without any degree, LLM fixed effects, and month-by-year fixed effects. The regressions in Panel A additionally control for hospital fixed effects and gender. The regressions in Panel B are weighted by the average number of live births in a LLM.

Table A.9: Cost of Illness by ICD-10 Diagnosis

ICD-10		Description	Hospital Statistics on Main Diagnoses (2015): Infants below age 1 (2015) Cases	in % (of I-XIX)	Avg. stay (days)	Deaths	Patients of all ages (2015) Cases	in % (of I-XIX)	Avg. stay (days)	Cost of illness in millions of Euro (2015), by type of provider: All providers*	Hospitals	Hospitals in total cost (%)	Avg. hospital cost per day
A. All diseases and sequelae of effects of external causes (A00-T98)													
I	AB	Certain infectious and parasitic diseases	25,461	6.7	3.6	27	618,031	3.3	7.5	7,081	2,572	36.3	555.2
II	CDa	Neoplasms	4,177	1.1	4.5	10	1,823,716	9.6	7.7	23,002	10,984	47.8	782.1
III	Db	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	967	0.3	7.0	7	134,727	0.7	6.5	2,395	555	23.2	630.2
IV	E	Endocrine, nutritional and metabolic diseases	1,873	0.5	7.2	16	517,959	2.7	7.6	15,609	2,114	13.5	535.5
V	F	Mental, Behavioral and Neurodevelopmental disorders	785	0.2	9.5	1	1,224,128	6.4	20.9	44,372	9,780	22.0	381.4
VI	G	Diseases of the nervous system	4,560	1.2	8.5	36	774,907	4.1	6.7	17,150	3,023	17.6	581.4
VII	Ha	Diseases of the eye and adnexa	1,079	0.3	3.9		338,687	1.8	3.1	11,186	774	6.9	734.5
VIII	Hb	Diseases of the ear and mastoid process	2,137	0.6	3.4		157,443	0.8	4.0	3,225	491	15.2	781.1
IX	I	Diseases of the circulatory system	1,500	0.4	10.5	54	2,888,142	15.2	7.8	46,436	17,056	36.7	761.4
X	J	Diseases of the respiratory system	48,627	12.8	4.3	50	1,287,378	6.8	7.0	16,544	5,197	31.4	576.9
XI	K	Diseases of the digestive system	11,735	3.1	3.7	13	1,922,135	10.1	5.8	41,620	7,244	17.4	651.4
XII	L	Diseases of the skin and subcutaneous tissue	3,934	1.0	5.3		292,564	1.5	7.0	5,199	1,039	20.0	507.6
XIII	M	Diseases of musculoskeletal system and connective tissue	479	0.1	7.6		1,782,500	9.4	7.4	34,193	8,750	25.6	664.2
XIV	N	Diseases of the genitourinary system	8,597	2.3	5.6	2	1,035,501	5.5	5.4	11,289	3,433	30.4	617.5
XV	O	Pregnancy, childbirth and the puerperium					1,012,108	5.3	3.9	4,229	2,326	55.0	
XVI	P	Certain conditions originating in the perinatal period	190,848	50.3	9.0	1,540	190,932	1.0	9.0	1,508	1,298	86.1	751.5
XVII	Q	Congenital malformations, deformations and chromosomal abnormalities	30,633	8.1	7.6	340	103,084	0.5	5.5	2,170	695	32.0	1,222.5
XVIII	R	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	17,126	4.5	2.7	54	944,170	5.0	4.0	20,026	2,011	10.0	536.6
XIX	ST	Injury, poisoning, other consequences of external causes	24,850	6.6	2.2	13	1,941,602	10.2	7.1	17,993	9,220	51.2	667.3
I - XIX	*A-T*	*All diseases and sequelae of effects of external causes*	*379,368*	*100.0*	*6.8*	*2,163*	*18,989,714*	*100*	*7.5*	*325,228*	*88,562*	*27.2*	*618.9*
XXI	Z	Z30-Z39 Health services in circumstances related to reproduction	529,157	3	3		529,157		3.1	2,183	512	23.5	307.5

Notes: Own calculations of average cost per hospital day based on official statistics of main diagnoses and cost of illness by ICD-10 code and type of provider, German Statistical Office (Bonn). *All providers includes hospitals, residential long-term care facilities, providers of ambulatory health care, providers of ancillary services, Retailers and other providers of medical goods, providers of preventive care, providers of health care system administration and financing, rest of the economy, rest of the world.

A.3 Health Outcomes from Hospital Diagnosis Data

A.3.1 Newborn Health Outcomes from Hospital Diagnosis Data

- **diag_nb**: Indicator for health-related hospital stay of a newborn

  I. **diag_AB**: Certain infectious and parasitic diseases (A00-B99)

  II. **diag_CDa**: Neoplasms (C00-D48)

  III. **diag_Db**: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89)

  IV. **diag_E**: Endocrine, nutritional and metabolic diseases (E00-E90)

  V. **diag_F**: Mental and behavioural disorders (F00-F99)

  VI. **diag_G**: Diseases of the nervous system (G00-G99)

  VII. **diag_Ha**: Diseases of the eye and adnexa (H00-H59)

  VIII. **diag_Hb**: Diseases of the ear and mastoid process (H60-H95)

  IX. **diag_I**: Diseases of the circulatory system (I00-I99)

  X. **diag_J**: Diseases of the respiratory system (J00-J99)

  XI. **diag_K**: Diseases of the digestive system (K00-K93)

  XII. **diag_L**: Diseases of the skin and subcutaneous tissue (L00-L99)

  XIII. **diag_M**: Diseases of the musculoskeletal system and connective tissue (M00-M99)

  XIV. **diag_N**: Diseases of the genitourinary system (N00-N99)

  XVI. **diag_P**: Certain conditions originating in the perinatal period excluding stillbirth (P00-P94, P96)

    (a) **P_mat**: Fetus and newborn affected by maternal factors and by complications of pregnancy, labor and delivery (P00-P04)

    (b) **abnorm**: Disorders related to length of gestation and fetal growth (P05-P08)

      – **elight**: Disorders related to slow fetal growth, fetal malnutrition or short gestation and low birth weight (P05, P07)

        * **P_05**: Slow fetal growth and fetal malnutrition (P05)
        * **P_07**: Disorders related to short gestation and low birth weight, not elsewhere classified (P07)

          · **P_1000**: Extremely low birth weight; below 1000g (P07.0)

          · **P_2500**: Low birth weight; below 2500g (P07.0, P07.1)

          · **P_prem**: Preterm infants; less than 37 weeks of gestation (P07.2, P07.3)

      – **hvylate**: Disorders related to long gestation and high birth weight (P08)

    (c) **P_traum**: Birth trauma (P10-P15)

(d) Respiratory and cardiovascular disorders specific to the perinatal period (P20-P29)
   - **P_resp**: Respiratory disorders specific to the perinatal period (P20, P22-P24, P25.0-P25.2, P25.8, P26-P28)
     * **P_22**: Respiratory distress of newborn (P22)
   - **P_21**: Birth asphyxia (P21)
   - **P_card**: Cardiovascular disorders specific to the perinatal period (P25.3, P29)

(e) **P_infec**: Infections specific to the perinatal period (P35-P39)

(f) **P_haemo**: Haemorrhagic and haematological disorders of fetus and newborn (P50-P61)
   - **P_jaun**: Neonatal jaundice (P58, P59)

(g) **P_metab**: Transitory endocrine and metabolic disorders specific to fetus and newborn (P70-P74)

(h) **P_diges**: Digestive system disorders of fetus and newborn (P75-P78)

(i) **P_tmprg**: Conditions involving the integument and temperature regulation of fetus and newborn (P80-P83)

(j) **P_other**: Other disorders originating in the perinatal period (P90-P94, P96)

XVII. **diag_Q**: Congenital malformations, deformations and chromosomal abnormalities (Q00-Q99)

(a) **Q_nerv**: Congenital malformations of the nervous system (Q00-Q07)

(b) **Q_face**: Congenital malformations of eye, ear, face and neck (Q10-Q18)

(c) **Q_circ**: Congenital malformations of the circulatory system (Q20-Q28)

(d) **Q_resp**: Congenital malformations of the respiratory system (Q30-Q34)

(e) **Q_cleft**: Cleft lip and cleft palate (Q35-Q37)

(f) **Q_diges**: Other congenital malformations of the digestive system (Q38-Q45)

(g) **Q_gntal**: Congenital malformations of genital organs (Q50-Q56)

(h) **Q_urin**: Congenital malformations of the urinary system (Q60-Q64)

(i) **Q_musc**: Congenital malformations and deformations of the musculoskeletal system (Q65-Q79)

(j) **Q_other**: Other congenital malformations (Q80-Q89)

(k) **chrmdis**: Chromosomal abnormalities, not elsewhere classified (Q90-Q99)

XVIII. **diag_R**: Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)

XIX. **diag_ST**: Injury, poisoning and certain other consequences of external causes (S00-T98)

- **diag_other**: Health-related diagnosis outside of the P and Q diagnosis groups

- (XXI.) Factors influencing health status and contact with health services (Z00-Z99)
  - **lbirths**: Liveborn infants according to place of birth (Z38)
  - **Z_other**: Remaining Z-diagnoses (Z00-Z37 & Z39-Z99)[1]
- **stillb**: Fetal death of unspecified cause (P95)
- **ndeaths**: Indicator for neonatal death
- **stay_all**: Length of hospital stay
- **stay_diag**: Length of hospital stay conditional on having a health-related stay (diag_nb=1)
- **resp**: Diseases of the respiratory system and respiratory disorders specific to the perinatal period (J00-J99,P20, P22-P24, P25.0-P25.2, P25.8, P26-P28)

A.3.2 Maternal Health Outcomes from Hospital Diagnosis Data

Note: These diagnoses are given relatively to 1000 live births (or 1000 live births in the respective age sub-group)

- V. Mental and behavioural disorders (F00-F99)
  - **F_mood**: Mood [affective] disorders (F30-F39)
- XV. Pregnancy, childbirth and the puerperium (O00-O99)
  - **miscar**: Miscarriage / spontaneous abortion (O03)
  - **O_preg**: Maternal health diagnosis during the pregnancy (O10-O16, O20-O29, O98-O99)
  - **O_del**: Maternal health diagnosis related to childbirth[2] (O31-O48, O60, O62-O75)
    - **O_del1**: Maternal care related to the fetus and amniotic cavity and possible delivery problems (O31-O48)
    - **O_del2**: Complications of labor and delivery (O60, O62-O75)

[1] excluded from sample

[2] Does neither include delivery as main diagnosis (O80-O82) nor multiple gestation (O30) or failed induction of labor (O61)

A.4 Local Labor Markets

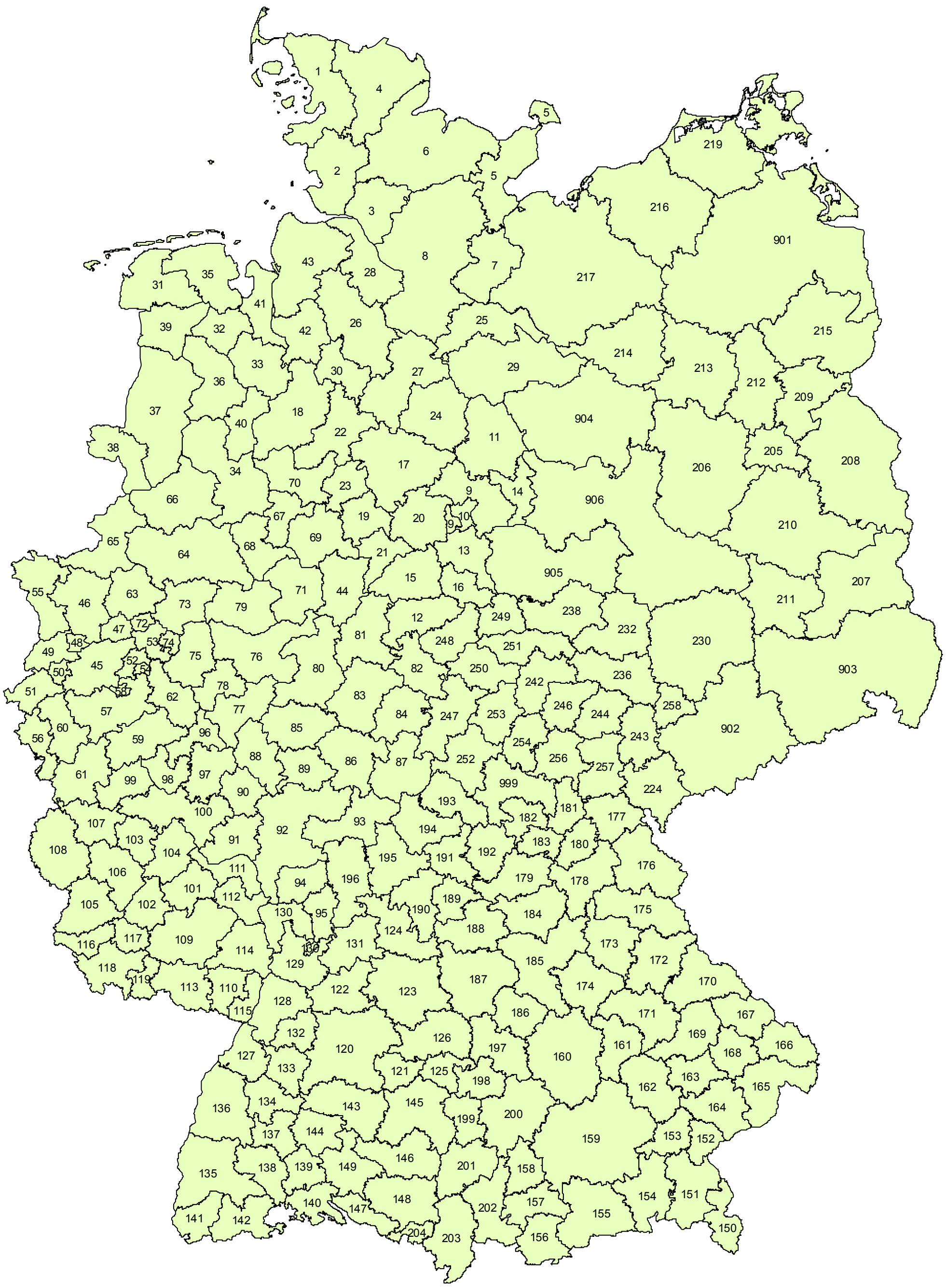


Figure A.4: Map of the Local Labor Markets (LLMs) in Germany

Notes: This map was created by means of a map of the Federal Agency for Cartography and Geodesy (BKG) on the administrative borders in 2014 (VG250). The LLMs follow the 2014 classification of the BBSR, but are modified to account for major district reforms during the observation period (901-906) and to comply with data protection guidelines (999).

