CERGE
Center for Economics Research and Graduate Education
Charles University Prague

Essays on Local Labor Markets

Jakub Grossmann

Dissertation

Prague, July 14, 2021

Jakub Grossmann

Essays on Local Labor Markets

Dissertation

Prague, July 14, 2021

Dissertation Committee

Prof. Ing. Štěpán Jurajda, Ph.D. (CERGE-EI, Chair)

Randall K. Filer, Ph.D. (CERGE-EI, Hunter College)

Nikolas Mittag, Ph.D. (CERGE-EI)

Doc. Ing. Daniel Münich, Ph.D. (CERGE-EI)

Christian Ochsner, Ph.D. (CERGE-EI)

Filip Pertold, Ph.D. (CERGE-EI)

Referees

Tor Eriksson, Ph.D. (Aarhus University)

Kamil Galuščák, Ph.D. (Czech National Bank)

Acknowledgments

I would like to express my gratitude to my supervisor Štěpán Jurajda for his excellent guidance, useful advice, and continuous support during the process of writing my thesis. I greatly appreciate his efforts and the time he devoted to make me a better researcher and economist. Štepán, thank you.

I am also grateful to the members of my dissertation committee: Randall K. Filer, Nikolas Mittag, Daniel Münich, Christian Ochsner and Filip Pertold, for their helpful comments and inspiring discussions. Furthermore, I would like to express my appreciation to Felix Rösel for hosting me during my research visit in Dresden. I also want to thank the many people who gave me feedback on my work, and to audiences of numerous conferences and workshops that I attended. I am thankful for the CERGE-EI community, which has supported me and made my studies pleasant and trouble-free.

My gratitude also goes to the thesis referees Tor Eriksson and Kamil Galuščák for their effort in reviewing and providing feedback on the thesis.

I especially want to thank my friends and family for their encouragement and support during my studies.

I gratefully acknowledge financial support from the Charles University GA UK project No. 672218. and from the Volkswagen Foundation project 'The Rise of Populist Parties in Europe'.

Jakub Grossmann

Abstract

This thesis studies local labor markets affected by policy reforms, and shocks to health and migration. The effects studied in the three essays operate locally and are important for workers' labor-market outcomes, for family health members, and for long-term identity of local communities. The thesis contributes to existing empirical research by proposing new identification approaches and using new sources of variation. The essays quantify policy effects, some of them multi-generational, and ask about the underlying mechanisms behind the estimated effects. Each chapter focuses on a specific topic related to local labor markets or local communities in the Czech Republic.

In the first paper, I study the employment effects of four minimum wage increases implemented in the Czech Republic during 2012-2017, which cumulatively increased the national minimum wage by 37 percent. I analyze outcomes at the level of firm-occupation-county-specific job cells and apply an intensity-treatment estimator similar to that of Machin et al. (2003). My preferred specifications suggest that minimum wage increases led to higher wages for low-paid workers and did not have significant impacts on their rates of employment.

The second paper argues that a system of sick-pay is critical for balancing the economic and health costs of infectious diseases. Surprisingly, most research on sick-pay reforms does not rely on variation in worker exposure to diseases when investigating absences from work. This paper studies the effects on work absences of changes in health-insurance coverage for the first three days of sickness. I explore geographic variations in the prevalence of infectious diseases, primarily the seasonal flu, to identify variations in the need for sickness insurance. Estimates based on the Czech Structure of Earnings Survey imply that when sickness insurance is not available, total hours of work missed are not affected, but employees rely on paid and unpaid leave instead of sick-leave. The substitution

effects are heterogenous across occupations and socio-demographic characteristics of employees, and suggest that workers do not spread infectious diseases at their workplaces as a result of a lack of sickness insurance coverage in the first three days of an illness.

In the third paper we study how staying minorities who evaded ethnic cleansing integrate into re-settled communities. After World War Two, three million ethnic Germans were expelled from Czechoslovakia's *Sudetenland*, but some were allowed to stay, many of whom were left-leaning anti-fascists. We study quasi-experimental local variation in the number of anti-fascist Germans staying in post-war Czechoslovakia and find a long-lasting footprint: Communist party support, party cell frequencies, and far-left values and social policies are more prevalent today in locations where anti-fascist Germans stayed in larger numbers. Our findings also suggest that political identity supplanted German ethnic identity among stayers who faced new local ethnic majorities.

Tato diplomová práce zkoumá dopady veřejných politik, změn v nemocnosti obyvatel a migrace na lokální pracovní trhy. Efekty popsané ve třech esejích dizertace fungují lokálně a přinášejí důležité informace o chování pracovníků a jejich ekonomické aktivitě, o jejich nemocnosti a o nemocnosti členů jejich domácnosti, a dlouhodobé identitě lokálních komunit. Práce přispívá k existující empirické literatuře tím, že navrhuje nové identifikační přístupy a využívá doposud opomíjené zdroje variací v datech. Eseje kvantifikují dopady veřejných politik, jedna z nich mezigeračně, a odkrývají mechanismy stojící za zmíněnými dopady. Každá z kapitol se zaměřuje na konkrétní téma lokálních pracovních trhů nebo lokálních komunit v České republice.

V prvním článku studuji dopady čtyř nárůstů minimální mzdy mezi lety 2012-2017, které činily celkem 37 %. Dopady jsou zkoumány na úrovni pracovních buněk, které jsou organizačními jednotkami vytvořenými jako kombinace firmy, povolání a okresu. Obdobně jako v případě Machin et al. (2003) je pro odhad efektů aplikována "treatment intensity" strategie. Preferovaná specifikace naznačuje, že nárůsty minimální mzdy měly za následek růst mezd nízkopříjmových pracovníků ale neměly výraznější vliv na jejich zaměstnanost.

Druhý článek argumentuje, že systém zdravotního pojištění je důležitý pro vybalancování ekonomických a zdravotních nákladů infekčních onemocnění. Překvapivě, většina literatury, zabývající se dopady reforem zdravotního pojištění na absenci zaměstnanců, nespoléhá na variaci ve vystavení nákaze. Tento článek studuje efekty změny zdravotního pojištění v karenční době. Využitím regionální variace ve výskytu nakažlivých chorob, zejména sezónní chřipky, identifikuji variaci v potřebě zdravotního pojištění. Odhady na datech ISPV naznačují, že v případech, kdy není zdravotní pojištění dostupné, celkové odpracované hodiny zaměstnanců vystavených nakažlivým nemocem se nezměnily. Změnila se však jejich struktura a zaměstnanci spoléhali více na placené a neplacené volno namísto nemocenské. Tyto substituční efekty jsou značně heterogenní napříč profesemi a socio-demografickými charakteristikami zaměstnanců a naznačují, že zaměstnanci nešíří infečkní nemoci na pracovišti z důvodu nižších náhrad mzdy v případě nemoci.

Ve třetím článku studujeme, jak se minority, které se vyhnuly etnickým čistkám, integrovaly do nově vznikajících společenství. Tři miliony etnických Němců bylo po druhé světové válce vyhnáno ze Sudet, avšak některým bylo dovoleno zůstat, přičemž mnoho z nich bylo levicově smýšlejícími antifašisty. V tomto článku studujeme kvazi-experimentální lokální variaci počtu německých antifašistů, kteří nebyli po válce odsunuti a dokládáme, že podpora Komunistické strany, četnost stranických buněk, levicové hodnoty a sociální politiky jsou dodnes více přítomny v oblastech, kde zůstalo více antifašistů. Naše výsledky také naznačují, že politická identita u německé menšiny, která nebyla odsunuta po druhé světové válce, převládá nad identitou národní.

Contents

Chapter 1

The Effects of Minimum Wage Increases in the Czech Republic

1.1 Introduction

A large literature studies the ramifications of minimum wages, including its central effect on employment. Most of this work concludes that minimum wage increases have minor to no disemployment effects.[1] A problem with extending this predominantly US-based research to European countries is that the European minimum wage legislation is typically applied at the national level, which limits the use of region-based difference-in-differences identification strategies. Two types of approaches have been devised to allow identification of national minimum wage effects. First, the bunching estimator compares the number of jobs created above the level of an increased minimum wage level with the number of jobs destroyed just below that level (e.g. Meyer and Wise, 1983; Harasztosi and Lindner, 2019; Cengiz et al., 2019), assuming that the wage distribution would remain the same in the absence of a minimum wage increase. Second, the treatment-intensity

[1] See Doucouliagos and Stanley (2009) for a meta-analysis of the estimated employment effects.

estimator relies on variations in treatment exposure to national minimum wages typically generated by the pre-existing share of workers whose wages are below the minimum wage level set for the next year. This strategy, applied at the firm and/or establishment level, allows researchers to compare firms that would have to increase their wage bills to varying degrees due to a given minimum wage increase (introduction) in order to keep all of their workers (Machin et al., 2003; Eriksson and Pytlikova, 2004; Harasztosi and Lindner, 2019).

In this paper, we employ both of these strategies and are the first to apply the treatment-intensity approach at the level of job cells, defined as the combination of employee's occupation and employer's location. We study the impacts of four increases in the national minimum wage (NMW) in the Czech Republic implemented during 2012 to 2017. These increases followed a period of 7 years in which the NMW was not increased despite concurrent cumulative 16% growth in mean nominal wages in the economy (The Czech Statistical Office, 2019).[2] Starting at the level of 320 Euro[3] (32% of the average wage) in the beginning of 2013, these changes altogether amount to a 37.5% increase in the national minimum wage, reaching 440 EUR in 2017 (46% of the average wage). The highest annual increase occurred in 2017, when the NMW increased by 11%. We consider these increases both separately and jointly, as one significant NMW increase.

We start our analysis by applying the intensity-treatment estimator at the firm level (as, e.g. Harasztosi and Lindner, 2019, do). That is, we measure the treatment exposure to a NMW increase by the share of employees of a firm who are currently paid a wage that is below the minimum wage level set in next period (we refer to this measure as "the Share") and by the proportion of an employers' total wage bill that corresponds to the sum of wages that would have to be increased so that all current employees are paid at least the minimum wage in next period ("the Gap" measure). A potential weakness of this strategy is, first, that minimum wage increases could be timed to correspond to demand shocks in industries that employ a heavy share of low-wage labor, i.e. that an unobserved demand shock at the firm level could make the minimum wage endogenous. Sec-

[2]This a similar situation to that of Hungary in 2001, studied in Harasztosi and Lindner (2019).

[3]We apply an approximate conversion rate of 25 CZK per 1 EUR throughout the paper.

ond, since the share of workers paid below a future NMW in the average firm is around 3%[4] in 2012, such a strategy mixes the effects of NMW on low-wage employment with the evolution of employment high above the NMW level, which is unlikely to be causally affected by the NMW. Indeed, some applications of the treatment-intensity approach focus on specific establishments characterized by a high exposure to NMW increases and homogenous labor composition (e.g. Machin et al., 2003 who study care homes), which, however, limits the generalizeability of the results.

In the second step of our analysis, we therefore employ treatment intensity variation at the job-cell level, where job cells are groups of workers in the same firm, in the same location and in the same occupation, and we study only low-wage job cells. This allows us to exploit variation in exposure to NMW increases at the level of homogenous groups of low-wage workers, i.e., we do not mix the NMW effects on employment of low- and high-wage employees. The strategy also allows us to control for firm-level evolution of employment by conditioning on firm FE. We supplement the employment analysis of NMW effects at the job-cell level by asking whether NMW increases affected job-cell worker turnover and employment structure in terms of education, gender, or worker firm-specific tenure.[5]

Third, we apply a bunching estimator. Assuming that real wage distributions would not change in the absence of NMW increases, we compare the number of jobs created above a new NMW level with the number of jobs destroyed below the new NMW level. The real wage distribution from a period preceding a NMW increase is used as a counterfactual. We inspect employment dynamics in a region of +/- 100 EUR (2,500 CZK) around a new NMW level, i.e. approximately 70-130% of the NMW level in 2013.

Our analysis relies on the Czech Structure of Earnings Survey (SES), which offers several advantages for a study of NMW effects.[6] The SES is a large panel of nearly

[4]For more details and the shares of workers paid below a future NMW in later years see Table 1.8 in the Appendix.

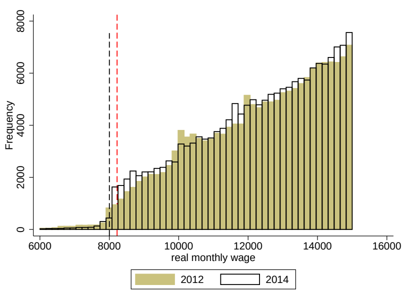
[5]This is important for understanding the employment effects. If low-educated low-wage workers are replaced by high-wage high-education workers as a result of NMW increases, we would detect no employment effects.

[6]The SES is the linked employer-employee dataset (LEED) designed to collect harmonised data on earnings in EU Member States.

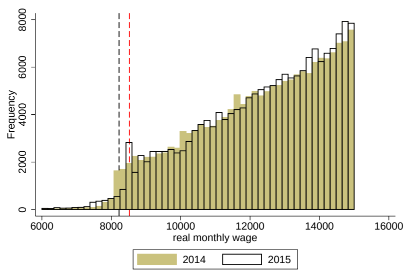
4 thousand firms which provides detailed information about *all* employees working in the firms surveyed. The SES covers approximately 1.5 million employees each year (out of about 3.5 million salaried employees in the Czech private employment sector). This allows us to observe a large number of homogenous job cells and to exploit significant variations in treatment exposure to NMW increases across these cells. The SES also allows us to study changes in hours worked in addition to employment changes.

Figure 1.1 shows that each of the four NMW increases did result in a shift in the wage distribution.[7] Individual graphs show wage distributions for years surrounding each NMW increase; the only exception is graph (a) which shows wage distributions for 2012 and 2014 because the NMW increase was implemented in the middle of 2013. Black horizontal lines denote the initial levels of the NMW and red lines indicate NMW levels after each increase.

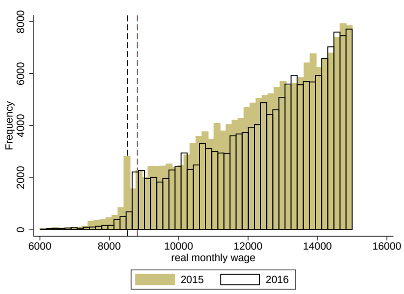
Figure 1.1: Changes in wage frequency distributions



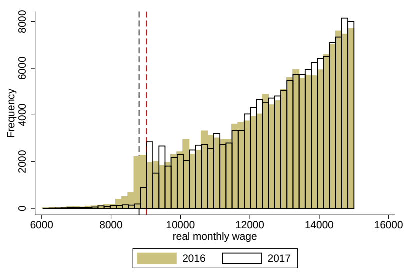
(a) NMW increase in 2013



(b) NMW increase in 2015



(c) NMW increase in 2016



(d) NMW increase in 2017

[7]To show changes in wage distributions net of a general price rise in economy, we discount nominal wages by the median wage growth, because inflation rates were very small during the years studied.

Our firm-level analysis produces mixed results. We find a significantly negative employment effect associated with the 2013 NMW increase, but positive employment effects associated with the NMW increases in 2015 and 2016. The employment elasticity with respect to minimum wage associated with the 2013 increase is -0.154[8] (the estimated coefficient is almost 25%). One possible explanation is that employers had already considered possible future NMW increases and adjusted employment accordingly. It could also be that our exposure measures correlate with employment trends of high-wage workers within a firm. It is worth stressing that there is another factor that may be behind such a large estimate. We analyze the NMW increase in 2013 using data from 2012 and 2014, and it is possible that the estimated coefficient also captures changes in employment that were not caused by the NMW increase. The size of the firm-level based estimate is large compared to previous work. For example, Eriksson and Pytlikova (2004) found in one specification that a NMW increase in 2000 in the Czech Republic caused a 14% decrease in employment. Our estimates show that subsequent NMW increases in 2015 and 2016 had opposite, i.e. positive, effects on employment, which both amounted to approximately 16% (the employment elasticity wrt. NMW are 0.03 and 0.08, respectively). We do not find any statistically significant employment effect associated with the 2017 NMW increase.

Compared to the firm-level results, our job-cell results are more consistent and in line with the existing literature. The estimated coefficients do not switch in signs across years. We find negative employment effects only for the 2013 NMW increase. However, this effect is economically small and below the level of estimates appearing in the recent literature. The NMW increase in 2013 caused a 11% drop in employment for job cells in which all employees were paid less than the NMW (corresponding to an elasticity of -0.066), which is small in comparison to previous research. We also find a small negative and statistically significant effect on employment in 2016 in one specification. Next, we focus on selected

[8] It is not obvious how to compare elasticities obtained from difference-in-differences with intensity-treatment estimate types directly. Therefore, different approaches to facilitate the comparison are used in the literature. For example, Harasztosi and Lindner (2019) adjust their estimated elasticity by 25 %, as this is the share of directly affected teenage employees in the US population. We multiply the estimated elasticities by the share of directly affected workers (i.e. workers paid below the NMW level set in the next period) in our data-set.

job cells. First, we analyze 5 of the most affected occupations separately.[9] We find no negative employment effects using these job cells. Second, we focus on job cells in accommodation and restaurants. Our preferred specification does not show any negative effects on employment. Third, manufacturing does not show negative effects on employment except for the NMW increase in 2013 when analyzed separately.

In line with our treatment-intensity analysis, the bunching-based estimates indicate that employment changes induced by the NMW increases were negligible. Some 4% of jobs around the minimum wage threshold were destroyed in 2017, which was the most extreme case. We find also positive employment effects associated with the 2013 and 2015 increases. However, the bunching estimates are sensitive to the chosen upper and lower bounds, which determine the region of wage distribution where employment changes are measured.