1. Husum
2. Heide
3. Itzehoe
4. Flensburg
5. Luebeck
6. Kiel
7. Ratzeburg
8. Hamburg
9. Braunschweig
10. Salzgitter
11. Wolfsburg
12. Göttingen
13. Goslar
14. Helmstedt
15. Einbeck
16. Osterode
17. Hannover
18. Sulingen
19. Hameln
20. Hildesheim
21. Holzminden
22. Nienburg
23. Stadthagen
24. Celle
25. Lüneburg
26. Zeven
27. Soltau
28. Stade
29. Uelzen
30. Verden
31. Emden
32. Westerstede
33. Oldenburg
34. Osnabrück
35. Wilhelmshaven
36. Cloppenburg
37. Lingen
38. Nordhorn
39. Leer
40. Vechta
41. Nordenham
42. Bremen
43. Bremerhaven
44. Höxter
45. Düsseldorf
46. Duisburg
47. Essen
48. Krefeld
49. Viersen
50. Mönchengladbach
51. Heinsberg
52. Wuppertal
53. Schwelm
54. Remscheid
55. Kleve
56. Aachen
57. Köln
58. Leverkusen
59. Bonn
60. Düren
61. Euskirchen
62. Gummersbach
63. Gelsenkirchen
64. Münster
65. Borken
66. Steinfurt
67. Bielefeld
68. Gütersloh
69. Detmold
70. Minden
71. Paderborn
72. Bochum
73. Dortmund
74. Hagen
75. Lüdenscheid
76. Meschede
77. Siegen
78. Olpe
79. Soest
80. Korbach
81. Kassel
82. Eschwege
83. Schwalm-Eder

84. Hersfeld
85. Marburg
86. Lauterbach
87. Fulda
88. Wetzlar
89. Gießen
90. Limburg
91. Wiesbaden
92. Frankfurt/Main
93. Hanau
94. Darmstadt
95. Erbach
96. Altenkirchen
97. Montabaur
98. Neuwied
99. Ahrweiler
100. Koblenz
101. Bad Kreuznach
102. Idar-Oberstein
103. Cochem
104. Simmern
105. Trier
106. Bernkastel-Wittlich
107. Daun
108. Bitburg
109. Kaiserslautern
110. Landau
111. Mainz
112. Alzey-Worms
113. Pirmasens
114. Ludwigshafen
115. Germersheim
116. Merzig
117. St. Wendel
118. Saarbrücken
119. Homburg/Saar
120. Stuttgart
121. Göppingen
122. Heilbronn
123. Schwäbisch Hall
124. Tauberbischofsheim
125. Heidenheim
126. Aalen
127. Baden-Baden
128. Karlsruhe
129. Heidelberg
130. Mannheim
131. Mosbach
132. Pforzheim
133. Calw
134. Freudenstadt
135. Freiburg
136. Offenburg
137. Rottweil
138. Villingen-Schwenningen
139. Tuttlingen
140. Konstanz
141. Lörrach
142. Waldshut
143. Reutlingen/Tübingen
144. Balingen
145. Ulm
146. Biberach
147. Friedrichshafen
148. Ravensburg
149. Sigmaringen
150. Bad Reichenhall
151. Traunstein
152. Burghausen
153. Mühldorf
154. Rosenheim
155. Bad Tölz
156. Garmisch-Partenkirchen
157. Weilheim
158. Landsberg
159. München
160. Ingolstadt
161. Kelheim-Mainburg
162. Landshut
163. Dingolfing
164. Eggenfelden/Pfarrkirchen
165. Passau
166. Freyung

167. Regen-Zwiesel
168. Deggendorf
169. Straubing
170. Cham
171. Regensburg
172. Schwandorf
173. Amberg
174. Neumarkt
175. Weiden
176. Marktredwitz
177. Hof
178. Bayreuth
179. Bamberg
180. Kulmbach
181. Kronach
182. Coburg
183. Lichtenfels
184. Erlangen
185. Nürnberg
186. Weißenburg-Gunzenhausen
187. Ansbach
188. Neustadt/Aisch
189. Kitzingen
190. Würzburg
191. Schweinfurt
192. Haßfurt
193. Bad Neustadt/ Saale
194. Bad Kissingen
195. Lohr am Main
196. Aschaffenburg
197. Donauwörth-Nördlingen
198. Dillingen
199. Günzburg
200. Augsburg
201. Memmingen
202. Kaufbeuren
203. Kempten
204. Lindau
205. Berlin
206. Potsdam-Brandenburg
207. Cottbus
208. Frankfurt/Oder
209. Eberswalde
210. Luckenwalde
211. Finsterwalde
212. Oranienburg
213. Neuruppin
214. Perleberg
215. Prenzlau
216. Rostock
217. Schwerin
219. Nordvorpommern
224. Vogtlandkreis
230. Leipzig
232. Halle
236. Burgenlandkreis
238. Mansfeld-Südharz
242. Erfurt
243. Gera
244. Jena
246. Weimar
247. Eisenach
248. Eichsfeld
249. Nordhausen
250. Mühlhausen
251. Sondershausen
252. Meiningen
253. Gotha
254. Arnstadt
256. Saalfeld
257. Pößneck
258. Altenburg
901. Mecklenburgische Seenplatte (218)
   Südvorpommern (220)
902. Chemnitz (221)
   Erzgebirgskreis (222)
   Mittelsachsen (223)
   Zwickau (225)
903. Dresden (226)
   Bautzen (227)
   Görlitz (228)
   Meißen (229)

904. Salzwedel (234)
Stendal (240)

905. Harz (237)
Salzlandkreis (239)

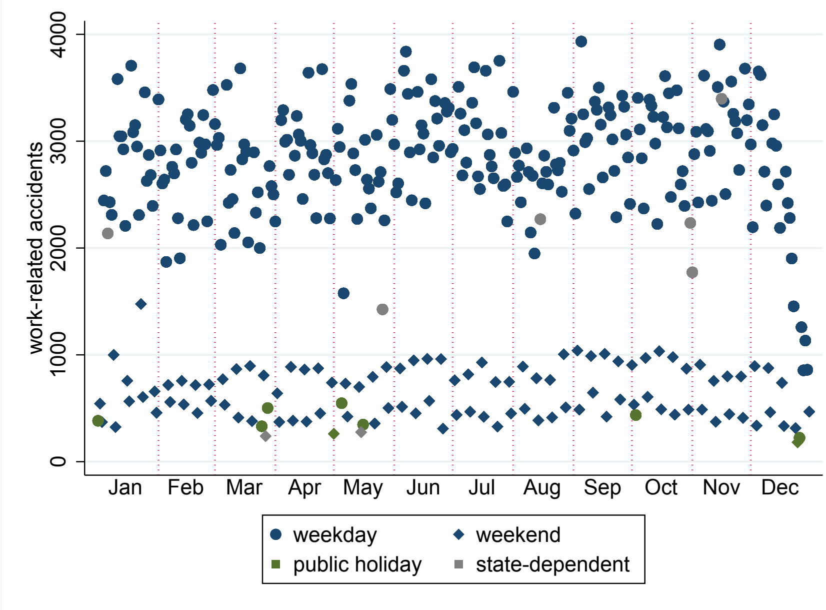
906. Dessau-Roßlau (231)
Magdeburg (233)
Anhalt-Bitterfeld (235)
Wittenberg (241)

999. Suhl (245)
Sonneberg (255)

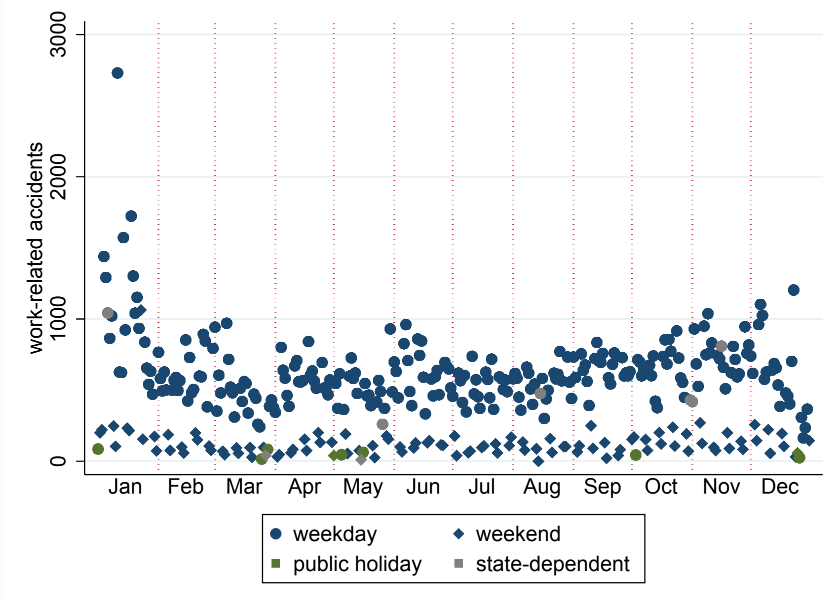
Appendix B

Out of the Dark, into the Light? The Impact of Seasonal Time Changes on Work-Related Accidents

B.1 Additional Figures



(a) Workplace Accidents in 2016



(b) Commuting Accidents in 2016

Figure B.1: Workplace Accidents and Commuting Accidents in 2016

Notes: The figure illustrates the absolute number of workplace accidents and commuting accidents per day over the year 2016. The red dashed lines represent the first day of a month.

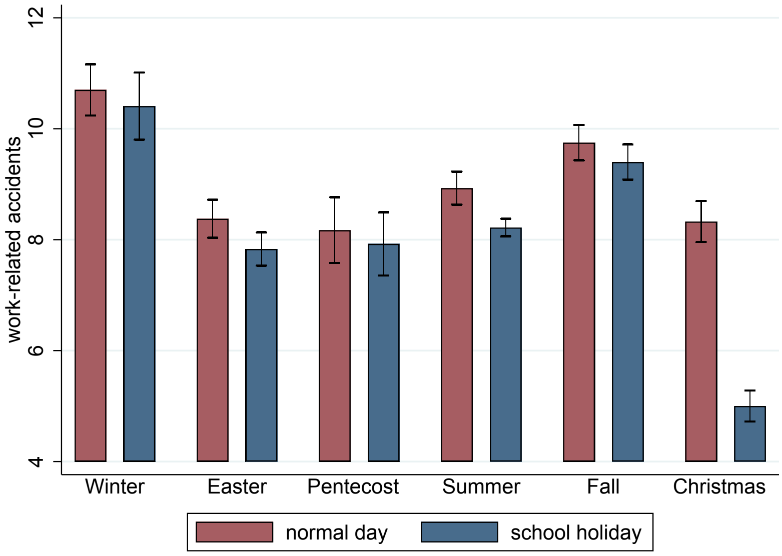
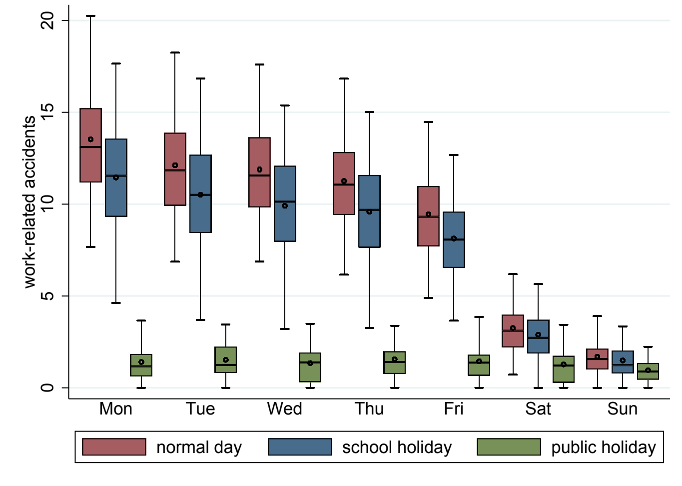
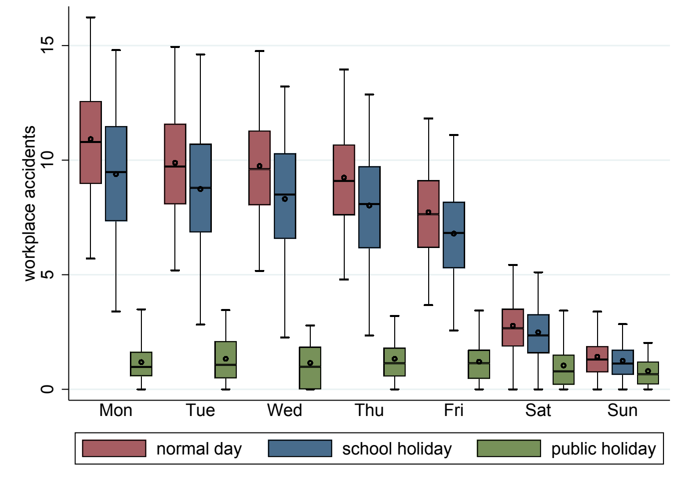


Figure B.2: Work-Related Accidents over Different School Holidays

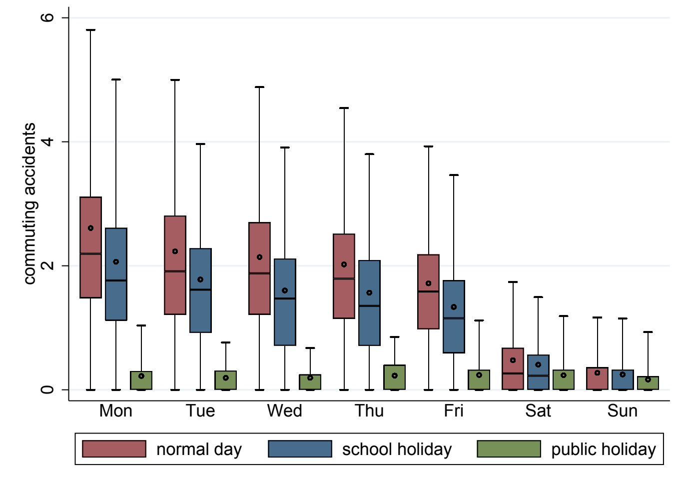
Notes: The figure illustrates the mean number of work-related accidents (per 100,000 employees) for different school holidays and surrounding ‘control weeks’, excluding public holidays. The whiskers indicate 95% confidence intervals. *normal day* refers to days which is not a school holiday but occurs 1 week before or after one. Numbers are weighted by the number of employees subject to social insurance. The figure is inspired by Cox (2009).



(a) Work-Related Accidents



(b) Workplace Accidents



(c) Commuting Accidents

Figure B.3: Distribution of Work-Related Accidents by Type of Day and Day of Week

Notes: The figure illustrates the distribution of the mean work-related accidents (per 100,000 employees) by type of day and day of week. *normal day* refers to days which are neither school holidays nor public holidays. Colored boxes indicate the lower and upper quartiles of the state-day observations. The median is represented by the line subdividing the box. The mean is added by the point symbol. Whiskers represent the 5th and 95th percentile of the distribution. Numbers are weighted by the number of employees subject to social insurance. The figure is inspired by Cox (2009).