We supplement our job-cell analysis by inspecting additional effects of NMW increases. We find that the NMW increases during 2013-2017 did not affect: i) hours worked, ii) turnover rates, iii) educational and gender composition of job cells. Furthermore, by estimating employment effects on specific parts of the job-cell employment distribution[10], we address a possible correlation between labor demand shocks and the employment evolution of exposed job cells. Our findings suggest that labor demand shocks did not affect our job-cell level results. We also estimate the effects of a hypothetical situation in which a sizable NMW increase is implemented. We combine individual NMW increases during 2012-2017 and treat them as one large hike in the NMW. Our estimates suggest that such a considerable increase would not have a negative effect on employment.

To better understand the sources of discrepancies between estimates on the firm and job-cell levels, we ask how well the firm-level exposure measure predicts employment changes separately for low- and high-paid employees. Our results show that there are no statistically significant relationships, i.e. the minimum wage in-

[9] We include 5 occupations with the highest mean value of the Share variable for each year. These are mostly employees with ISCO codes 5 and 9 (elementary occupations, service and sales workers).

[10] We consider the job-cell employment distributions without the upper and lower quartiles to exclude positive and negative demand shocks.

creases did not cause decreases in the employment of low-paid employees, which is in contradiction to our firm-level estimates. A closer look at the estimated coefficients for control variables shows that they differ significantly in both subsamples. This finding suggests that the firm-level aggregated characteristics fail to sufficiently control for specific characteristics of low-paid employees and likely correlate with the firm-level minimum wage exposure. Therefore, we do not find our firm-level estimates credible and prefer the results estimated on the job-cell level. A lesson learned is that the application of the treatment intensity strategy to estimate the effects of minimum wage increases should be carried out on homogenous groups of employees; otherwise, the estimated effects may be biased.

This paper relates to the minimum wage literature in several ways. First, we contribute to the work on the employment effects of minimum wage increases. Similarly to recent EU analyses, we find no or small negative effects of the NMW increases we study on employment. Second, our paper contributes to the part of the literature that applies a treatment-intensity estimator to estimate the employment effects of NMW increases. Often, this approach is the only possible means to identify the causal effects of nation-wide minimum wage increases. This type of research is usually carried out at the firm level (e.g., as in Harasztosi and Lindner 2019, and Eriksson and Pytlikova 2004, who use SES data), or researchers use occupation-specific organizational units (as in Machin et al. 2003, who analyzed employment patterns in the care-homes industry during the introduction of a NMW in the UK). The major advantage of using SES or similar data is that such data-sets contain characteristics of a large number of firms and their employees. However, the firms surveyed are large heterogenous organizational units that may employ only a small fraction of workers exposed to a NMW increase.[11] Therefore, measuring the exposure to a NMW increase by the share of affected employees in firms is likely to be imprecise. Moreover, this approach mixes the employment trends of low- and high-paid workers. The other alternative used in the literature is to focus on specific occupational units that group employees with similar wages and characteristics. Therefore, it is possible to precisely zoom

[11] Approximately only 3% of employees were affected by the 2013 NMW increase in the Czech Republic. For more details see Table 1.8.

in on low-paid employees and study their employment changes. The drawback is that it is difficult to collect such data-sets. Furthermore, the estimated effects of an NMW increase are not generalizable. In this paper, we focus on firm-occupation-county-specific job cells, combining the benefits of both approaches. Using SES data provides a large number of observations, making it possible to focus on specific labor-market segments, i.e. to target subgroups of employees that are likely to be affected by a NMW increase and to compare job cells that are similar. Job cells are small homogenous units and their characteristics derived from information on individuals are more accurate than these of firms. Moreover, when NMW increases are small, the higher variance in exposure at the job-cell level facilitates more precise estimations. We believe that using job cells allows us to compare employees who are in the same part of the wage distribution, and who have similar individual characteristics, but who differ in exposure to NMW increases. Therefore, we estimate the true effects of NMW increases whilst the firm-level analysis also includes the effects on employees who are paid well above the minimum wage. Third, we contribute to the literature on the effects of NMW increases in the Czech Republic. There have been only a few papers studying the effects of minimum wage changes in the Czech Republic (Eriksson and Pytlikova 2004; Fialová and Mysíková 2009; Duspivová et al. 2013). Eriksson and Pytlikova (2004) study relatively large increases - varying from 11.1% to 35.8% - in the minimum wage in the Czech Republic during 1999-2002. Using Czech Structure of Earnings Survey (SES) data, they estimate the effects of NMW increases on wages and employment of low-paid employees at the firm level. They follow the approach used in Card (1992a) and construct two variables which measure exposure to NMW increases. Their findings suggest that legislative changes had a positive effect on wages, and there was a small negative effect on employment in some specifications, especially for small firms.

This paper proceeds as follows. Section 1.2 presents the institutional context. Section 1.3 describes the data. Section 1.4 presents our units of interest - job cells. Section 1.5 shows the methods we apply. Section 1.6 presents the employment analysis. Section 1.7 presents wage effects. Section 1.8 shows the bunching-based estimates. Section 1.9 concludes.

1.2 Institutional Context

The existence of a minimum wage is anchored in the Czech Labor Code. The minimum wage has been changed 20 times during the last 25 years (Ministry of Labour and Social Affairs, 2018). The majority of the changes occurred during 1998-2007, when the minimum wage more than tripled to 320 Euro - 8,000 Czech crowns per month. This level remained stable until 2013. Since then, the minimum wage has been adjusted every year except in 2014.[12] The percentage increases with monthly minimum wage levels in brackets are 6.25% (340 EUR - 8,500 CZK); 8.2% (368 EUR - 9,200 CZK); 7.6% (396 EUR - 9,900 CZK); 11.1% (440 EUR -11,000 CZK) in August 2013; January 2015; January 2016; January 2017 respectively. The minimum wage in the Czech Republic is established by Act no. 262/2006 Sb. (Labor Code) and the minimum wage levels are set by Government Decrees. Plans for increasing the minimum wage level are usually discussed publicly, and include representatives of employees and employers, but it is the Government which ultimately decides about the minimum wage levels in the Czech Republic. A Government Decree becomes binding by inscription into the Legal Code. Minimum wage Government Decrees are usually published in the Legal Code during the fall season, and become effective as of January 1st, allowing employers time to prepare for the wage increases to some extent.[13]

Graph 1.2 shows the evolution of the monthly minimum wage in the Czech Republic together with minimum wage to average wage and median wage ratios. Red bars indicate the timing of minimum wage changes. The minimum wage to average wage ratio varies from approximately 0.3 to 0.42 during the observed period; the rates are similar to those in neighboring countries in the region (OECD, 2018).

The Czech Republic is a central European post-communist country with economic activities distributed unequally across regions. The diversity results in the minimum wage setting being most effective in certain industry, occupation, or regional-specific clusters. The simple (unweighted) average wage in NUTS-4

[12] Our analysis ends by the increase in 2017.

[13] For example, the government decree affecting the minimum wage level in 2017 was published in the Legal Code on October 5th, 2016.

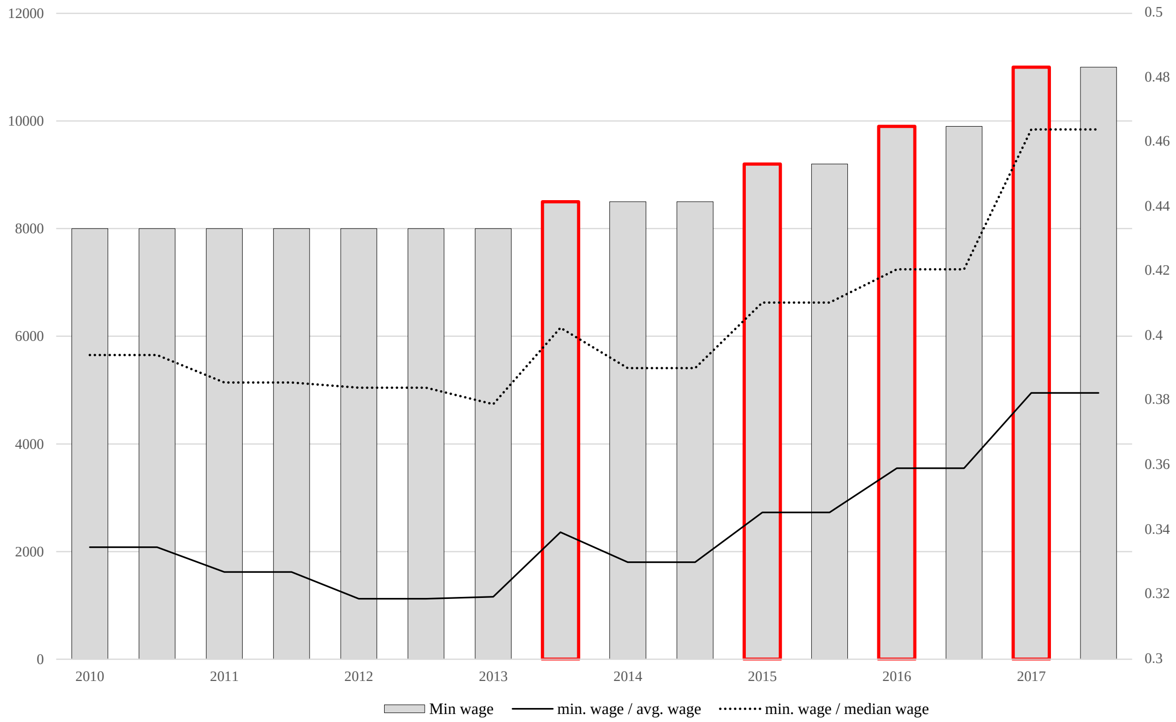


Figure 1.2: Minimum wage evolution

regions in the Czech Republic is 1,150 EUR (28,761 CZK) with a standard deviation of 113.3 (2,833); the average industry wage (according to 19 CZ-NACE groups) is 1,114 EUR (27,857 CZK) with a standard deviation of 359.8 (8,919); the average wage of CZ-ISCO major groups is 1,041 EUR (26,027 CZK) with a standard deviation of 340 (8,492) (The Czech Statistical Office, 2017). The lowest wages are traditionally in accommodation and food service, and administrative and support service industries. Occupations with the lowest wages in the Czech Republic are mainly elementary occupations, and service and sales workers with an average monthly wage of 619 EUR (15,466 CZK) and 670 EUR (16,755 CZK) respectively in 2016. According to the Ministry of Labor and Social affairs (2018), the average unemployment rate across 77 NUTS4 counties was 3.8% with a standard deviation of 1.45 at the end of 2017.

1.3 Data

We use information on private-sector employees from the Czech Structure of Earnings Survey for 2012-2017.[14] The data include an unbalanced panel of firms with repeated cross-section observations of all workers employed by each firm.[15] We work with annual data, i.e. we use aggregated data that were reported in each quarter of the year. We use data for 2012 and 2014 to analyze the first increase in NMW, as it happened in the middle of 2013. For the rest of the increases, which occurred in January of each year, we use data on the two years around each NMW increase. Firm-level data provide information about firms´ location (NUTS4 classification), a 4-digit NACE industry code, and the presence of a collective agreement. Data on employees include gender, age, place of work (NUTS4), hours worked, salary, 4-digit ISCO occupation, education, and tenure in the job. The number of observations in the annual surveys vary from 1.23 to 1.31 million per year during the period studied. After dropping observations with missing values and keeping only full-time workers, we are left with approximately 95% of the original data in each year.

1.4 Job Cells

Our main units of interest are firm-occupation-county-specific job cells. Job cells are groups of employees with similar skills and wages, and we argue that they are more homogenous than firms in terms of employment dynamics. There are 3,655 firms and 51,977 job cells in our data in 2012. The higher number of job-cell observations allows us to zoom in on the bottom segment of the job-cell wage distribution, where low-paid workers are sorted. Assuming that employment changes of high-paid and low-paid employees differ, focusing only on low-paid job cells brings us closer to satisfying the identifying assumption of the same

[14] The European Union Structure of Earnings Survey is designed to collect harmonized data on the relationships between the level of remuneration and individual characteristics of employees in EU Member States, including the Czech Republic. For details see `https://ec.europa.eu/eurostat/web/microdata/structure-of-earnings-survey`.

[15] Firms in the Czech SES are selected using the stratified sampling method with following strata: firm size - 4 groups, industry - 6 groups, region - 14 regions. The Czech SES covers 1.2% of firms up to 9 employees, 4.5% of firms with 10-49 employees, 15% of firms with 50-249 employees, and 100% of firms with 250 and more employees. For more details see `www.ispv.cz`.

employment evolution in organizational units which are and are not affected by the NMW increase.

We generate job-cell characteristics from individual characteristics of employees who belong to the same job cell. Most importantly, we create log(cell wage) as the mean of individual log(monthly wages) and log cell employment for each year. Similarly, we generate shares of females, Czech nationals, average age, tenure, and their squared counterparts.

1.5 Methods

Unlike in US-based research, we cannot apply region-based difference-in-differences identification strategies. Therefore, we use the treatment-intensity estimator applied in, e.g., Machin et al. (2003). This approach compares organizational units with different treatment exposures to a NMW increase, to estimate the causal effects of NMW increases on changes in employment and wages. The underlying identification assumption is that affected and non-affected firms / job cells would follow the same employment evolution in the absence of NMW increases.[16] The treatment exposure is typically measured by the share of employees who are paid below the NMW level set for the next year. Below, we define two measures representing the exposure to NMW changes, which we subsequently use in regressions in which changes in wages and employment are on the LHS and the exposure measures are on the RHS of our regression equations. This approach allows us to identify what parts of changes in employment and wages are induced by increases in the NMW.

We define an indicator for a worker i who is paid in period t below a new NMW level set in period *t+1*:

$$affected_{i,t} = \begin{cases} 1 & \text{if } \mathrm{NMW}_{t+1} > \text{monthly wage}_{i,t} \\ 0 & \text{otherwise} \end{cases}$$

[16] It is not possible to test this assumption directly. However, we test for different employment evolution for firms / job cells which were / were not affected by the NMW increase in 2013 during 2011-2012, and find that employment trends were the same for firms with different exposure to the 2013 NMW increase. Results are available upon request.

We follow by considering a simple share of workers paid below a new NMW level (extensive margin) and call it the *Share*,

$$Share_{j,c,o,t} = \frac{\sum_i affected_{i,j,c,o,t}}{N_{j,c,o,t}} \tag{1.1}$$

where a subscript j states for firms, c county, o occupation and N is the total number of workers within a specific firm / job cell. The share measure equals the share of workers below the new NMW and captures the extent to which a firm / job cell is exposed to an upcoming change in the NMW. However, the Share measure does not inform us how intense the effects would be. Therefore, we construct another variable called *Gap* which measures the size of this exposure (intensive margin) on the firm / job-cell level. [17] This measure also reflects the size of job cells, being larger for job cells with more workers provided that the values of the Share variable are similar.

$$Gap_{j,c,o,t} = \frac{\sum_i max(w_{t+1}^{min} - w_{ij}, 0)}{\sum_i w_{ij}} \tag{1.2}$$

Where the w_{t+1}^{min} variable is a NMW level expressed in terms of monthly wage and the w_{ij} variable is average monthly wage computed as the total money paid to a worker during the period observed, divided by the number of months an employee worked. See the Appendix for more details about the construction of the monthly wage. The Gap variable measures the proportion of employers' wage bills that must be increased so that all workers are paid at least a NMW level set in next period.

More than 21,000 employees are directly affected by the 2013 NMW increase in our dataset in 2012, corresponding to less than 2% of employees (unweighted). These workers must either be paid an increased wage or be laid off. The share of employees affected by the subsequent increase was 0.85%. The share was growing

[17] There are several options to measure the intensity, however, we stick to this widely used measure from the literature, e.g. in Machin et al. (2003).

since then, accounting for 2.09% in 2015, and 2.37% in 2016. Mean distances between the monthly wages of affected workers and the new NMW level in the next period vary around 23-36 EUR (580-890 CZK) across the NMW increases. Detailed summary statistics at the individual level can be found in the Appendix, Table 1.3.[18]

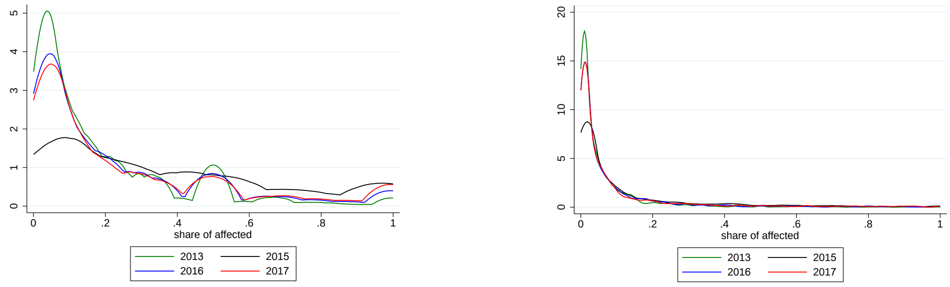
Our data show that affected employees are concentrated in specific occupations and industries. Table 1.5 in the Appendix summarizes shares of employees who were affected by NMW increases for each specific ISCO group and year. The groups that experienced the highest shares are *Elementary Occupations* and *Service and Sales Workers*. Focusing on specific industries, the highest shares of affected employees are in the *Hotel and Food Services* and *Real Estate* industries. Detailed statistics can be found in Table 1.6 in the Appendix.