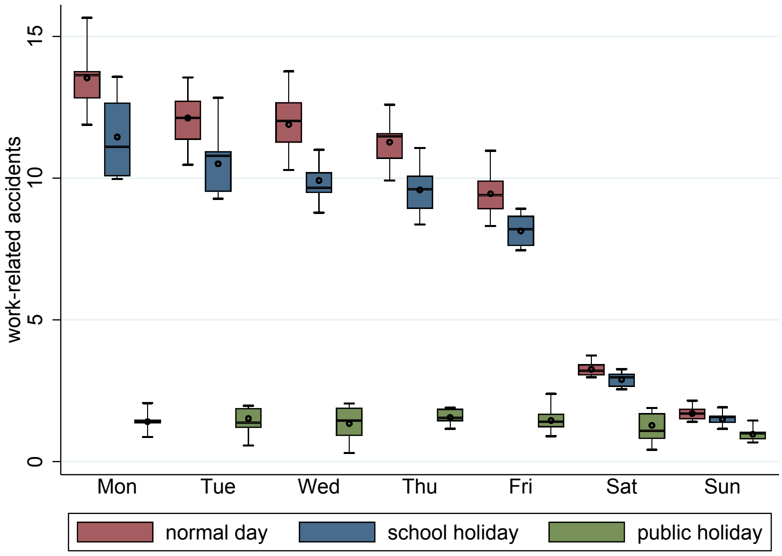


Figure B.4: State Mean Work-Related Accidents by Type of Day and Day of Week

Notes: The figure illustrates the distribution of the mean work-related accidents (per 100,000 employees) by type of day and day of week. *normal day* refers to days which are neither school holidays nor public holidays. The colored boxes indicate the lower and upper quartiles of the state means. The median is represented by the line subdividing the box. The mean is added by the point symbol. The whiskers represent the 5th and 95th percentile of the distribution. Numbers are weighted by the number of employees subject to social insurance. The figure is inspired by Cox (2009).

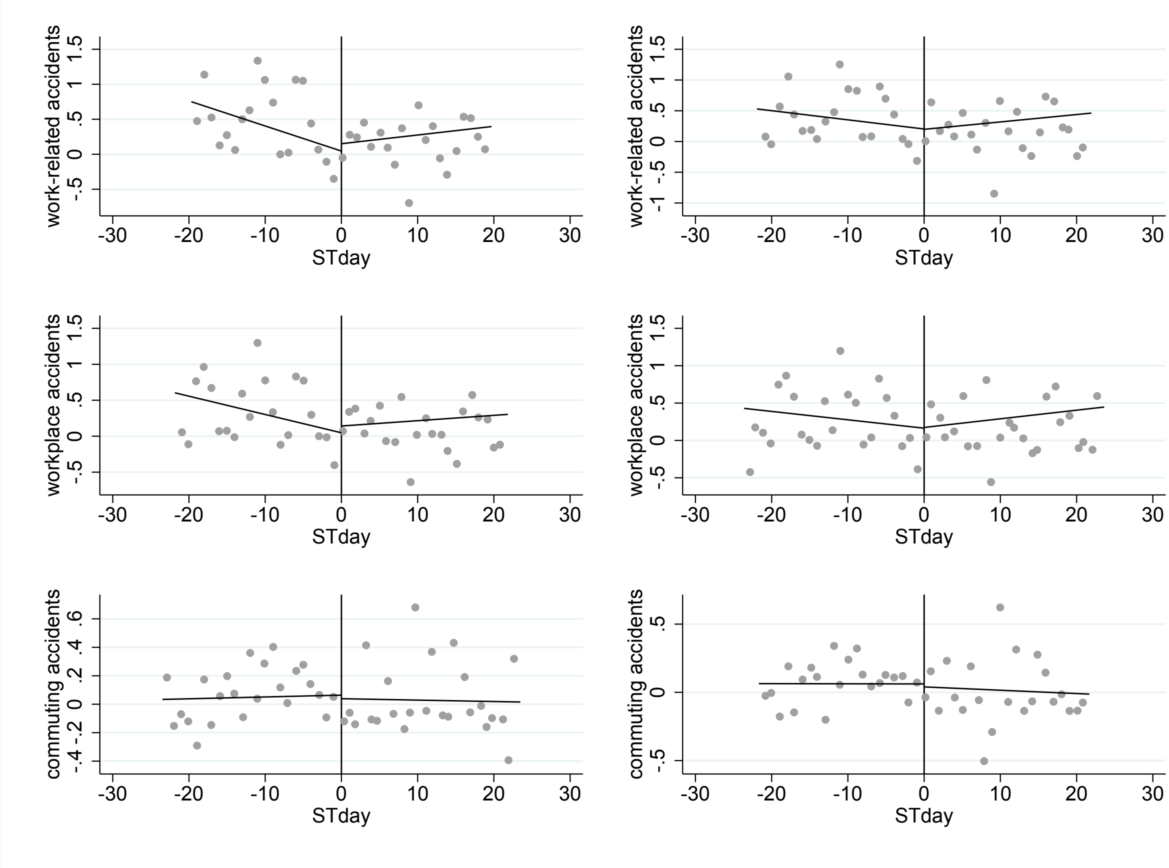


Figure B.5: Impact of the Transition to ST on Work-Related Accidents

Notes: The figure provides RD plots, corresponding to the lower part of Table 2.2. In the plots on the left side, the dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week. In the plots on the right side, the dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week, and weather variables. The points represent the weighted average of residuals per day relative to the transition out of DST. All specifications use the common MSE-optimal bandwidth selector for the RD treatment effect by CCFT, a first-order polynomial, and a triangular kernel.

B.2 Additional Tables

Table B.1: Summary Statistics (Unweighted)

	Mean	Std Dev	min	max	N
Work-related accidents (absolute numbers)					
(all) work-related accidents	163.18	199.54	0.00	2,106.43	29,216
workplace accidents	134.13	165.03	0.00	1,324.03	29,216
commuting accidents	29.05	44.33	0.00	1,257.60	29,216
Labor market variables					
Number of employees (in Mill.)	1.93	1.80	0.30	6.86	29,216
Work-related accidents (per 100,000 employees)					
(all) work-related accidents	8.67	6.19	0.00	152.35	29,216
workplace accidents	7.06	5.08	0.00	51.37	29,216
work-related travel accidents	1.60	2.22	0.00	100.98	29,216
Holiday indicators					
public holiday	0.03	0.17	0.00	1.00	29,216
school holiday	0.23	0.42	0.00	1.00	29,216
christmas holiday	0.03	0.17	0.00	1.00	29,216
day before public holiday	0.03	0.16	0.00	1.00	29,216
day after public holiday	0.03	0.16	0.00	1.00	29,216
Weather variables					
mean of air temperature (in °C)	10.12	7.01	-12.04	29.82	29,216
maximum of air temperature (in °C)	14.37	8.35	-9.39	37.84	29,216
minimum of air temperature (in °C)	5.91	6.12	-17.82	21.51	29,216
min. air ground (at 5 cm) temp. (in °C)	3.97	6.20	-19.85	19.71	29,216
precipitation height (in mm)	1.93	3.68	0.00	117.90	29,216
snow depth (in cm)	0.41	1.52	0.00	23.44	29,216
sunshine duration (in h)	4.44	4.12	0.00	16.01	29,216
maximum of wind gust(in m/s)	10.32	3.61	2.13	32.38	29,216
mean of wind speed (m/s)	3.56	1.51	0.74	13.14	29,216
mean of cloud cover	5.64	1.86	0.00	8.00	29,216
mean of vapor pressure (in hPa)	10.23	4.00	1.98	23.54	29,216
mean of pressure (in hPa)	992.44	20.25	923.50	1,041.28	29,216
mean of relative humidity (in %)	79.09	10.92	34.58	99.89	29,216

Notes: Data sources: German Social Accident Insurance (DGUV), German Federal Employment Agency, standing conference of the ministers of education and cultural affairs of the Länder in the Federal Republic of Germany, schulferien.org, and German Meteorological Service (DWD) Climate Data Center.

Table B.2: Impact of the Transition to ST with Alternative Bandwidth Selectors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All accidents								
ST	-0.1842	-0.1462	0.0656	0.0052	0.0656	-0.0053	0.2010	0.0446
	(0.2093)	(0.1755)	(0.2429)	(0.2248)	(0.2429)	(0.2239)	(0.2442)	(0.2245)
Observations	6800	7360	3600	3280	3600	3440	2640	2960
Dep. var. (mean)	8.445	8.445	8.445	8.445	8.445	8.445	8.445	8.445
Relative effect	-.022	-.017	.008	.001	.008	-.001	.024	.005
Workplace accidents								
ST	-0.0675	-0.0718	0.0904	0.0159	0.0904	0.0101	0.1412	0.0370
	(0.1631)	(0.1554)	(0.2246)	(0.2120)	(0.2246)	(0.2099)	(0.2296)	(0.2142)
Observations	8080	7920	3600	3600	3600	3760	2960	3120
Dep. var. (mean)	6.94	6.94	6.94	6.94	6.94	6.94	6.94	6.94
Relative effect	-.01	-.01	.013	.002	.013	.001	.02	.005
Commuting accidents								
ST	-0.0741	-0.0358	-0.0812	-0.0143	-0.0593	-0.0143	-0.0191	-0.0063
	(0.0763)	(0.0587)	(0.0691)	(0.0575)	(0.0716)	(0.0575)	(0.0746)	(0.0634)
Observations	5120	5440	5360	4240	4720	4240	3120	2960
Dep. var. (mean)	1.505	1.505	1.505	1.505	1.505	1.505	1.505	1.505
Relative effect	-.049	-.024	-.054	-.009	-.039	-.009	-.013	-.004
Bandwidth selector	msetwo	msetwo	msesum	msesum	msec2	msec2	cerrd	cerrd
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State-dow FE	✓	✓	✓	✓	✓	✓	✓	✓
Holiday-dow FE	✓	✓	✓	✓	✓	✓	✓	✓
Weather variables		✓		✓		✓		✓

Notes: Dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week (and weather variables). All specifications use the common MSE-optimal bandwidth selectors by CCFT, a first-order polynomial and a triangular kernel. *msetwo*: different MSE-optimal bandwidth selectors to the two sides of the cutoff. *msesum* one common bandwidth selector which minimizes the asymptotic MSE of the sum of the conditional expectation functions. *msec2:* two separate bandwidths based on median(mserd, msetwo, msesum). *cerrd:* one common coverage error rate (CER) optimal bandwidth selector for the RD estimator ST is the estimate of the discontinuity in work-related accidents at the transition out of DST. Nearest neighbor based standard errors are clustered at the state level (in parentheses). *Relative effect* reports the estimate relative to the sample mean of the dependent variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.3: Impact of the Transition to ST on Work-Related Accidents with respect to Kernel and Transition Date

	baseline		scaling day of transition		excluding day of transition		uniform kernel		epanechnikov kernel	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All accidents										
ST	0.1048	-0.0053	0.0862	-0.0328	0.1901	0.0694	0.1651	0.0677	0.1073	-0.0293
	(0.2440)	(0.2239)	(0.2436)	(0.2226)	(0.2709)	(0.2466)	(0.2755)	(0.2370)	(0.2536)	(0.2294)
Observations	3120	3440	3120	3600	3040	3200	2480	2480	2800	3280
Dep. var. (mean)										
Relative effect	.012	-.001	.01	-.004	.022	.008	.02	.008	.013	-.003
Workplace accidents										
ST	0.0943	0.0101	0.0789	-0.0158	0.1393	0.0625	0.1804	0.0774	0.1037	0.0028
	(0.2257)	(0.2099)	(0.2252)	(0.2088)	(0.2551)	(0.2349)	(0.2553)	(0.2314)	(0.2343)	(0.2165)
Observations	3440	3760	3600	3920	3360	3520	2320	2320	2960	3280
Dep. var. (mean)										
Relative effect	.014	.001	.011	-.002	.02	.009	.026	.011	.015	0
Commuting accidents										
ST	-0.0251	-0.0216	-0.0266	-0.0249	0.0088	-0.0031	-0.0104	-0.0065	-0.0324	-0.0334
	(0.0722)	(0.0618)	(0.0710)	(0.0616)	(0.0787)	(0.0707)	(0.0789)	(0.0753)	(0.0759)	(0.0629)
Observations	3760	3440	3920	3440	3200	3200	2320	2160	3280	3120
Dep. var. (mean)										
Relative effect	-.017	-.014	-.018	-.017	.006	-.002	-.007	-.004	-.022	-.022
Transition day	Yes	Yes	24/25	24/25	No	No	Yes	Yes	Yes	Yes
Kernel	tri	tri	tri	tri	tri	tri	uni	uni	epa	epa
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
State-dow FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Holiday-dow FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Weather variables		✓		✓		✓		✓		✓

Notes: Dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week (and weather variables). All specifications use the common MSE-optimal bandwidth selector for the RD treatment effect by CCFT and a first-order polynomial. *tri* refers to triangular kernel weights, *uni* to uniform kernel weights, and *epa* to epanechnikov kernel weights. 24/23 multiplies the accidents on the day of the transition by 24/23. *No* excludes the day of the transition from the estimation. ST is the estimate of the discontinuity in work-related accidents at the transition out of DST. Nearest neighbor based standard errors are clustered at the state level (in parentheses). *Relative effect* reports the estimate relative to the sample mean of the dependent variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.4: Bias-Corrected RD Estimates with Robust Confidence Intervals: Transition to ST

	All accidents		Workplace accidents		Commuting accidents	
	(1)	(2)	(3)	(4)	(5)	(6)
<u>*Transition to DST*</u>						
ST	0.2377	0.0873	0.1975	0.0877	0.0266	0.0001
	(0.2684)	(0.2581)	(0.2614)	(0.2497)	(0.0763)	(0.0734)
Observations	3120	3440	3440	3760	3760	3440
Dep. var. (mean)	8.445	8.445	6.94	6.94	1.505	1.505
Relative effect	.028	.01	.028	.013	.018	0
Year FE	✓	✓	✓	✓	✓	✓
State-dow FE	✓	✓	✓	✓	✓	✓
Holiday-dow FE	✓	✓	✓	✓	✓	✓
Weather variables		✓		✓		✓