We also compare characteristics of the job cells that were / were not affected. We call a job cell (firm) “affected" if at least one employee is paid below the NMW level effective in the next period. Affected jobs cells are larger: the average size of the affected job cells is 66 employees vs. only 18 employees for unaffected job cells in 2012. The affected job cells include significantly higher shares of females, slightly older employees in some of the years studied, a lower share of Czech nationals, and less time in the job. The affected job cells have a significantly lower number of hours worked compared to those which were not affected. Not surprisingly, the affected job cells contain more people with primary and vocational education. Overall summary statistics at the job-cell level are presented in the Appendix, Table 1.7. Firm level characteristics show similar patterns to the job-cell level; detailed summary statistics can be found in the Appendix, Table 1.8.

Figures 1.3a and 1.3b show the distributions of the Share variable for all affected job cells and firms. Each color represents the distribution for a particular NMW increase. The Share measure has higher variation in the case of job cells; the standard deviations across the years studied vary in the range of 0.21-0.32 and 0.11-0.19 in the case of job cells and firms, respectively. The Gap variable is

[18]Table 1.4 in the Apendix shows the minimum wage coverage by employee'characteristics in our sample.

distributed similarly to the Share measure.



(a) Job-cell level (low-wage job cells) **(b)** Firm level (all firms)

Figure 1.3: Kernel density of the Share measure: Firm vs. Job-cell levels

1.6 Employment Effects

To estimate the effects of the NMW increases on employment, we estimate Equation 2.1

$$\Delta log(employment)_{j,c,o,t} = \alpha_1 + \beta_1 * Share_{j,c,o,t-1} + \beta_2 * Gap_{j,c,o,t-1} + \delta_1 * X_{j,c,o,t-1} + \psi_{j,c,o,t} \tag{1.3}$$

where the dependent variable is the change in log(firm / job-cell employment). Our coefficients of interest are β_1, β_2 (estimated separately), X are control variables and ψ are firm and county two-way cluster-robust errors. We use the Share and Gap variables to measure the NMW exposure. We also weight Eq. 3 by the number of employees in firms / job cells. The Share coefficient (β_1) states the average (dis)employment effect for a firm / job cell, where all employees are affected (i.e. paid below the NMW level effective in the next period); the comparison groups are firms / job cells where no employee is affected.[19] Similarly, the Gap coefficient (β_2) states the average (dis)employment effect if employers would have to double their wage bills as a consequence of a NMW increase.

[19] There is no firm and only a small number of job-cells that contain exclusively employees who are affected by a NMW increase. For example, among the affected units in 2012, the average share of affected employees was 19% for job cells and 6% for firms (see Tables 1.7 and 1.8 in the Appendix). However, we adhere to this interpretation to make our results comparable with previous research.

We use only firms and job cells that exist in both periods for most of our analysis, i.e. we do not include newly created or destroyed units.[20] We analyze job cells from the first quartile of the job-cell wage distribution as these units are most likely to be affected by a NMW increase, and they are covariates in terms of their characteristics. Table 1.1 summarizes the estimated employment effects. Rows represent different econometric specifications. Columns show estimates for three different exposure measures associated with each of the four NMW increases. We focus on the estimated Share coefficients in our interpretation below. In cases where the Gap and the weighted Share estimates are statistically different from zero, the estimated effects have the same signs as the Share estimates. The size of the estimates is similar to our Share estimates in our job-cells specifications.

We first carry out analysis at the firm level, as this is usually the level used in the literature. Our results are presented on the first line of Table 1.1. The firm-level results suggest that there is a negative effect on employment associated with the NMW increase in 2013 and there are small positive effects on employment in 2015 and 2016. The estimated negative coefficient amounts to almost 25% (employment elasticity[21] with respect to minimum wage is -0.154).[22] One explanation for such a large effect may be that employers expected future increases in minimum wage levels and, thus, adjusted their decisions about employment accordingly. Another possible explanation is that the Czech Republic was still experiencing the end of the Great Recession in 2013 and this was a different economic situation compared to subsequent years when the NMW was increased during the economic boom (The World Bank Group, 2019). To address a possible effect of the economic cycle, we test whether firm-level employment and wages are more procyclical in firms with a higher share of low-paid employees.[23] We do not find

[20] We perform a robustness check by putting 0 for destroyed and created job cells to account for possible employment effects, which are not captured in our estimates.

[21] We compute employment elasticity with respect to minimum wage as: (%Δ employment due to the NMW increase (i.e., β_1 from Eq. 3) / % increase in the NMW (computed as the increase in the NMW relative to median wage in the economy))*(share of directly affected employees in our data).

[22] Importantly, this effect is driven by small firms; for detailed results see Table 1.9.

[23] We use Czech SES semiannual data for 2007-2012, i.e. for the years when the NMW was not raised. We construct a panel of firms and estimate regression equations where the dependent variables are firm-level percentage changes in wages and employment. Our independent variables are shares of low-paid workers - measured as a share of workers whose wages belong to

that firms with more low-wage workers have different employment patterns during economic booms and busts.[24] The full set of results can be found in Tables 1.17 and 1.18 in the Appendix. It is also possible that more exposed firms were affected more severely by a 2013-specific negative labor demand shock.[25] Finally, it is possible that the firm-level estimates also capture the employment trends of high-wage employees. To address this concern, we estimate the firm-level equation using the employment changes of low- and high-paid employees as separate dependent variables.[26] Our estimates in Table 1.15 show that the Share exposure measure on the firm level does not predict employment changes for low- and high-paid employees separately, though the estimates are precisely estimated. A closer inspection reveals that the estimates of control variables significantly differ across these two subsamples. For example, the average industry FE for low-paid employees in 2013 is 0.073, -0.185 for high-paid employees, and the difference is -0.255, which is a sizable effect similar to the firm-level estimate of the Share measure in the same year (see Table 1.1). This suggests that the application of the treatment intensity estimator on heterogenous groups of employees can lead to biased estimated effects of the minimum wage increases. For this reason, we prefer our job-cell estimates to the firm-level ones.

Results based on our preferred units of interest - job cells - present a different picture. Our estimates suggest that there were no or only small negative employment effects. Although we are not able to directly test the identifying assumption of equal employment trends in the absence of NMW increases, we aim to get as close as possible to satisfying this assumption by considering only job cells from the 1st quartile of the job-cell wage distribution. Table 1.1 shows that there were negative effects on employment only in 2013, and that these were negligible. The employment effects in 2013 are comparable for various specifications, amounting

the 1st decile, 1st quartile, or bottom half of the wage distribution. We approximate economic performance by the industry-specific growth rate in production based on OECD (2019) STAN data. We use standard controls and include firm fixed-effects.

[24] We have also found that firm-level wages are procyclical, the percentage growth is faster in firms with higher shares of low-paid employees and the interaction of a share of low-paid and the economic growth is associated with negative effects on wage growth.

[25] To account for this issue, we run regression equations with firm fixed-effects on the job-cell level.

[26] To stay consistent with our job-cell approach, we consider employees from the first quartile of job-cell wage distribution to be low-paid employees, and those that belong to the 2nd -4th quartiles to be high-paid employees

to -10.6% in our baseline specification and -13.1% in the specification with firm fixed-effects; the associated employment elasticities with respect to the minimum wage are -0.066 and -0.081, respectively. These effects are small and comparable to previous findings (e.g., Harasztosi and Lindner, 2019, found employment elasticities around -0.035). We found consistently statistically significant employment effects only in 2013. Connecting job cells across all years, we are able to estimate the employment effects of the NMW increases, controlling for job-cell fixed-effects. Table 1.12 presents these estimates. We find a small negative effect on employment which amounts to -5.72% with this specification.

Figure 1.4 presents the Share estimates for different specifications (the first three rows in Table 1.1) together with 95% confidence intervals. The horizontal axis indicates the year of a NMW increase and the vertical axis shows the size of our coefficient estimates. As recent research argues (Brewer et al., 2019), relevant public policy recommendations should consider not only a failure to reject the null hypothesis but also the range of estimated effects on employment. Figure 1.4 shows that the firm-level estimates are less precise than their job-cell level counterparts. The absolute values of the point estimates are usually higher in the case of firms, however, they are not statistically different from the job-cell level estimates. Our job-cell level estimates are consistent and economically small.

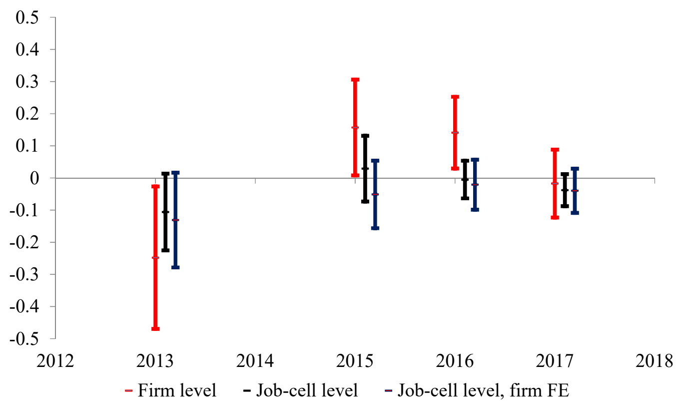


Figure 1.4: Estimates comparison - the Share measure

The employment effects of minimum wage increases are probably the most often studied, however, there are other effects of interest closely related to minimum wage increases.

Specific occupations - Some low-skilled occupations are more likely to be affected by minimum wage increases. Instead of focusing on one specific occupation, we select five occupations with the highest mean values of the Share exposure measure in each year.[27] The most affected occupations belong mainly to groups 5 and 9 (*Elementary Occupations*, and *Service and Sales Workers*). Our results do not show any negative employment effects during the years studied (Table 1.1, line “5 most affected occupations").

Specific industries - We focus on job cells in specific industries, which are likely to be affected by NMW increases. We are interested in responses of the hotel and food service industries to a rise in NMW.[28] Surprisingly, we find no statistically significant decrease in employment associated with any of the years studied. Estimates appear in Table 1.1, line “Accommodation and food services". We also estimate the employment effects using job cells from the manufacturing industry separately. This is the largest industry in our study, accounting for almost 40% of salaried employees in the Czech Republic. The results presented in Table 1.1, line “Manufacturing" do not show any disemployment effects, except for the 2013 increase. Our estimate, in terms of size, is similar to the firm-level estimate in the same year, -27% (employment elasticity with respect to minimum wage is -0.106). Similarly to the employment effects associated with the NMW increase in 2013 at the firm level, these results are driven by small units (see Table 1.10).

Destroyed / created job cells[29] - Considering only job cells that are observable in both periods surrounding the NMW increases in our analysis may neglect systematic closure and / or creation of job cells. Therefore, we carry out a robustness exercise, where we assign 0 for employment when a job cell is missing in our data. The results appear on the line “0 if missing" in Table 1.1. The results indicate that there are negative employment effects mainly associated with the

[27] We use two digit ISCO classification in this case. Table 1.5 presents shares of affected employees on the one digit ISCO level.

[28] There is anecdotal evidence that some employees are officially paid exactly the minimum wage level but receive additional pay ‘off the books’.

[29] By construction, we are not able to determine the exposure to a NMW increase for job cells missing in the period before a NMW increase, as we do not observe wages of employees working in these job cells.

NMW increases in 2013, 2015, and 2016.[30] Nevertheless, similarly to our baseline specification, these estimates are rather small.

110 pcnt. of NMW levels - To account for possible spillovers, we arbitrarily set the NMW levels to 110 percent of their original levels and estimate the employment regressions. We do not find evidence that the NMW increases affected employment of workers paid above the new NMW levels.

Hours worked - One reasonable concern is that employers may not lay employees off entirely, but may instead reduce their working hours. To investigate this issue, we estimate equations with changes in log hours worked as the dependent variable. The estimates in the row "LHS: Hours worked" in Table 1.1 show that we do not see such behavior in our data.

Turnover - It is possible that employers replace employees paid less than new NMW levels with new workers who are paid higher wages. In this case, the employment level could remain the same and the job cell would show higher average cell wages. To explore this issue, we run regressions with turnover as the dependent variable. The results are presented in Table 1.1. We do not find any systematic evidence that employers substitute workers more in job cells where they are paid below the NMW levels.

Skill substitution -It can also happen that employers who are forced to raise wages engage in skills substitution to mitigate their costs: i.e., replacing low-skilled workers with better-skilled new employees. To inspect this scenario, we run a set of regressions with average education as the dependent variable (the education category variable ranges from 1 to 6 according to the highest education attained). We do not find that skills substitutions in fact occur.

Gender composition - We ask whether employers change the gender composition of job cells as a result of NMW increases. We run regressions with the percentage change in the share of females in job cells as the dependent variable. We do not find that females are systematically replaced by male workers or *vice versa*. The only significant gender-related results are associated with the NMW

[30] The size of the negative effect associated with the 2013 increase is comparable to our firm-level estimate in the same year.

increases in 2015 and 2016, and are economically small.

Job-cell fixed-effects - To inspect job-cell-specific effects, we create a panel of job cells for 2012-2017 and estimate a regression equation with job-cell fixed-effects. Results appear in Table 1.12. We see a small negative employment effect of 5.7%. The size of the estimate is in line with our job-cells results. We also observe a positive effect on wage growth of 10.5%.

Unemployment - To address the concern that employment effects in regions with high unemployment rates may be different to those in regions with low unemployment, we control for county-specific unemployment rates and interact our exposure measures with unemployment rates. We find that the levels of unemployment are important controls only in 2015. Specifically, counties with higher unemployment rates exhibited higher increases in job-cell employment than counties with low unemployment rates. This might suggest that the supply of workers in regions with low unemployment was already depleted due to the economic boom and only firms in counties with high unemployment had opportunities to hire. Detailed results can be found in Table 1.11 in the Appendix.

One large increase in the NMW - The NMW increases we study in the CR are rather small. To simulate a hypothetical situation when the minimum wage level is increased significantly, we combine job-cell data from 2012 and 2017[31], recode our exposure measures, and evaluate one large increase in NMW. We do not find that this artificial increase in NMW had any effect on the employment of low-paid employees (see Table 1.13 for detailed results).

Finally, we would like to stress that we are aware of some aspects that we are not able to control. One is a wage-benefits substitution. It can happen that some employers reduce employee benefits so they can afford to increase their wages (see e.g. Babecký et al., 2019 who show that employers use non-base wage components to adjust labor costs during economic shocks). Another potential issue can be that employers convert regular employee contracts into self-employment (contractor) contracts. This would show as a disemployment effect in our data, although these workers might not loose income.

[31] We use only data on job cells which we observe in both periods, i.e. only job cells that survived all the studied increases are included.

Table 1.1: Employment effects

	Δ log Employment 2012-14			Δ log Employment 2014-15			Δ log Employment 2015-16			Δ log Employment 2016-17		
	Share	Gap	Share (wght)	Share	Gap	Share (wght)	Share	Gap	Share (wght)	Share	Gap	Share (wght)
Firm level	-0.248**	0.0436	-0.232*	0.157**	2.330*	0.114	0.155**	2.817**	-0.0282	-0.0173	0.578	0.112
R2 adj. (n.obs.: 2206; 2182; 2218; 2218)	0.082	0.074	0.188	0.031	0.031	0.204	0.024	0.022	0.274	0.05	0.05	0.191
Job-cell level	-0.106*	-0.229	0.0332	0.029	0.651	0.0494	-0.00498	-0.0288	-0.0999*	-0.0379	-0.375	-0.116
R2 adj. (n.obs.: 6961; 6879; 6902; 7350)	0.034	0.034	0.292	0.021	0.023	0.224	0.012	0.012	0.148	0.03	0.03	0.149
Job-cell level (Firm FE)	-0.131*	-0.481*	-0.0507	-0.0512	0.269	0.0379	-0.0209	-0.0719	0.0112	-0.0397	-0.397	-0.1
R2 adj. (n.obs.: 6961; 6879; 6902; 7350)	0.238	0.237	0.799	0.17	0.17	0.698	0.075	0.075	0.298	0.145	0.146	0.581
5 most affected occupations	-0.044	0.151	0.073	0.0408	2.366***	0.0241	0.0023	0.136	-0.0946	-0.0403	-1.33	-0.0551
R2 adj. (n.obs.: 1940; 1921; 1925; 1950)	0.029	0.029	0.468	0.035	0.055	0.232	0.009	0.009	0.15	0.011	0.019	0.174
Accommodation and food service	-0.534	-9.255	-3.294*	0.0966	1.226	0.498*	-0.0494	-2.613	-0.045	-0.0807	-2.292*	0.0195
R2 adj. (n.obs.: 311; 311; 311; 311)	0.115	0.107	0.824	0.071	0.069	0.242	-0.076	-0.074	0.008	0.211	0.217	0.4
Manufacturing	-0.269***	-1.252***	-0.112	0.355	2.028***	0.604**	0.0256	0.223	0.12	-0.00521	-0.198	0.0444
R2 adj. (n.obs.: 10819; 10819; 10819; 10819)	0.039	0.038	0.212	0.032	0.034	0.186	0.029	0.029	0.115	0.027	0.027	0.083
0 if missing	-0.288***	-0.379*	-0.109	-0.0813*	-0.121	0.00931	-0.176***	0.141	-0.572**	-0.0126	-0.0916	-0.0797
R2 adj. (n.obs.: 14041; 14598; 14747; 12986)	0.069	0.068	0.144	0.035	0.035	0.112	0.085	0.084	0.184	0.016	0.016	0.119
110 pct of MW	-0.0363	0.023	-0.0178	0.0002	0.363	0.0301	-0.015	-0.0323	-0.0976**	-0.0357*	-0.232	-0.0577
R2 adj. (n.obs.: 6961; 6879; 6902; 7350)	0.034	0.034	0.292	0.021	0.022	0.224	0.012	0.012	0.149	0.03	0.03	0.147
LHS: Hours worked	-0.0288	-0.0131	0.106	0.0421	0.925*	0.0407	0.0118	-0.175	-0.0631	-0.0153	-0.223	-0.138
R2 adj. (n.obs.: 6961; 6879; 6902; 7350)	0.045	0.045	0.266	0.025	0.028	0.213	0.019	0.019	0.144	0.024	0.025	0.122
LHS: Turnover	0.141	-0.1	-0.0405	-0.146	-1.697	-0.0947	-0.15	-0.394	0.0499	0.0431	0.0614	0.0772
R2 adj. (n.obs.: 6961; 6879; 6902; 7350)	0	0	0.055	-0.005	-0.005	-0.005	0.002	0.001	0.047	0.05	0.05	0.136
LHS: Δ Education	0.0162	0.229	-0.0222	-0.0166	-0.0352	-0.0321**	0.00629	-0.0292	-0.0134	-0.0265*	-0.358*	-0.0133
R2 adj. (n.obs.: 6862; 6775; 6806; 7282)	0.121	0.121	0.166	0.065	0.065	0.093	0.073	0.073	0.073	0.055	0.055	0.052
LHS: Δ Gender composition	0.002	0.118	0.0162	-0.016	0.0125	-0.0155*	0.0131*	0.0345	-0.0002	0.002	0.126	0.0284
R2 adj. (n.obs.: 6961; 6879; 6902; 7350)	0.007	0.007	0.162	0.01	0.01	0.087	-0.002	-0.002	0.034	0.002	0.003	0.059