Notes: Dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, holiday-day-of-week (and weather variables). All specifications use the common MSE-optimal bandwidth selector for the RD treatment effect by CCFT, a first-order polynomial, a triangular kernel, and bias-correction (Calonico et al., 2014b). ST is the estimate of the discontinuity in work-related accidents at the transition out of DST. Robust nearest neighbor based standard errors are clustered at the state level (in parentheses). *Relative effect* reports the estimate relative to the sample mean of the dependent variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.5: Additional Robustness: Transition to ST

	baseline		explaining variables				functional form		sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
All accidents									
ST	0.1048	-0.0053	0.0665	0.1838	0.0676	0.1254	0.0993	0.0552	0.0410
	(0.2440)	(0.2239)	(0.2443)	(0.2351)	(0.2405)	(0.2445)	(0.2428)	(0.2375)	(0.2365)
Observations	3120	3440	3440	3280	3120	3280	3120	3440	3760
Dep. var. (mean)	8.445	8.445	8.445	8.445	8.445	8.445	8.445	8.445	8.512
Relative effect	.012	-.001	.008	.022	.008	.015	.012	.007	.005
Workplace accidents									
ST	0.0943	0.0101	0.0633	0.1737	0.0509	0.1190	0.0885	0.0758	0.0711
	(0.2257)	(0.2099)	(0.2279)	(0.2211)	(0.2233)	(0.2258)	(0.2252)	(0.2212)	(0.2241)
Observations	3440	3760	3760	3440	3440	3600	3440	3920	3920
Dep. var. (mean)	6.94	6.94	6.94	6.94	6.94	6.94	6.94	6.94	7.09
Relative effect	.014	.001	.009	.025	.007	.017	.013	.011	.01
Commuting accidents									
ST	-0.0251	-0.0216	0.0102	-0.0080	-0.0208	-0.0103	-0.0286	-0.0366	-0.0414
	(0.0722)	(0.0618)	(0.0699)	(0.0652)	(0.0672)	(0.0723)	(0.0650)	(0.0681)	(0.0658)
Observations	3760	3440	2960	3440	4080	3280	4080	3280	3760
Dep. var. (mean)	1.505	1.505	1.505	1.505	1.505	1.505	1.505	1.505	1.422
Relative effect	-.017	-.014	.007	-.005	-.014	-.007	-.019	-.024	-.029
b/a Pub. hol.-dow FE	✓	✓	✓	✓		✓	✓	✓	✓
Christmas Hol.-dow FE	✓	✓	✓	✓	✓	✓	✓	✓	
Weather variables		✓							
Fewer weather var.			✓						
State-Hol.-dow FE				✓					
Bridge days					✓				
2b/2a Pub. hol.-dow FE						✓			
Elastic net							✓		
Zero-inflated neg. bin.								✓	
w/o Jan. & Dec.									✓

Notes: Dependent variable is the residual number of work-related accidents per 100,000 employees. All specifications are demeaned by year, state-day-of-week, public holiday-day-of-week, and school holiday-day-of-week. Further *b/a Pub. Hol.-dow FE* includes an interaction between indicators for the day before and the day after a public holiday interacted with day-of-week, *Christmas Hol.-dow FE* separate Christmas holiday-day-of-week indicators, *Bridge days* includes indicators for a Monday before a public holiday and a Friday after a public holiday, *2b/2a Pub. hol.-dow FE* includes an interaction between indicators for the second day before and the second day after a public holiday interacted with day-of-week, *weather variables* includes weather measures interacted with state-day-of-week and holiday-day-of-week, *Fewer weather var.* includes weather measures interacted with day-of-week, state, and the five holiday indicators. *Elastic net* uses an elastic net regularization instead of a simple OLS regression and *Zero-inflated neg. bin.* estimates a zero-inflated negative binomial model. *w/o Jan & Dec.* excludes the months January from the estimation sample. All specifications use the common MSE-optimal bandwidth selector for the RD treatment effect by CCFT, a first-order polynomial, and a triangular kernel. ST is the estimate of the discontinuity in work-related accidents at the transition out of DST. Nearest neighbor based standard errors are clustered at the state level (in parentheses). *Relative effect* reports the estimate relative to the sample mean of the dependent variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.6: Robustness of the Variance Estimation: Transition to DST

	(1)	(2)	(3)	(4)	(5)
All accidents					
DST	0.2868	0.2871	0.2868	0.2868	0.2868
	[.331]	[.245]	[.246]	[.324]	[.379]
Observations	3600	3600	3600	3600	3600
Dep. var. (mean)	8.445	8.445	8.445	8.445	8.445
Relative effect	.034	.034	.034	.034	.034
Workplace accidents					
DST	0.1560	0.1531	0.1560	0.1560	0.1560
	[.473]	[.412]	[.276]	[.366]	[.551]
Observations	4240	4080	4240	4240	4240
Dep. var. (mean)	6.94	6.94	6.94	6.94	6.94
Relative effect	.022	.022	.022	.022	.022
Commuting accidents					
DST	0.1376	0.1213	0.1376	0.1376	0.1376
	[.123]	[.123]	[.194]	[.324]	[.238]
Observations	4240	4720	4240	4240	4240
Dep. var. (mean)	1.505	1.505	1.505	1.505	1.505
Relative effect	.091	.081	.091	.091	.091
Year FE	✓	✓	✓	✓	✓
State-dow FE	✓	✓	✓	✓	✓
Holiday-dow FE	✓	✓	✓	✓	✓
Nearest neighbors	✓				
Plugin residuals		✓			
Bootstrapping			✓	✓	
State cluster	✓	✓	✓	✓	✓
Day-of-year cluster				✓	
Permutation test					✓

Notes: Dependent variable is the number of work-related accidents per 100,000 employees demeaned by year, state-day-of-week, and holiday-day-of-week. All specifications use the common MSE-optimal bandwidth selector for the RD treatment effect by CCFT, a first-order polynomial, and a triangular kernel. DST is the estimate of the discontinuity in work-related accidents at the transition into DST. *Relative effect* reports the estimate relative to the sample mean of the dependent variable. All variance estimations account for state level clustering. “Nearest neighbors” indicates cluster-robust nearest neighbor variance estimation, “Plugin residuals” cluster-robust plug-in residuals variance estimation (CCFT). “Bootstrapping” refers to linear regressions, bootstrapped for state clusters (Caskey, Caskey; Cameron et al., 2008, 2011). “Permutation test” re-estimates the baseline specification for all days with respect to the transition in to DST, which neither include the 14 first days of DST, nor a year change in the original bandwidth (cf. Smith (2016), p.77). P-values are provided in brackets.

Appendix C

All Geared Towards Success? Cultural Origins of Gender Gaps in Student Achievement

C.1 Additional Figures

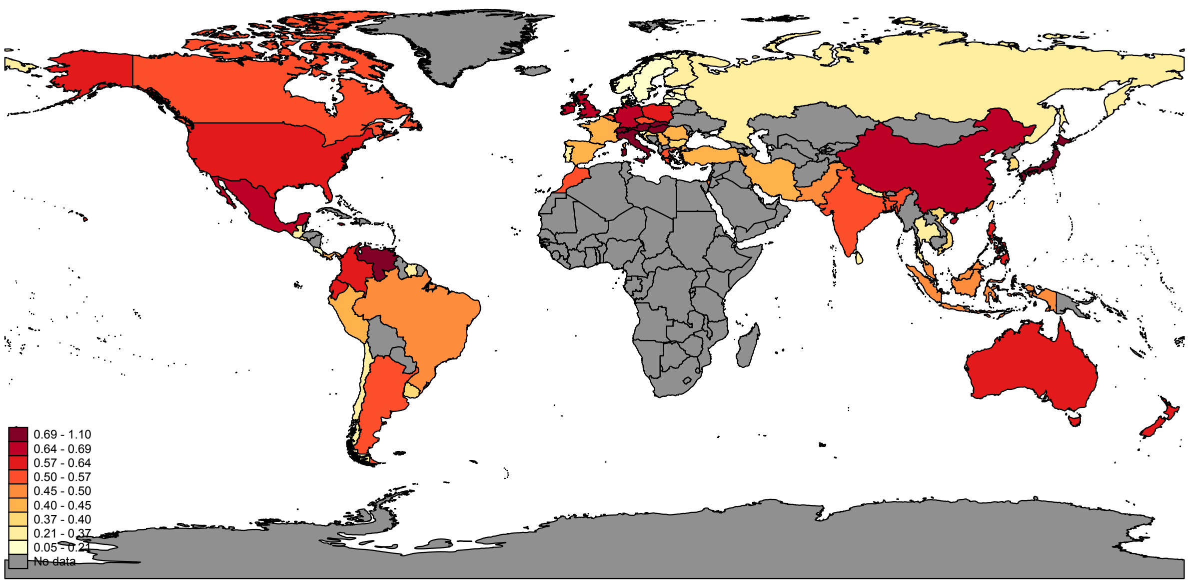


Figure C.1: Distribution of MAS around the World

Notes: This figure maps the distribution of Hofstede et al.'s (2010) cultural dimension MAS around the world. The figure is inspired by Figlio et al. (2019).

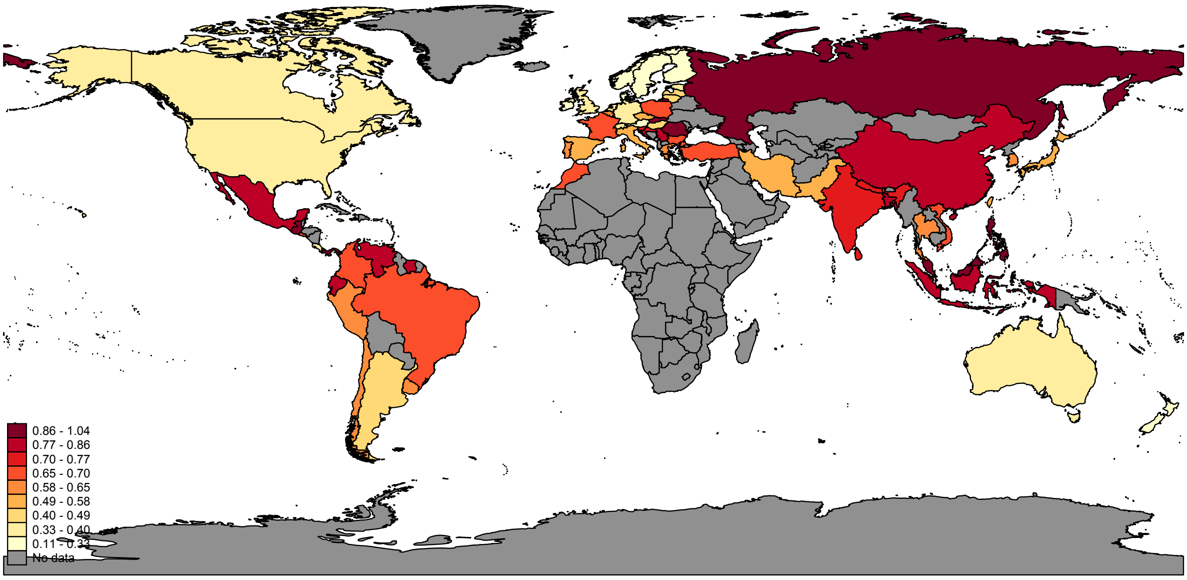


Figure C.2: Distribution of PDI around the World

Notes: This figure maps the distribution of Hofstede et al.'s (2010) cultural dimension PDI around the world. The figure is inspired by Figlio et al. (2019).

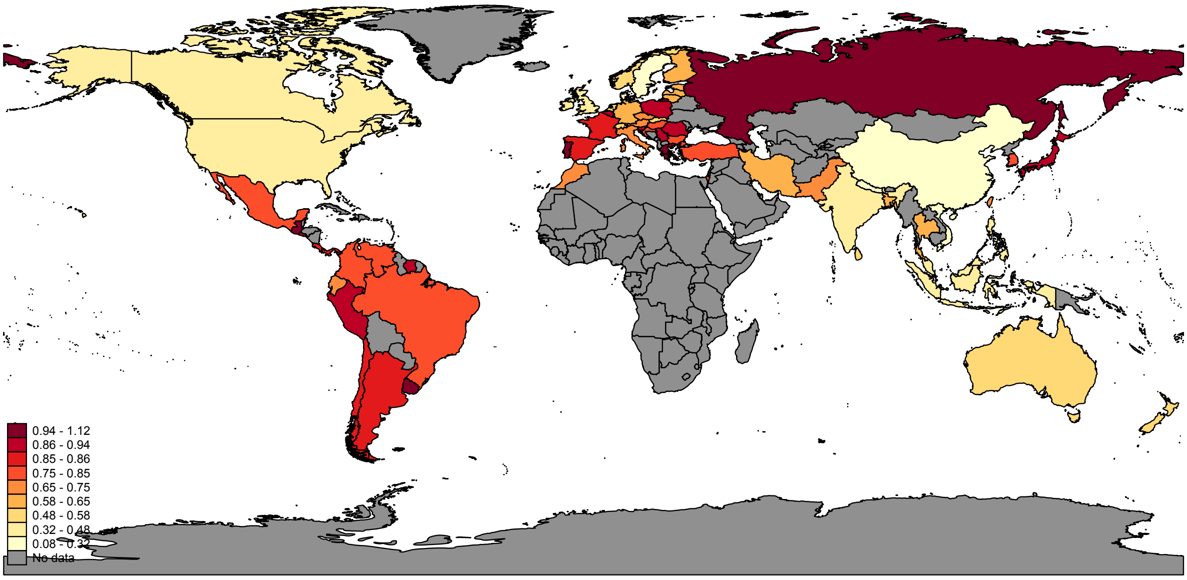


Figure C.3: Distribution of UAI around the World

Notes: This figure maps the distribution of Hofstede et al.'s (2010) cultural dimension UAI around the world. The figure is inspired by Figlio et al. (2019).

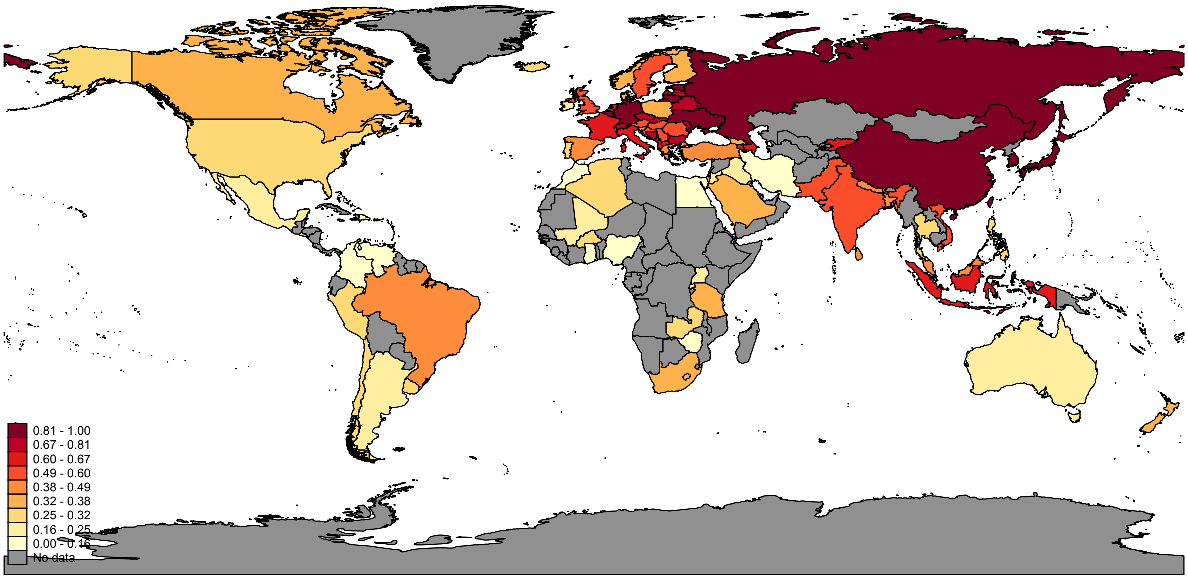


Figure C.4: Distribution of LTO around the World

Notes: This figure maps the distribution of Hofstede et al.'s (2010) cultural dimension LTO around the world. The figure is inspired by Figlio et al. (2019).