Note: The table reports the employment effects of NMW increases for 2013-2017. Columns state the estimated β coefficients from Eq. 3 related to each NMW increase i.e., the estimated coefficients associated with the Share and Gap measures, and the Share measure where linear regressions are weighted by the number of employees within cells or firms. Rows represent different specifications. Controls included: age, length of employment, gender, share of Czech employees, educ. cat, firm size cat., county, industry, occupation (industry substituted by firm in Firm FE regressions). Observations - Job cell level: 1st quartile (except “5 most affected occupations", “Manufacturing", “Accommodation and food service" reg. where are all JCs are used) ; Firm level: all firms.
Share (wght) - Coefficient from a regression weighted by the number of employees within units.
P-values ***0.01, **0.05, *0.1.

1.7 Wage Effects

The next step of our analysis is to inspect how the 2013-2017 NMW increases affected the wages of low-paid workers. Figures 1.1a, 1.1b, 1.1c, 1.1d show visible bunching around new NMW levels, indicating that the NMW increases were binding and, thus, they should have had a positive effect on the wages of low-paid employees. Similarly to the employment effects, we estimate the wage effects of the NMW increases by Equation 1.4

$$\Delta log(cellwage)_{j,c,o,t} = \alpha_2 + \beta_3 * Share_{j,c,o,t-1} + \beta_4 * Gap_{j,c,o,t-1} + \delta_2 * X_{j,c,o,t-1} + \epsilon_{j,c,o,t} \tag{1.4}$$

where the dependent variable is a percentage growth in wages, our coefficients of interest are β and the regression equations include controls for age, length of employment, gender, share of Czech employees, educ. cat, firm size category, county, industry, and occupation. Table 1.2 summarizes the results. Columns represent different exposure measures (the Share, the Gap, and the Share weighted by the number of workers employed within units) for each of the NMW increases. Rows represent various levels of our analysis and econometric specifications. The exposure measures capture wage growth better at the job-cell level, being comparable for all specifications. Our baseline results suggest that NMW increases caused a raise in the wages of directly affected employees by 8%, 9.4%, 2.8%, and 5.5% in 2013, 2015, 2016, and 2017, respectively. Similarly to the interpretation of the employment effects, these are the effects for job cells in which all employees are affected and their wages are increased, i.e. none of them is laid off. The growth in wages caused by the NMW increase is slightly smaller when firm fixed effects are included. The highest estimated coefficients across all specifications are associated with the NMW increase in 2015. Table 1.16 in the Appendix shows the job-cell level β_3 estimates for different parts of the wage distribution in each year. A comparison between years shows that the estimates are the highest in 2014 and the lowest in 2015.

Furthermore, we estimate the wage effects on a subsample of the 5 most affected occupations. The estimates are comparable in size to our baseline specification based on all job cells from the 1st quartile of the job-cell wage distribution.

We also present estimates for *Accomodation and food services* and *Manufacturing* industries. There is anecdotal evidence that restaurant employees are often paid only the minimum wage and the rest of their remuneration is paid off the books. If we accept that this is the case, we would observe a significant increase in wages and no negative effects on employment. However, we do not find this pattern in our data. This may be because the Czech SES contains mostly firms with a higher number of employees, and this type of behavior is not practiced.

Table 1.2: Wage effects

	Δ log Wage 2012-14			Δ log Wage 2014-15			Δ log Wage 2015-16			Δ log Wage 2016-17		
	Share	Gap	Share (wght)	Share	Gap	Share (wght)	Share	Gap	Share (wght)	Share	Gap	Share (wght)
Firm level	0.0299	-0.206***	0.000721	0.0375	0.634	0.0465**	0.0199	0.654***	0.0101	0.0496**	0.736**	0.0398
R2 adj. (n.obs.: 2206; 2182; 2218; 2218)	0.128	0.131	0.236	0.019	0.02	0.204	0.078	0.079	0.251	0.107	0.107	0.207
Job-cell level	0.0804***	0.306**	0.0583	0.0937***	0.516	0.0457**	0.0275***	0.13	0.00506	0.0545***	0.467**	0.0870***
R2 adj. (n.obs.: 6961; 6879; 6902; 7350)	0.087	0.082	0.305	0.099	0.102	0.143	0.092	0.091	0.209	0.12	0.119	0.174
Job-cell level (Firm FE)	0.0603**	0.228*	0.062	0.0753***	0.347	0.0684	0.0403***	0.0986	0.0330***	0.0510***	0.399*	0.0906***
R2 adj. (n.obs.: 6961; 6879; 6902; 7350)	0.318	0.316	0.759	0.393	0.393	0.572	0.334	0.331	0.598	0.374	0.375	0.584
5 most affected occupations	0.0447*	0.621***	0.0812*	0.0779***	1.160***	0.0319***	0.0431***	0.473**	0.0195**	0.0458***	0.812***	0.0737***
R2 adj. (n.obs.: 1940; 1921; 1925; 1950)	0.076	0.08	0.569	0.088	0.13	0.253	0.124	0.123	0.239	0.133	0.155	0.242
Accommodation and food service	0.116	-1.388	0.691*	0.103*	2.154	0.190**	-0.0249	-0.0617	-0.0236	0.00417	0.896	0.0144
R2 adj. (n.obs.: 311; 311; 311; 311)	-0.043	-0.049	0.804	-0.012	-0.024	0.132	0.013	0.012	0.031	0.187	0.196	0.234

Note: The table reports the wage effects of NMW increases for 2013-2017. Columns state the estimated β coefficients from Eq. 3 related to each NMW increase i.e., the estimated coefficients associated with the Share and Gap measures, and the Share measure where linear regressions are weighted by the number of employees within cells or firms. Rows represent different specifications. Controls included: age, length of job, gender, share of Czech employees, educ. cat, firm size cat., county, industry, occupation (industry substituted by firm in Firm FE regressions). Observations - Job cell level: 1st quartile (except "5 most affected occupations", "Manufacturing", "Accommodation and food service" reg. where are all JCs are used) ; Firm level: all firms.

Share (wght) - Coefficient from a regression weighted by the number of employees within units.

P-values ***0.01, **0.05, *0.1.

1.8 Bunching

The third step in our analysis is to apply the bunching estimator, a common tool applied in the economic literature estimating the effects of minimum wage changes (e.g. Meyer and Wise, 1983, Harasztosi and Lindner, 2019, or Cengiz et al., 2019).[32] This approach aims to shed light on employment changes of workers who are paid around a minimum wage threshold. The bunching estimator allows researchers to clarify how the number of "missing" jobs (compared to a counterfactual wage distribution) below a minimum wage threshold relates to jobs added in a new wage distribution above the threshold. The identifying assumption behind the bunching estimator is that the wage distributions would be the same in absence of NMW increases. Usually, counterfactual wage distributions are based on wage distributions in periods prior to NMW changes or they are artificially created as, for example, in Friedman et al. (2011), who applied a polynomial fit to current period distributions. Upper and lower bounds defining the region of interest are set arbitrarily.

Figures 1.1a, 1.1b, 1.1c, 1.1d show that the minimum wage changes in the Czech Republic were small and the bunching around the new NMW levels is moderate. We use a real wage distribution, where wages are discounted by the median growth in wages (with a base in 2013), as the rates of inflation were very low during 2012-2017 (varying in the range of 0.3-3.3) and average wages grew much faster than inflation. We set the upper and lower bounds as + /- 100 EUR (2,500 CZK) around a new NMW level, i.e. we capture the employment of everyone whose monthly wage is in a range of approximately 70 - 130 % of the NMW level in 2013. We use the same approach to choose bounds for subsequent increases. This is in line with Harasztosi and Lindner (2019), who use 20%, 35%, and 50% ranges of the new minimum wage.[33]

Analyzing the 2013 increase, we see that there is approximately 1,300 extra jobs, which means that for every 100 jobs in our range, three new jobs were created.