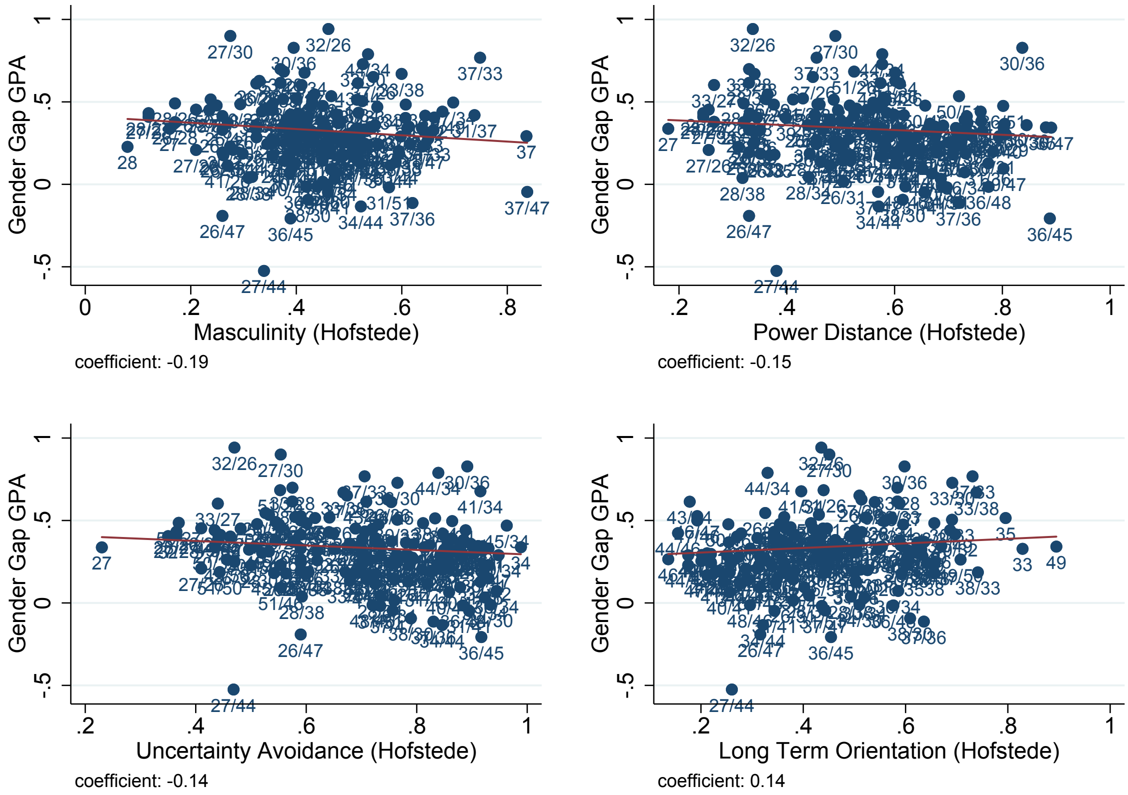


Figure C.5: Hofstede's Cultural Dimensions and Gender GPA Gap

Notes: This figure presents plots of the gender gap in student achievement averaged by second-generation immigrant groups and cultural dimension $C \in \{MAS, PDI, UAI, LTO\}$. For data protection reasons, we only include immigrant groups with at least 50 observations here. The figure is inspired by Figlio et al. (2019).

C.2 Additional Tables

Table C.1: Gender Math Gap and Cultural Dimensions, Sensitivity Checks

	(1)	(2)	(3)	(4)	(5)	(6)
MAS * Female	-0.1775***	-0.1761***	-0.1798**	-0.1912***	-0.1287	-0.1877**
	(0.0560)	(0.0560)	(0.0703)	(0.0588)	(0.0797)	(0.0923)
PDI * Female	-0.0072	-0.0128	-0.0802	-0.0918	-0.1426	-0.0028
	(0.0521)	(0.0521)	(0.1544)	(0.1202)	(0.1699)	(0.1951)
UAI * Female	-0.1275**	-0.1319***	-0.1129**	-0.0886	-0.0691	-0.0689
	(0.0504)	(0.0503)	(0.0562)	(0.0709)	(0.0742)	(0.0844)
LTO * Female	0.1374***	0.1414***	0.1486***	0.1542***	0.1290***	0.1103*
	(0.0414)	(0.0415)	(0.0450)	(0.0469)	(0.0491)	(0.0563)
Observations	78040	78040	78040	78040	77702	73448
R-squared	.636	.638	.636	.636	.643	.682
Dependent var. (mean)	-.008	-.008	-.008	-.008	-.008	-.01
Dependent var. (sd)	.997	.997	.997	.997	.997	.996
Number of clusters	30018	30018	30018	30018	29898	28201
Gender Gap	.015	.015	.015	.015	.015	.011
Family FE	✓	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓	✓
Birth variables		✓			✓	✓
Birth var. * Fem		✓			✓	✓
Individualism * Fem.			✓		✓	✓
Indulgence * Fem.			✓		✓	✓
LogGDPpc2000 * Fem.				✓	✓	✓
Municipality FE					✓	
Mun. FE * Fem.					✓	
Neighborhood FE						✓
Neighb. FE * Fem.						✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. The dependent variable is normalized to be mean 0 and standard deviation 1 relative to the universe of all second-generation immigrant students. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Birth variables include dummies for the month of birth and birth order. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.2: Gender Gap in Swedish and Cultural Dimensions, Sensitivity Checks

	(1)	(2)	(3)	(4)	(5)	(6)
MAS * Female	-0.1833***	-0.1829***	-0.3406***	-0.1655***	-0.2455***	-0.2885***
	(0.0569)	(0.0568)	(0.0707)	(0.0596)	(0.0791)	(0.0928)
PDI * Female	-0.2984***	-0.2958***	-0.0894	-0.1879	-0.0850	-0.0818
	(0.0528)	(0.0527)	(0.1558)	(0.1206)	(0.1713)	(0.1980)
UAI * Female	-0.0970*	-0.1017**	-0.1048*	-0.1480**	-0.0992	-0.0420
	(0.0501)	(0.0500)	(0.0560)	(0.0713)	(0.0744)	(0.0867)
LTO * Female	0.3091***	0.2954***	0.3097***	0.2872***	0.2513***	0.2584***
	(0.0421)	(0.0422)	(0.0454)	(0.0470)	(0.0492)	(0.0567)
Observations	77601	77601	77601	77601	77263	73022
R-squared	.615	.618	.615	.615	.623	.664
Dependent var. (mean)	-.015	-.015	-.015	-.015	-.015	-.017
Dependent var. (sd)	.993	.993	.993	.993	.993	.991
Number of clusters	29865	29865	29865	29865	29745	28051
Gender Gap	.471	.471	.471	.471	.471	.468
Family FE	✓	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓	✓
Birth variables		✓			✓	✓
Birth var. * Fem		✓			✓	✓
Individualism * Fem.			✓		✓	✓
Indulgence * Fem.			✓		✓	✓
LogGDPpc2000 * Fem.				✓	✓	✓
Municipality FE					✓	
Mun. FE * Fem.					✓	
Neighborhood FE						✓
Neighb. FE * Fem.						✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. The dependent variable is normalized to be mean 0 and standard deviation 1 relative to the universe of all second-generation immigrant students. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Birth variables include dummies for the month of birth and birth order. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.3: Gender Gap in Quality and Type of School Attendance, Sensitivity Checks (Municipality Level)

	(1) Res. school quality	(2) Private school
MAS * Female	0.4043	0.0517**
	(0.5268)	(0.0215)
PDI * Female	-0.7672	0.0438
	(1.1201)	(0.0462)
UAI * Female	-0.1968	-0.0530**
	(0.5228)	(0.0211)
LTO * Female	-0.4922	-0.0380**
	(0.3518)	(0.0155)
Observations	57445	57104
R-squared	.624	.768
Dependent var. (mean)	.207	.096
Dependent var. (sd)	6.193	.294
Number of clusters	22726	22607
Gender Gap	.101	.005
Family FE	✓	✓
Grad. year FE	✓	✓
Age	✓	✓
Age * Female	✓	✓
Birth variables	✓	✓
Birth var. * Fem	✓	✓
Individualism * Fem.	✓	✓
Indulgence * Fem.	✓	✓
LogGDPpc2000 * Fem.	✓	✓
Municipality FE		
Mun. FE * Fem.	✓	✓
Nonmover sample	✓	✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. *Nonmover sample* additionally restricts the sample to families who lived in the same neighborhood at graduation of all their children. Residual school quality measures the average peer achievement by school and graduation year, after netting out variation across schools that is explained by children's gender, age, and birth country as well as mothers' and fathers' education, earnings, birth country and immigration age. Private school is a binary variable which indicates whether the student attends a private school in the year of graduation. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Birth variables include dummies for the month of birth and birth order. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.4: Gender Gap in Quality and Type of School Attendance, Sensitivity Checks (Neighborhood Level)

	(1) Res. school quality	(2) Private school
MAS * Female	0.3119	0.0442**
	(0.5353)	(0.0223)
PDI * Female	-1.0356	0.0371
	(1.1411)	(0.0484)
UAI * Female	0.0338	-0.0455**
	(0.5343)	(0.0219)
LTO * Female	-0.3681	-0.0337**
	(0.3550)	(0.0159)
Observations	57692	57348
R-squared	.615	.757
Dependent var. (mean)	.205	.096
Dependent var. (sd)	6.194	.295
Number of clusters	22817	22697
Gender Gap	.101	.005
Family FE	✓	✓
Grad. year FE	✓	✓
Age	✓	✓
Age * Female	✓	✓
Birth variables	✓	✓
Birth var. * Fem	✓	✓
Individualism * Fem.	✓	✓
Indulgence * Fem.	✓	✓
LogGDPpc2000 * Fem.	✓	✓
Neighborhood FE		
Neighb. FE * Fem.	✓	✓
Nonmover sample	✓	✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. *Nonmover sample* additionally restricts the sample to families who lived in the same neighborhood at graduation of all their children. Residual school quality measures the average peer achievement by school and graduation year, after netting out variation across schools that is explained by children's gender, age, and birth country as well as mothers' and fathers' education, earnings, birth country and immigration age. Private school is a binary variable which indicates whether the student attends a private school in the year of graduation. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Birth variables include dummies for the month of birth and birth order. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.5: Mechanisms, Sensitivity Checks (Municipality Level)

	(1) Res. school quality	(2) Private school	(3) Predicted GPA	(4) Par. time in Sweden	(5) Traditional LFP
MAS	1.2613***	0.0331*	-0.0157	-3.1100***	0.0970***
	(0.3169)	(0.0177)	(0.0250)	(0.3194)	(0.0358)
PDI	-1.0990***	-0.1035***	-0.2380***	-12.0561***	0.0314
	(0.3210)	(0.0168)	(0.0223)	(0.2968)	(0.0362)
UAI	-1.0175***	-0.0009	-0.0339*	3.4192***	-0.0631*
	(0.2837)	(0.0153)	(0.0200)	(0.2050)	(0.0326)
LTO	1.7731***	0.0606***	0.2909***	8.3474***	-0.0363
	(0.2453)	(0.0144)	(0.0182)	(0.2116)	(0.0271)
Observations	57540	57316	77822	66260	77822
R-squared	.109	.12	.41	.234	.02
Dependent var. (mean)	.206	.096	-.231	5.615	.058
Dependent var. (sd)	6.193	.294	.468	4.668	.616
Number of clusters	22821	22819	30015	25527	30015
Gender Gap	.072	.003	.349	-.068	.005

	(1) Std. GPA	(2) Std. GPA	(3) Std. GPA	(4) Std. GPA	(5) Std. GPA
Res. sch. qual. * Female	-0.0019*				
	(0.0011)				
Res. school quality	0.0196***				
	(0.0010)				
Priv. school * Female		-0.0781***			
		(0.0217)			
Private school		0.1904***			
		(0.0245)			
Pred. GPA * Female			-0.0300**		
			(0.0147)		
Par. time Swe. * Female				-0.0007	
				(0.0013)	
Trad. LFP * Female					-0.0128
					(0.0091)
Observations	57445	57104	77702	66150	77702
R-squared	.692	.687	.681	.679	.681
Dependent var. (mean)	.039	.039	-.007	.023	-.007
Dependent var. (sd)	.982	.983	.996	.986	.996
Mechanism (mean)	.207	.096	-.231	5.613	.058
Mechanism (sd)	6.193	.294	.468	4.666	.616
Mechanism * Fem. (beta)	-.012	-.023	-.014	-.003	-.008
Number of clusters	22726	22607	29898	25420	29898
Gender Gap	.313	.314	.314	.31	.314
Family FE	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓
Birth variables	✓	✓	✓	✓	✓
Birth var. * Fem	✓	✓	✓	✓	✓
Individualism * Fem.	✓	✓	✓	✓	✓
Indulgence * Fem.	✓	✓	✓	✓	✓
LogGDPpc2000 * Fem.	✓	✓	✓	✓	✓
Municipality FE			✓	✓	✓
Mun. FE * Fem.	✓	✓	✓	✓	✓
Nonmover sample	✓	✓			

Notes: See table notes of Table 3.8. Birth variables include dummies for the month of birth and birth order. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.6: Mechanisms, Sensitivity Checks (Neighborhood Level)

	(1) Res. school quality	(2) Private school	(3) Predicted GPA	(4) Par. time in Sweden	(5) Traditional LFP
MAS	2.4051***	0.1200***	-0.0461*	-2.7014***	0.1152***
	(0.3312)	(0.0175)	(0.0249)	(0.3317)	(0.0397)
PDI	-2.0150***	-0.1229***	-0.1295***	-11.5452***	0.0536
	(0.3255)	(0.0160)	(0.0234)	(0.3170)	(0.0413)
UAI	-0.2746	0.0305**	0.0117	3.2932***	-0.0517
	(0.3025)	(0.0154)	(0.0199)	(0.2192)	(0.0357)
LTO	1.4346***	0.0204	0.1547***	7.5248***	-0.0387
	(0.2609)	(0.0145)	(0.0183)	(0.2231)	(0.0303)
Observations	57697	57472	74672	63385	74672
R-squared	.015	.065	.54	.372	.126
Dependent var. (mean)	.205	.096	-.234	5.568	.055
Dependent var. (sd)	6.195	.295	.466	4.624	.617
Number of clusters	22822	22821	29186	24771	29186
Gender Gap	.072	.003	.353	-.034	.006
	(1) Std. GPA	(2) Std. GPA	(3) Std. GPA	(4) Std. GPA	(5) Std. GPA
Res. sch. qual. * Female	-0.0017				
	(0.0011)				
Res. school quality	0.0199***				
	(0.0010)				
Priv. school * Female		-0.0749***			
		(0.0217)			
Private school		0.2064***			
		(0.0245)			
Pred. GPA * Female			-0.0133		
			(0.0178)		
Par. time Swe. * Female				-0.0001	
				(0.0015)	
Trad. LFP * Female					-0.0073
					(0.0102)
Observations	57692	57348	73448	62256	73448
R-squared	.689	.684	.715	.716	.715
Dependent var. (mean)	.038	.038	-.007	.022	-.007
Dependent var. (sd)	.982	.983	.996	.985	.996
Mechanism (mean)	.205	.096	-.237	5.553	.055
Mechanism (sd)	6.194	.295	.465	4.615	.617
Mechanism * Fem. (beta)	-.011	-.022	-.006	-.001	-.004
Number of clusters	22817	22697	28201	23880	28201
Gender Gap	.313	.314	.313	.309	.313
Family FE	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓
Birth variables	✓	✓	✓	✓	✓
Birth var. * Fem	✓	✓	✓	✓	✓
Individualism * Fem.	✓	✓	✓	✓	✓
Indulgence * Fem.	✓	✓	✓	✓	✓
LogGDPpc2000 * Fem.	✓	✓	✓	✓	✓
Neighborhood FE			✓	✓	✓
Neighb. FE * Fem.	✓	✓	✓	✓	✓
Nonmover sample	✓	✓			

Notes: See table notes of Table 3.8. Birth variables include dummies for the month of birth and birth order. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.7: Mechanisms as Control Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MAS * Female	-0.1840***	-0.1818***	-0.1759***	-0.1805***	-0.1413***	-0.1837***	-0.1360**
	(0.0503)	(0.0500)	(0.0506)	(0.0504)	(0.0545)	(0.0503)	(0.0545)
PDI * Female	-0.0523	-0.0452	-0.0490	-0.0625	-0.0944*	-0.0516	-0.0846*
	(0.0463)	(0.0460)	(0.0466)	(0.0466)	(0.0510)	(0.0463)	(0.0512)
UAI * Female	-0.0560	-0.0530	-0.0615	-0.0593	-0.0491	-0.0568	-0.0571
	(0.0427)	(0.0424)	(0.0428)	(0.0426)	(0.0448)	(0.0427)	(0.0446)
LTO * Female	0.1123***	0.1156***	0.1106***	0.1224***	0.1204***	0.1118***	0.1234***
	(0.0376)	(0.0373)	(0.0377)	(0.0377)	(0.0410)	(0.0376)	(0.0409)
Res. sch. qual. * Female		-0.0012					-0.0000
		(0.0010)					(0.0011)
Res. school quality		0.0203***					0.0191***
		(0.0008)					(0.0009)
Priv. school * Female			-0.0619***				-0.0650***
			(0.0191)				(0.0214)
Private school			0.1970***				0.1055***
			(0.0209)				(0.0224)
Pred. GPA * Female				-0.0332**			-0.0445***
				(0.0135)			(0.0148)
Par. time Swe. * Female					-0.0013		-0.0004
					(0.0013)		(0.0013)
Trad. LFP * Female						-0.0108	-0.0123
						(0.0090)	(0.0096)
Observations	78040	78020	77550	78040	66454	78040	66016
R-squared	.674	.68	.675	.674	.672	.674	.679
Dependent var. (mean)	-.007	-.007	-.007	-.007	.023	-.007	.024
Dependent var. (sd)	.996	.996	.996	.996	.986	.996	.985
Number of clusters	30018	30013	29861	30018	25530	30018	25389
Gender Gap	.313	.312	.313	.313	.309	.313	.313
Family FE	✓	✓	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓	✓	✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. The dependent variable is normalized to be mean 0 and standard deviation 1 relative to the universe of all second-generation immigrant students. Residual school quality measures the average peer achievement by school and graduation year, after netting out variation across schools that is explained by children's gender, age, and birth country as well as mothers' and fathers' education, earnings, birth country and immigration age. Private school is a binary variable which indicates whether the student attends a private school in the year of graduation. Predicted GPA is obtained by regressing GPA on parents' education and earnings, age, a female indicator and graduation year dummies. Parental time in Sweden captures the host country experience prior to the birth of the oldest sibling. Traditional LFP takes 1 if only the father is working, 0 if both/none work, and -1 if only the mother is working. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.8: Mechanisms as Control Variables, Sensitivity Checks (Municipality Level)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MAS * Female	-0.1865***	-0.2009***	-0.1919***	-0.1924***	-0.1319*	-0.1840***	-0.1668**
	(0.0693)	(0.0687)	(0.0695)	(0.0694)	(0.0751)	(0.0693)	(0.0747)
PDI * Female	-0.1495	-0.1384	-0.1496	-0.1188	-0.1997	-0.1518	-0.1567
	(0.1480)	(0.1467)	(0.1481)	(0.1486)	(0.1600)	(0.1480)	(0.1596)
UAI * Female	-0.0153	-0.0063	-0.0149	-0.0247	0.0035	-0.0160	-0.0017
	(0.0634)	(0.0626)	(0.0634)	(0.0635)	(0.0676)	(0.0633)	(0.0672)
LTO * Female	0.1060**	0.1156***	0.1085**	0.1098**	0.1147**	0.1056**	0.1257***
	(0.0430)	(0.0426)	(0.0431)	(0.0430)	(0.0464)	(0.0430)	(0.0461)
Res. sch. qual. * Female		-0.0016					-0.0004
		(0.0010)					(0.0012)
Res. school quality		0.0201***					0.0190***
		(0.0009)					(0.0010)
Priv. school * Female			-0.0538***				-0.0555**
			(0.0197)				(0.0220)
Private school			0.1818***				0.0951***
			(0.0211)				(0.0226)
Pred. GPA * Female				-0.0327**			-0.0460***
				(0.0148)			(0.0160)
Par. time Swe. * Female					-0.0013		-0.0006
					(0.0013)		(0.0013)
Trad. LFP * Female						-0.0125	-0.0126
						(0.0091)	(0.0097)
Observations	77702	77682	77216	77702	66150	77702	65717
R-squared	.681	.687	.682	.681	.68	.681	.686
Dependent var. (mean)	-.007	-.007	-.007	-.007	.023	-.007	.024
Dependent var. (sd)	.996	.996	.996	.996	.986	.996	.985
Number of clusters	29898	29893	29742	29898	25420	29898	25281
Gender Gap	.314	.313	.314	.314	.31	.314	.314
Family FE	✓	✓	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓	✓	✓
Birth variables	✓	✓	✓	✓	✓	✓	✓
Birth var. * Fem	✓	✓	✓	✓	✓	✓	✓
Individualism * Fem.	✓	✓	✓	✓	✓	✓	✓
Indulgence * Fem.	✓	✓	✓	✓	✓	✓	✓
LogGDPpc2000 * Fem.	✓	✓	✓	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓	✓	✓	✓
Mun. FE * Fem.	✓	✓	✓	✓	✓	✓	✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. The dependent variable is normalized to be mean 0 and standard deviation 1 relative to the universe of all second-generation immigrant students. Residual school quality measures the average peer achievement by school and graduation year, after netting out variation across schools that is explained by children's gender, age, and birth country as well as mothers' and fathers' education, earnings, birth country and immigration age. Private school is a binary variable which indicates whether the student attends a private school in the year of graduation. Predicted GPA is obtained by regressing GPA on parents' education and earnings, age, a female indicator and graduation year dummies. Parental time in Sweden captures the host country experience prior to the birth of the oldest sibling. Traditional LFP takes 1 if only the father is working, 0 if both/none work, and -1 if only the mother is working. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Birth variables include dummies for the month of birth and birth order. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.9: Mechanisms as Control Variables, Sensitivity Checks (Neighborhood Level)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MAS * Female	-0.2090**	-0.2266***	-0.2126***	-0.2126***	-0.1076	-0.2075**	-0.1436
	(0.0816)	(0.0807)	(0.0820)	(0.0817)	(0.0899)	(0.0817)	(0.0895)
PDI * Female	0.0026	0.0212	0.0167	0.0128	-0.0812	0.0012	-0.0276
	(0.1742)	(0.1725)	(0.1747)	(0.1747)	(0.1917)	(0.1742)	(0.1910)
UAI * Female	-0.0320	-0.0253	-0.0311	-0.0338	-0.0059	-0.0322	-0.0070
	(0.0746)	(0.0737)	(0.0747)	(0.0746)	(0.0815)	(0.0746)	(0.0809)
LTO * Female	0.0943*	0.0972*	0.0939*	0.0951*	0.1187**	0.0941*	0.1154**
	(0.0504)	(0.0499)	(0.0506)	(0.0504)	(0.0550)	(0.0504)	(0.0548)
Res. sch. qual. * Female		-0.0006					0.0005
		(0.0012)					(0.0014)
Res. school quality		0.0195***					0.0187***
		(0.0010)					(0.0011)
Priv. school * Female			-0.0599***				-0.0736***
			(0.0227)				(0.0259)
Private school			0.1831***				0.0971***
			(0.0232)				(0.0252)
Pred. GPA * Female				-0.0162			-0.0379*
				(0.0178)			(0.0195)
Par. time Swe. * Female					-0.0006		0.0000
					(0.0015)		(0.0015)
Trad. LFP * Female						-0.0066	-0.0057
						(0.0102)	(0.0109)
Observations	73448	73426	72973	73448	62256	73448	61820
R-squared	.715	.721	.716	.715	.716	.715	.722
Dependent var. (mean)	-.007	-.007	-.007	-.007	.022	-.007	.022
Dependent var. (sd)	.996	.996	.996	.996	.985	.996	.985
Number of clusters	28201	28195	28047	28201	23880	28201	23738
Gender Gap	.313	.311	.312	.313	.309	.313	.311
Family FE	✓	✓	✓	✓	✓	✓	✓
Grad. year FE	✓	✓	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓	✓	✓
Age * Female	✓	✓	✓	✓	✓	✓	✓
Birth variables	✓	✓	✓	✓	✓	✓	✓
Birth var. * Fem	✓	✓	✓	✓	✓	✓	✓
Individualism * Fem.	✓	✓	✓	✓	✓	✓	✓
Indulgence * Fem.	✓	✓	✓	✓	✓	✓	✓
LogGDPpc2000 * Fem.	✓	✓	✓	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓	✓	✓
Neighb. FE * Fem.	✓	✓	✓	✓	✓	✓	✓

Notes: The table reports estimates of equation (3.1) on a sample of second-generation immigrant students with opposite-sex siblings. The dependent variable is normalized to be mean 0 and standard deviation 1 relative to the universe of all second-generation immigrant students. Residual school quality measures the average peer achievement by school and graduation year, after netting out variation across schools that is explained by children's gender, age, and birth country as well as mothers' and fathers' education, earnings, birth country and immigration age. Private school is a binary variable which indicates whether the student attends a private school in the year of graduation. Predicted GPA is obtained by regressing GPA on parents' education and earnings, age, a female indicator and graduation year dummies. Parental time in Sweden captures the host country experience prior to the birth of the oldest sibling. Traditional LFP takes 1 if only the father is working, 0 if both/none work, and -1 if only the mother is working. All regressions include the female dummy (non-reported). Age is captured as the difference between the year of graduation and the year of birth. Birth variables include dummies for the month of birth and birth order. Standard errors are adjusted for clustering at the family level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.10: Gender GPA Gap and Cultural Dimensions, Sensitivity Checks for PISA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MAS * Female	-0.2921**	-0.3084**	-0.2231*	-0.2773*	-0.2231*		-0.2644**
	(0.1267)	(0.1355)	(0.1160)	(0.1411)	(0.1197)		(0.1261)
PDI *Female	-0.1717*	-0.1946*	-0.1901*	-0.2037	-0.2323		-0.1896**
	(0.0893)	(0.0978)	(0.1093)	(0.1311)	(0.1402)		(0.0892)
UAI * Female	-0.0819	-0.0634	-0.0769	-0.0632	-0.0517		-0.0668
	(0.1033)	(0.1138)	(0.1038)	(0.1240)	(0.1350)		(0.1048)
LTO * Female	-0.0046	-0.0609	0.0040	-0.0046	-0.0532		-0.0350
	(0.0934)	(0.1042)	(0.0953)	(0.0938)	(0.1051)		(0.0921)
MAS (moth.) *Female						-0.2315*	
						(0.1329)	
PDI (moth.) * Female						-0.1705*	
						(0.1004)	
UAI (moth.) * Female						-0.1109	
						(0.1160)	
LTO (moth.) *Female						-0.0281	
						(0.1012)	
Observations	35347	35347	35347	35347	35347	33750	35754
R-squared	.399	.418	.399	.399	.418	.408	.398
Dependent var. (mean)	.001	.001	.001	.001	.001	-.005	.001
Dependent var. (sd)	.965	.965	.965	.965	.965	.965	.965
Number of Clusters	73	73	73	73	73	40/50	72
Gender Gap	-.011	-.008	-.011	-.011	-.008	-.01	-.012
Year FE	✓	✓	✓	✓	✓	✓	✓
Grade FE	✓	✓	✓	✓	✓	✓	✓
Anc. Country FE	✓	✓	✓	✓	✓	✓	✓
Host Country FE	✓	✓	✓	✓	✓	✓	✓
Host Country FE * Fem.	✓	✓	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓	✓	✓
Age * Fem.	✓	✓	✓	✓	✓	✓	✓
Parental Variables		✓			✓		
Parental Variables * Fem.		✓			✓		
Individualism * Fem.			✓		✓		
Indulgence * Fem.			✓		✓		
LogGDPpc2000 * Fem.				✓	✓		
no diff. betw. m/f & f/m							✓

Notes: The table reports estimates of equation (3.2) on a sample of second-generation immigrant students tested in PISA studies 2003, 2006, 2009, 2012, and 2015. The dependent variable is a student's PISA grade-point average, computed as the average normalized test score of mathematics, science, and reading. Each subject score is normalized to be mean 0 and standard deviation 1 in our estimation sample. All regressions include the female dummy (non-reported). Parental variables include parents' education. Column 6 assigns cultural dimensions accoding to the mother's country of ancestry. Standard errors are adjusted for clustering at parents' country-of-origin level (combining mother's and father's origin and distinguishing between the two in columns 1-6). Column 7 relaxes the distinguishment between these combinations (treating m/f like f/m). ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.11: Gender Math Gap and Cultural Dimensions, PISA Data

Dependent Variable:	*Standardized PISA Test Score in Mathematics*				
	(1)	(2)	(3)	(4)	(5)
MAS * Female	-0.1849**				-0.2910**
	(0.0918)				(0.1162)
PDI * Female		-0.0986			-0.1608*
		(0.0934)			(0.0916)
UAI * Female			0.0126		-0.0835
			(0.0811)		(0.0967)
LTO * Female				-0.0533	0.0248
				(0.0693)	(0.0891)
Observations	35512	35512	35512	35347	35347
R-squared	.399	.399	.399	.4	.4
Dependent var. (mean)	0	0	0	.001	.001
Dependent var. (sd)	1	1	1	1.001	1.001
Cultural var. (mean)	.563	.700	.554	.664	
Cultural var. (sd)	.137	.154	.292	.23	
Cultural var. * Fem. (beta)	-.025	-.015	.004	-.012	
Number of Clusters	74	74	74	73	73
Gender Gap	-.195	-.195	-.195	-.195	-.195
Year FE	✓	✓	✓	✓	✓
Grade FE	✓	✓	✓	✓	✓
Anc. Country FE	✓	✓	✓	✓	✓
Host Country FE	✓	✓	✓	✓	✓
Host Country FE * Fem.	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Age * Fem.	✓	✓	✓	✓	✓

Notes: The table reports estimates of equation (3.2) on a sample of second-generation immigrant students tested in PISA studies 2003, 2006, 2009, 2012, and 2015. Each subject score is normalized to be mean 0 and standard deviation 1 in our estimation sample. All regressions include the female dummy (non-reported). Standard errors are adjusted for clustering at parents' country-of-origin level (combining mother's and father's origin and distinguishing between the two). ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.12: Gender Science Gap and Cultural Dimensions, PISA Data

Dependent Variable:	*Standardized PISA Test Score in Science*				
	(1)	(2)	(3)	(4)	(5)
MAS * Female	-0.1997**				-0.3058**
	(0.0929)				(0.1294)
PDI * Female		-0.0966			-0.1454
		(0.0852)			(0.0888)
UAI * Female			-0.0129		-0.1515
			(0.0763)		(0.1091)
LTO * Female				-0.0801	-0.0228
				(0.0618)	(0.0889)
Observations	35512	35512	35512	35347	35347
R-squared	.386	.386	.386	.388	.388
Dependent var. (mean)	0	0	0	.001	.001
Dependent var. (sd)	1	1	1	1	1
Cultural var. (mean)	.563	.700	.554	.664	
Cultural var. (sd)	.137	.154	.292	.23	
Cultural var. * Fem. (beta)	-.027	-.015	-.004	-.018	
Number of Clusters	74	74	74	73	73
Gender Gap	-.103	-.103	-.103	-.103	-.103
Year FE	✓	✓	✓	✓	✓
Grade FE	✓	✓	✓	✓	✓
Anc. Country FE	✓	✓	✓	✓	✓
Host Country FE	✓	✓	✓	✓	✓
Host Country FE * Fem.	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Age * Fem.	✓	✓	✓	✓	✓

Notes: The table reports estimates of equation (3.2) on a sample of second-generation immigrant students tested in PISA studies 2003, 2006, 2009, 2012, and 2015. Each subject score is normalized to be mean 0 and standard deviation 1 in our estimation sample. All regressions include the female dummy (non-reported). Standard errors are adjusted for clustering at parents' country-of-origin level (combining mother's and father's origin and distinguishing between the two). ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table C.13: Gender Reading Gap and Cultural Dimensions, PISA Data

Dependent Variable:	*Standardized PISA Test Score in Reading*				
	(1)	(2)	(3)	(4)	(5)
MAS * Female	-0.2297**				-0.2796*
	(0.0927)				(0.1421)
PDI * Female		-0.1320			-0.2089**
		(0.1062)			(0.0984)
UAI * Female			0.0981		-0.0107
			(0.0803)		(0.1185)
LTO * Female				-0.1193	-0.0157
				(0.0772)	(0.1090)
Observations	35512	35512	35512	35347	35347
R-squared	.37	.37	.37	.371	.371
Dependent var. (mean)	0	0	0	0	0
Dependent var. (sd)	1	1	1	1	1
Cultural var. (mean)	.563	.700	.554	.664	
Cultural var. (sd)	.137	.154	.292	.23	
Cultural var. * Fem. (beta)	-.031	-.02	.029	-.027	
Number of Clusters	74	74	74	73	73
Gender Gap	.266	.266	.266	.265	.265
Year FE	✓	✓	✓	✓	✓
Grade FE	✓	✓	✓	✓	✓
Anc. Country FE	✓	✓	✓	✓	✓
Host Country FE	✓	✓	✓	✓	✓
Host Country FE * Fem.	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Age * Fem.	✓	✓	✓	✓	✓

Notes: The table reports estimates of equation (3.2) on a sample of second-generation immigrant students tested in PISA studies 2003, 2006, 2009, 2012, and 2015. Each subject score is normalized to be mean 0 and standard deviation 1 in our estimation sample. All regressions include the female dummy (non-reported). Standard errors are adjusted for clustering at parents' country-of-origin level (combining mother's and father's origin and distinguishing between the two). ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Bibliography

Aizer, A., L. Stroud, and S. Buka (2016). Maternal Stress and Child Outcomes: Evidence from Siblings. *Journal of Human Resources 51*(3), 523–555.

Algan, Y., C. Dustmann, A. Glitz, and A. Manning (2010). The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom. *Economic Journal 120*(542), F4–F30.

Almond, D. and J. Currie (2011a). Human Capital Development Before Age Five. *Handbook of Labor Economics 4B*, 1315–1486.

Almond, D. and J. Currie (2011b). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives 25*(3), 153–72.

Almond, D., J. Currie, and V. Duque (2018). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature 56*(4), 1360–1446.

Angrist, J., P. Azoulay, G. Ellison, R. Hill, and S. F. Lu (2017). Economic Research Evolves: Fields and Styles. *American Economic Review, Papers & Proceedings 107*(5), 293–97.

Aparicio, A., L. González, and J. V. Castelló (2020). Newborn Health and the Business Cycle: The Role of Birth Order. *Economics & Human Biology 37*, 100836.

Aries, M. B. and G. R. Newsham (2008). Effect of Daylight Saving Time on Lighting Energy Use: A Literature Review. *Energy Policy 36*(6), 1858–1866.

Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2016). School Quality and the Gender Gap in Educational Achievement. *American Economic Review, Papers & Proceedings 106*(5), 289–95.

Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2019). Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes. *American Economic Journal: Applied Economics 11*(3), 338–81.

Aydemir, A., W.-H. Chen, and M. Corak (2009). Intergenerational Earnings Mobility Among the Children of Canadian Immigrants. *Review of Economics and Statistics 91*(2), 377–397.

Barnes, C. M. and D. T. Wagner (2009). Changing to Daylight Saving Time Cuts into Sleep and Increases Workplace Injuries. *Journal of Applied Psychology 94*(5), 1305.

Berk, M., S. Dodd, K. Hallam, L. Berk, J. Gleeson, and M. Henry (2008). Small Shifts in Diurnal Rhythms Are Associated with an Increase in Suicide: The Effect of Daylight Saving. *Sleep and Biological Rhythms 6*(1), 22–25.

Bertrand, M. (2018). Coase Lecture–The Glass Ceiling. *Economica 85*(338), 205–231.

Bertrand, M., S. E. Black, S. Jensen, and A. Lleras-Muney (2019). Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway. *Review of Economic Studies 86*(1), 191–239.

Bertrand, M. and J. Pan (2013). The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. *American Economic Journal: Applied Economics 5*(1), 32–64.

Bhalotra, S. (2010). Fatal Fluctuations? Cyclicality in Infant Mortality in India. *Journal of Development Economics 93*(1), 7–19.

Bleakley, H. and A. Chin (2008). What Holds Back the Second Generation? The Intergenerational Transmission of Language Human Capital Among Immigrants. *Journal of Human Resources 43*(2), 267–298.

Bünnings, C. and V. Schiele (0). Spring Forward, Don't Fall Back: The Effect of Daylight Saving Time on Road Safety. *Review of Economics and Statistics 0*(forthcoming), 1–45.

Bozzoli, C. and C. Quintana-Domeque (2014). The Weight of the Crisis: Evidence from Newborns in Argentina. *Review of Economics and Statistics 96*(3), 550–562.

Bundesregierung (1977). Entwurf eines Gesetzes über die Zeitbestimmung (Zeitgesetz — ZeitG). In *Deutscher Bundestag, 8. Wahlperiode*, Number 258 in 8, pp. 1–8. Bonner Universitäts-Buchdruckerei, 5300 Bonn.

Burlando, A. (2014). Transitory Shocks and Birth Weights: Evidence from a Blackout in Zanzibar. *Journal of Development Economics 108*, 154–168.

Callaghan, W. M., M. F. MacDorman, S. A. Rasmussen, C. Qin, and E. M. Lackritz (2006). The Contribution of Preterm Birth to Infant Mortality Rates in the United States. *Pediatrics 118*(4), 1566–1573.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2018). On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association 113*(522), 767–779.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2020). Optimal Bandwidth Choice for Robust Bias-Corrected Inference in Regression Discontinuity Designs. *Econometrics Journal 23*(2), 192–210.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for Regression Discontinuity Designs. *Stata Journal 17*(2), 372–404.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019). Regression Discontinuity Designs Using Covariates. *Review of Economics and Statistics 101*(3), 442–451.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a). Robust Data-Driven Inference in the Regression-Discontinuity Design. *Stata Journal 14*(4), 909–946.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014b). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica 82*(6), 2295–2326.

Camacho, A. (2008). Stress and Birth Weight: Evidence from Terrorist Attacks. *American Economic Review, Papers & Proceedings 98*(2), 511–15.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics 90*(3), 414–427.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust Inference with Multiway Clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Card, D., J. DiNardo, and E. Estes (2000, January). The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s. In G. J. Borjas (Ed.), *Issues in the Economics of Immigration*, Issues in the Economics of Immigration, pp. 227–270. National Bureau of Economic research.

Carlson, K. (2015). Fear Itself: The Dffects of Distressing Economic News on Birth Outcomes. *Journal of Health Economics 41*, 117–132.

Carroll, C. D., B.-K. Rhee, and C. Rhee (1994). Are There Cultural Effects on Saving? Some Cross-Sectional Evidence. *Quarterly Journal of Economics 109*(3), 685–699.

Caskey, J. User written STATA do file: cgmwildboot.ado.

Chadi, A. (2014). Regional Unemployment and Norm-Induced Effects on Life Satisfaction. *Empirical Economics 46*(3), 1111–1141.

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. *Quarterly Journal of Economics 118*(3), 1121–1167.

Chiswick, B. R. (1977). Sons of Immigrants: Are They at an Earnings Disadvantage? *American Economic Review, Papers & Proceedings 67*(1), 376–380.

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. (2003). Unemployment as a Social Norm: Psychological Evidence from Panel Data. *Journal of Labor Economics 21*(2), 323–351.

Coren, S. (1996). Daylight Savings Time and Traffic Accidents. *New England Journal of Medicine 334*(14), 924.

Cox, N. J. (2009). Speaking Stata: Creating and Varying Box Plots. *Stata Journal 9*(3), 478–496.

Currie, J. and M. Rossin-Slater (2013). Weathering the Storm: Hurricanes and Birth Outcomes. *Journal of Health Economics 32*(3), 487 – 503.

Currie, J. and E. Tekin (2015). Is There a Link Between Foreclosure and Health? *American Economic Journal: Economic Policy 7*(1), 63–94.

Currie, J., J. G. Zivin, J. Mullins, and M. Neidell (2014). What Do We Know About Short- and Long-Term Effects of Early-Life Exposure to Pollution? *Annual Review of Resource Economics 6*(1), 217–247.

Dahl, G. B., C. Felfe, P. Frijters, and H. Rainer (2020, January). Caught Between Cultures: Unintended Consequences of Improving Opportunity for Immigrant Girls. NBER Working Paper 26674, National Bureau of Economic Research.

De Cao, E., B. McCormick, and C. Nicodemo (2019). Does Unemployment Worsen Babies' Health? A Tale of Siblings, Maternal Behaviour and Selection. IZA Discussion Paper 12568, Institute of Labor Economics (IZA).

Dee, T. S. (2005). A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? *American Economic Review, Papers & Proceedings 95*(2), 158–165.

Dehejia, R. and A. Lleras-Muney (2004). Booms, Busts, and Babies' Health. *Quarterly Journal of Economics 119*(3), 1091–1130.

Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014). School Choice, School Quality, and Postsecondary Attainment. *American Economic Review 104*(3), 991–1013.

Deutsche Gesetzliche Unfallversicherung e.V. (DGUV) (2017a). Arbeitsunfallgeschehen 2016. November. `https://publikationen.dguv.de/zahlen-fakten/schwerpunkt-themen/3382/arbeitsunfallgeschehen-2016`, accessed on February 20, 2020.

Deutsche Gesetzliche Unfallversicherung e.V. (DGUV) (2017b). *DGUV-Statistiken für die Praxis - Aktuelle Zahlen und Zeitreihen aus der Deutschen Gesetzlichen Unfallversicherung*. Bonifatius GmbH, Druck · Buch · Verlag, Paderborn.

Deutsche Gesetzliche Unfallversicherung e.V. (DGUV) (2017c). *Geschäfts- und Rechnungsergebnisse der gewerblichen Berufsgenossenschaften und Unfallversicherungsträger der öffentlichen Hand 2016*. Bonifatius GmbH, Druck · Buch · Verlag, Paderborn.

Dieterich, C., E. Herrmann, and M. Parzeller (2020). Tod bei der Arbeit–eine Analyse tödlicher Arbeitsunfälle von 2005 bis 2016 im Obduktionsgut des Instituts für Rechtsmedizin in Frankfurt am Main. *Rechtsmedizin 30*, 144–152.

Doleac, J. L. and N. J. Sanders (2015). Under the Cover of Darkness: How Ambient Light Influences Criminal Activity. *Review of Economics and Statistics 97*(5), 1093–1103.

Dustmann, C., T. Frattini, and G. Lanzara (2012). Educational Achievement of Second-Generation Immigrants: An International Comparison. *Economic Policy 27*(69), 143–185.

Eiríksdóttir, V. H., T. L. Ásgeirsdóttir, R. I. Bjarnadóttir, R. Kaestner, S. Cnattingius, and U. A. Valdimarsdóttir (2013). Low Birth Weight, Small for Gestational Age and Preterm Births Before and After the Economic Collapse in Iceland: A Population Based Cohort Study. *PloS ONE 8*(12), e80499.

Ericsson, S. (2020). Cultural Gender Norms and the Gender Gap in Math. Mimeo, Lund Univeristy, Department of Economics and Centre for Economic Demography.

Erikson, R., O. N. Skans, A. Sjögren, and O. Åslund (2007). Ungdomars och invandrades inträde på arbetsmarknaden 1985-2003. Report 2007:18, IFAU, Upssala.

European Parliament (2019a). Discontinuing Seasonal Changes of Time ***I - European Parliament Legislative Resolution of 26 March 2019 on the Proposal for a Directive of the European Parliament and of the Council Discontinuing Seasonal Changes of Time and Repealing Directive 2000/84/EC (COM(2018)0639 – C8-0408/2018 – 2018/0332(COD)). March. `http://www.europarl.europa.eu/doceo/document/TA-8-2019-0225_EN.html`, accessed on January 23, 2020.

European Parliament (2019b). Minutes of the Sitting of 13 September 2018. *Official Journal of the European Union C227, 05/07/2019*, 227–306.

European Parliament and Council (1994). Seventh Directive 94/21/EC of the European Parliament and of the Council of 30 May 1994 on Summer-Time Arrangements. *Official Journal of the European Communities* L(164), 1 – 2.

European Parliament and Council (1997). Eighth Directive 97/44/EC of the European Parliament and of the Council of 22 July 1997 on Summer-Time Arrangements. *Official Journal of the European Communities L*(206), 62 – 63. 01/08/1997.

European Parliament and Council (2001). Directive 2000/84/EC of the European Parliament and of the Council of 19 January 2001 on Summer-Time Arrangements. *Official Journal of the European Communities L*(31), 21 – 22. 2.2.2001.

Evans, D. K., M. Akmal, and P. Jakiela (2019). Gender in Education: The Long View. CGD Working Paper 523, Center for Global Development.

Fernandez, R. (2007). Women, Work, and Culture. *Journal of the European Economic Association 5*(2-3), 305–332.

Fernández, R. and A. Fogli (2006). Fertility: The Role of Culture and Family Experience. *Journal of the European Economic Association 4*(2-3), 552–561.

Fernández, R. and A. Fogli (2009). Culture: An Empirical Investigation of Beliefs, Work, and Fertility. *American Economic Journal: Macroeconomics 1*(1), 146–77.

Fernández, R. (2011). Does Culture Matter? In J. Benhabib, A. Bisin, and M. O. Jackson (Eds.), *Handbook of Social Economics*, Volume 1, pp. 481 – 510. North-Holland.

Figlio, D., P. Giuliano, U. Özek, and P. Sapienza (2019, November). Long-Term Orientation and Educational Performance. *American Economic Journal: Economic Policy 11*(4), 272–309.

Finseraas, H. and A. Kotsadam (2017). Ancestry Culture and Female Employment—An Analysis Using Second-Generation Siblings. *European Sociological Review 33*(3), 382–392.

Forslund, A., L. Liljeberg, and O. Åslund (2011). Flykting—och anhöriginvandrades etablering på den svenska arbetsmarknaden. Report 2017:14, IFAU, Upssala.

Gibson, M. and J. Shrader (2018). Time Use and Labor Productivity: The Returns to Sleep. *Review of Economics and Statistics 100*(5), 783–798.

Giuliano, P. (2007). Living Arrangements in Western Europe: Does Cultural Origin Matter? *Journal of the European Economic Association 5*(5), 927–952.

Giuntella, O., W. Han, and F. Mazzonna (2017). Circadian Rhythms, Sleep, and Cognitive Skills: Evidence from an Unsleeping Giant. *Demography 54*(5), 1715–1742.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics 118*(3), 1049–1074.

Goldin, C. (2006). The Quiet Revolution that Transformed Women's Employment, Education, and Family. *American Economic Review, Papers & Proceedings 96*(2), 1–21.

Grimm, A. M., G. Hoffmann, R. Ebertin, and M. Puettjer (1994). *Die geographischen Positionen Europas* (12 ed.). Ebertin Verlag, Freiburg im Breisgau.

Guiso, L., F. Monte, P. Sapienza, and L. Zingales (2008). Culture, Gender, and Math. *Science 320*(5880), 1164–1165.

Hansen, B. T., K. M. Sønderskov, I. Hageman, P. T. Dinesen, and S. D. Østergaard (2017). Daylight Savings Time Transitions and the Incidence Rate of Unipolar Depressive Episodes. *Epidemiology 28*(3), 346–353.

Harrison, Y. (2013). The Impact of Daylight Saving Time on Sleep and Related Behaviours. *Sleep Medicine Reviews 17*(4), 285–292.

Hastings, J. S., T. J. Kane, and D. O. Staiger (2006). Gender and Performance: Evidence from School Assignment by Randomized Lottery. *American Economic Review, Papers & Proceedings 96*(2), 232–236.

Havranek, T., D. Herman, and Z. Irsova (2018). Does Daylight Saving Save Electricity? A Meta-Analysis. *Energy Journal 39*(2), 35–61.

Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values.* Thousand Oaks, CA: Sage.

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations.* Thousand Oaks, CA: Sage.

Hofstede, G., G. J. Hofstede, and M. Minkov (2010). *Cultures and Organizations: Software of the Mind: Intercultural Cooperation and its Importance for Survival.* New York: McGraw-Hill.

Holland, N. and J. Hinze (2000). Daylight Savings Time Changes and Construction Accidents. *Journal of Construction Engineering and Management 126*(5), 404–406.

House, R. J., P. J. Hanges, M. Javidan, P. W. Dorfman, and V. Gupta (2004). *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies.* Sage publications.

Janszky, I., S. Ahnve, R. Ljung, K. J. Mukamal, S. Gautam, L. Wallentin, and U. Stenestrand (2012). Daylight Saving Time Shifts and Incidence of Acute Myocardial Infarction–Swedish Register of Information and Knowledge About Swedish Heart Intensive Care Admissions (RIKS-HIA). *Sleep Medicine 13*(3), 237–242.

Janszky, I. and R. Ljung (2008). Shifts to and from Daylight Saving Time and Incidence of Myocardial Infarction. *New England Journal of Medicine 359*(18), 1966–1968.

Jiddou, M. R., M. Pica, J. Boura, L. Qu, and B. A. Franklin (2013). Incidence of Myocardial Infarction with Shifts to and from Daylight Savings Time. *American Journal of Cardiology 111*(5), 631–635.

Jin, L. and N. R. Ziebarth (2020). Sleep, Health, and Human Capital: Evidence from Daylight Saving Time. *Journal of Economic Behavior & Organization 170*, 174–192.

Kantermann, T., M. Juda, M. Merrow, and T. Roenneberg (2007). The Human Circadian Clock's Seasonal Adjustment is Disrupted by Daylight Saving Time. *Current Biology 17*(22), 1996–2000.

Kaplan, E. K., C. A. Collins, and F. A. Tylavsky (2017). Cyclical Unemployment and Infant Health. *Economics & Human Biology 27*, 281–288.

Kellogg, R. and H. Wolff (2008). Daylight Time and Energy: Evidence from an Australian Experiment. *Journal of Environmental Economics and Management 56*(3), 207–220.

Kotchen, M. J. and L. E. Grant (2011). Does Daylight Saving Time Save Energy? Evidence from a Natural Experiment in Indiana. *Review of Economics and Statistics 93*(4), 1172–1185.

Kountouris, Y. and K. Remoundou (2014). About Time: Daylight Saving Time Transition and Individual Well-Being. *Economics Letters 122*(1), 100–103.

Kramer, M. S. (1987). Determinants of Low Birth Weight: Methodological Assessment and Meta-Analysis. *Bulletin of the World Health Organization 65*(5), 663.

Kuehnle, D. and C. Wunder (2016). Using the Life Satisfaction Approach to Value Daylight Savings Time Transitions: Evidence from Britain and Germany. *Journal of Happiness Studies 17*(6), 2293–2323.

Lahti, T., E. Nysten, J. Haukka, P. Sulander, and T. Partonen (2010). Daylight Saving Time Transitions and Road Traffic Accidents. *Journal of Environmental and Public Health 2010*, 1–3.

Lahti, T., J. Sysi-Aho, J. Haukka, and T. Partonen (2011). Work-Related Accidents and Daylight Saving Time in Finland. *Occupational medicine 61*(1), 26–28.

Lahti, T. A., S. Leppämäki, J. Lönnqvist, and T. Partonen (2006). Transition to Daylight Saving Time Reduces Sleep Duration plus Sleep Efficiency of the Deprived Sleep. *Neuroscience letters 406*(3), 174–177.

Lindo, J. M. (2011). Parental Job Loss and Infant Health. *Journal of Health Economics 30*(5), 869–879.

Lindo, J. M. (2015). Aggregation and the Estimated Effects of Economic Conditions on Health. *Journal of Health Economics 40*, 83–96.

Luttmer, E. F. and M. Singhal (2011). Culture, Context, and the Taste for Redistribution. *American Economic Journal: Economic Policy 3*(1), 157–79.

Machin, S. and S. McNally (2008). The Literacy Hour. *Journal of Public Economics 92*(5-6), 1441–1462.

Manfredini, R., F. Fabbian, A. De Giorgi, B. Zucchi, R. Cappadona, F. Signani, N. Katsiki, D. P. Mikhailidis, et al. (2018). Daylight Saving Time and Myocardial Infarction: Should We Be Worried? A Review of the Evidence. *European Review for Medical and Pharmacological Sciences 22*, 750–755.

Menclova, A. K. (2013). The Effects of Unemployment on Prenatal Care Use and Infant Health. *Journal of Family and Economic Issues 34*(4), 400–420.

Michelson, W. (2011). Sleep Time: Media Hype vs. Diary Data. *Social Indicators Research 101*(2), 275–280.

Miller, D. L., M. E. Page, A. H. Stevens, and M. Filipski (2009). Why Are Recessions Good for Your Health ? *American Economic Review, Papers & Proceedings 99*(2), 122–127.

Morassaei, S. and P. M. Smith (2010). Switching to Daylight Saving Time and Work Injuries in Ontario, Canada: 1993–2007. *Occupational and Environmental Medicine 67*(12), 878–880.

Niederle, M. and L. Vesterlund (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *Quarterly Journal of Economics 122*(3), 1067–1101.

Niederle, M. and L. Vesterlund (2010). Explaining the Gender Gap in Math Test Scores: The Role of Competition. *Journal of Economic Perspectives 24*(2), 129–44.

Nollenberger, N., N. Rodríguez-Planas, and A. Sevilla (2016). The Math Gender Gap: The Role of Culture. *American Economic Review, Papers & Proceedings 106*(5), 257–61.

Olafsson, A. (2016). Household Financial Distress and Initial Endowments: Evidence From the 2008 Financial Crisis. *Health Economics 25*, 43–56.

Orsini, C. and M. Avendano (2015). Macro-Economic Conditions and Infant Health: A Changing Relationship for Black and White Infants in the United States. *PLoS ONE 10*(5), e0123501.

Österreichische Nationalbibliothek (Ed.) (1916). *Reichsgesetzblatt 1916*, pp. 247. Aus der kais. kön. Hof- und Staatdruckerei.

Poland, M., I. Sin, and S. Stillman (2019). Why Are There More Accidents on Mondays? Economic Incentives, Ergonomics or Externalities. IZA Discussion Paper 12850, Institute of Labor Economics (IZA).

Pope, D. G. and J. R. Sydnor (2010). Geographic Variation in the Gender Differences in Test Scores. *Journal of Economic Perspectives 24*(2), 95–108.

Quentin, W., A. Geissler, D. Scheller-Kreinsen, and R. Busse (2010). DRG-Type Hospital Payment in Germany: The G-DRG System. *Euro Observer 12*(3), 4–6.

Reardon, S. F., E. M. Fahle, D. Kalogrides, A. Podolsky, and R. C. Zárate (2019). Gender Achievement Gaps in US School Districts. *American Educational Research Journal 56*(6), 2474–2508.

Robb, D. and T. Barnes (2018). Accident Rates and the Impact of Daylight Saving Time Transitions. *Accident Analysis & Prevention 111*, 193–201.

Rodríguez-Planas, N. and N. Nollenberger (2018). Let the Girls Learn! It Is Not Only About Math… It's About Gender Social Norms. *Economics of Education Review 62*, 230–253.

Ruhm, C. J. (2000). Are Recessions Good for Your Health? *Quarterly Journal of Economics 115*(2), 617–650.

Ruhm, C. J. (2003). Good Times Make You Sick. *Journal of Health Economics 22*(4), 637–658.

Ruhm, C. J. (2015). Recessions, Healthy No More? *Journal of Health Economics 42*, 17–28.

Schwartz, S. H. (2000, January). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, C. Triandis, C. Kagitcibasi, S.-C. Choi, and G. Yoon (Eds.), *Individualism and Collectivism: Theory, Method, and Applications*, Volume 18 of *Cross-cultural research and methodology*, pp. 85–119. Sage Publications, Inc.

Sexton, A. L. and T. K. Beatty (2014). Behavioral Responses to Daylight Savings Time. *Journal of Economic Behavior & Organization 107*, 290–307.

SGB VII (1996). *Das Siebte Buch Sozialgesetzbuch – Gesetzliche Unfallversicherung – (Artikel 1 des Gesetzes vom 7. August 1996, BGBl. I S. 1254), das zuletzt durch Artikel 35 des Gesetzes vom 12. Dezember 2019 (BGBl. I S. 2652) geändert worden ist.* Bundesministerium der Justiz und für Verbraucherschutz sowie Bundesamt für Justiz.

Sipilä, J. O. T., J. O. Ruuskanen, P. Rautava, and V. Kytö (2016). Changes in Ischemic Stroke Occurrence Following Daylight Saving Time Transitions. *Sleep Medicine 27*, 20–24.

Smith, A. C. (2016). Spring Forward at Your Own Risk: Daylight Saving Time and Fatal Vehicle Crashes. *American Economic Journal: Applied Economics 8*(2), 65–91.

Sood, N. and A. Ghosh (2007). The Short and Long Run Effects of Daylight Saving Time on Fatal Automobile Crashes. *BE Journal of Economic Analysis & Policy 7*(1, Article 11), 1–20.

Taras, V., J. Rowney, and P. Steel (2009). Half a Century of Measuring Culture: Review of Approaches, Challenges, and Limitations Based on the Analysis of 121 Instruments for Quantifying Culture. *Journal of International Management 15*(4), 357–373.

Torche, F. (2011). The Effect of Maternal Stress on Birth Outcomes: Exploiting a Natural Experiment. *Demography 48*(4), 1473–1491.

Toro, W., R. Tigre, and B. Sampaio (2015). Daylight Saving Time and Incidence of Myocardial Infarction: Evidence from a Regression Discontinuity Design. *Economics Letters 136*, 1–4.

Trivers, R. L. and D. E. Willard (1973). Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring. *Science 179*(4068), 90–92.

United States Congress (2019). Sunshine Protection Act of 2019. *116th Congress* (S.670).

van den Berg, G. J., A. Paul, and S. Reinhold (2020). Economic Conditions and the Health of Newborns: Evidence from Comprehensive Register Data. *Labour Economics 63*, 101795.

van der Klaauw, B. (2014). From Micro Data to Causality: Forty Years of Empirical Labor Economics. *Labour Economics 30*, 88–97.

Varughese, J. and R. P. Allen (2001). Fatal Accidents Following Changes in Daylight Savings Time: The American Experience. *Sleep Medicine 2*(1), 31–36.

Vincent, A. (1998). Effects of Daylight Savings Time on Collision Rates. *New England Journal of Medicine 339*(16), 1167.

Wadsworth, E. and D. Walters (2019, April). *Safety and Health at the Heart of the Future of Work: Building on 100 Years of Experience*. Switzerland: International Labour Office.

Wigglesworth, E. (2006). Occupational Injuries by Hour of Day and Day of Week: A 20-Year Study. *Australian and New Zealand Journal of Public Health 30*(6), 505–508.

Wolff, H. and M. Makino (2012). Extending Becker's Time Allocation Theory to Model Continuous Time Blocks: Evidence from Daylight Saving Time. IZA Discussion Paper 6787, Institute of Labor Economics (IZA).

Zhao, D., L. Zou, X. Lei, and Y. Zhang (2017). Gender Differences in Infant Mortality and Neonatal Morbidity in Mixed-Gender Twins. *Scientific Reports 7*(8736), 1–6.