[32] For a review of the bunching literature see Kleven (2016).

[33] Harasztosi and Lindner (2019) do not use a lower bound because the NMW increase in Hungary amounted to approximately 60%, i.e. setting the lower bound symmetrically around the new NMW level would lead to leaving out some workers. However, the NMW increases in the Czech Republic are much smaller (approximately 37%) and we set the lower bound to exclude potential outliers on the very bottom end of the wage distribution.

The size and direction of the effect related to the NMW increase in 2015 is almost identical. We observe an increase that amounts to approximately 1,300 jobs (for each 100 jobs 3 new ones were created). In 2016 and 2017 there were very small decreases in employment; 1,260 and 1,440 jobs respectively, which means that for every 100 jobs 3 jobs were destroyed due to the 2016 increase and 4 jobs were destroyed due to the 2017 increase. Our results are similar in magnitude to previous research findings. For example, Harasztosi and Lindner (2019) found that a NMW increase in Hungary in 2001 caused 3 out of 100 workers to lose their employment.

Our bunching estimates are small (the highest unemployment estimate based on the bunching estimator is approximately three times smaller than the treatment-intensity estimate on the job-cell level in 2013) and they do not suggest that the NMW increases we study resulted in consistent disemployment effects.

1.9 Conclusion

This paper studies four recent increases in the NMW in the Czech Republic from 2013 to 2017. Constructing exposure measures similar to Machin et al. (2003), we inspect the effects of the NMW increases on employment and wages on a job-cell level. Compared to previous studies, which work with firm-level observations, we are able to study exposure more precisely for homogenous groups of workers. We show that the application of the treatment intensity estimator on heterogenous groups of people may lead to biased estimates.

Our findings suggest that the NMW increases in our study had positive effects on wage growth of low-paid workers and had no or small negative effects on employment. In our preferred specification, the employment elasticities with respect to the minimum wage vary between -0.0093 and 0.0017, which are rather small estimates compared to others in the existing literature. Our results are confirmed by several robustness tests.

1.10 Appendix

Wage measure

Wage measures available in our dataset are based on the total money paid to employees by employers. However, for the purpose of minimum wage analysis, it is necessary to clean the wage data in order to include only the amount of wages that are directly regulated by the minimum wage law[34]. The monthly wage available in the dataset is defined as

$$\text{monthly wage} = \frac{\text{money paid since January } 1^{\text{st}}}{\text{number of months worked since January } 1^{\text{st}}} \tag{1.5}$$

We strictly follow the minimum wage legislation and adjust the nominator according to the definition of wage in Government decree no. 89/2012 Sb. We also adjust the number of months worked since January 1^{st} so that it does not include overtime hours. This definition is as close as possible to the definition in legislation.

[34] This wage measure does not include overtime pay, extra pay for hard work, etc. For more information, please see Government decree no. 89/2012 Sb.

Tables

Table 1.3: Employees affected by NMW increases

year	Number of affected by MW in sample	Percent in sample	Fraction in firms with < 100 employees	Mean distance to the new MW level (in CZK)	Median distance to the new MW level (in CZK)
2012	21,659	1.77 %	0.52	890.5	420.9
				(1331.8)	
2014	10,560	0.85 %	0.54	583.2	427.6
				(649.2)	
2015	26,548	2.09 %	0.51	738.3	613.7
				(700.2)	
2016	30,507	2.37 %	0.56	890.1	785.0
				(744.7)	

* Standard deviations in parenthesis.

Note: The table reports descriptive statistics on employees who were / were not exposed to NMW increases in each year. The last two columns report the distances (among employees affected by the NMW increases) to the NMW level in the next period.

Table 1.4: Individual level characteristics - percentages of employees affected by NMW

	Category	2012	2014	2015	2016
Gender					
	Male	1.45	0.69	1.69	1.90
	Female	2.22	1.09	2.64	3.00
Age Group					
	<20	4.01	1.36	4.19	3.46
	21-30	1.91	0.61	1.97	2.22
	31-40	1.73	0.56	1.77	2.00
	41-50	1.56	0.87	1.89	2.17
	51-60	1.67	1.35	2.57	2.96
	>61	3.34	1.63	3.72	3.93
Education					
	Primary	2.84	2.98	4.77	5.42
	Apprenticeship	1.81	1.19	2.38	2.85
	Secondary	1.46	0.39	1.44	1.66
	College	1.52	0.12	1.24	1.41
	Post-graduate	1.69	0.06	1.25	1.14
Tenure in the job					
	<1 year	3.57	2.27	4.08	4.17
	1-5 years	2.25	1.26	2.95	3.32
	>5 years	1.16	0.36	1.18	1.40

Note: The table reports percentages of employees who were exposed to NMW increases in each year (by various characteristics).

Table 1.5: Occupations: percentages of affected employees in each occupational group and year

ISCO group	2012	2014	2015	2016
Managers	0.87	0.11	0.80	0.97
Professionals	1.35	0.02	1.11	1.34
Technicians and Associate Professionals	1.16	0.06	0.86	1.01
Clerical Support Workers	1.05	0.20	1.03	1.39
Services and Sales Workers	4.63	3.52	6.98	6.27
Skilled Agricultural, Forestry and Fishery Workers	3.01	0.48	2.12	2.38
Craft and Related Trades Workers	1.24	0.19	1.02	1.33
Plant and Machine Operators and Assemblers	1.40	0.43	1.28	1.66
Elementary Occupations	4.92	6.47	9.10	11.35

Note: The table reports percentages of workers affected by NMW increases in each year and occupational group (2 digit ISCO classification).

Table 1.6: Industry: percentages of affected employees in each industry group and year

NACE group	2012	2014	2015	2016
Aggr., Forest., Fish.	2.15	0.60	1.78	2.16
Mining and Metalurgy	1.13	0.16	0.50	0.80
Manufacturing	1.13	0.33	1.04	1.27
Utilities	0.45	0.40	0.71	0.84
Construction	1.18	0.24	1.70	1.57
Retail	2.10	0.51	2.90	2.30
Hotels, food serving	7.70	4.05	5.37	6.72
Transport	1.49	0.08	0.92	1.22
Banks, insurance	1.14	0.02	0.74	1.03
Real Estate, R&D	6.06	6.41	9.35	10.81
Public Admin, defense	0.39	0.00	1.22	0.64
Education	0.86	0.38	1.49	1.43
Health	1.72	1.94	3.36	5.15
Other Services	3.32	1.43	3.13	4.24
Communications	1.30	0.01	1.22	1.28

Note: The table reports percentages of employees affected by NMW increases in each year and industry (2 digit NACE classification).

Table 1.7: Job-cell characteristics

	2012		2014		2015		2016	
	Not affected	Affected	Not affected	Affected	Not affected	Affected	Not affected	Affected
Share	-	0.1870	-	0.3679	-	0.2302	-	0.2532
		(0.214)		(0.319)		(0.261)		(0.287)
Intensity measure	-	0.0078	-	0.0263	-	0.008	-	0.0126
		(0.103)		(0.129)		(0.035)		(0.041)
JC employement	18.3	66.4	22.08	45.2	18.6	64.4	18.5	64.6
	(59.7)	(174.7)	(76.7)	(125.3)	(62.7)	(166.6)	(57.2)	(183.5)
Net hrs worked	1655	1549	1642	1503	1619	1517	1619	1508
	(314)	(404)	(325)	(444)	(342)	(419)	(348)	(431)
Female	0.419	0.521	0.431	0.594	0.425	0.543	0.426	0.532
	(0.39)	(0.36)	(0.39)	(0.36)	(0.39)	(0.36)	(0.39)	(0.36)
Age	41.9	41.3	42	44.3	42.3	41.7	42.3	42.2
	(7.0)	(6.5)	(7.0)	(6.6)	(7.0)	(6.7)	(6.9)	(6.8)
Tenure in the job	9.3	7.4	9.5	5.7	9.7	7.0	9.5	7.1
	(6.8)	(5.5)	(6.7)	(5.9)	(6.8)	(5.1)	(6.7)	(5.71)
Primary educ.	0.043	0.082	0.041	0.180	0.040	0.095	0.040	0.098
	(0.12)	(0.15)	(0.11)	(0.23)	(0.11)	(0.17)	(0.11)	(0.17)
Apprenticeship educ.	0.311	0.360	0.301	0.513	0.295	0.376	0.287	0.369
	(0.34)	(0.32)	(0.33)	(0.30)	(0.33)	(0.32)	(0.33)	(0.32)
Secondary educ.	0.406	0.341	0.406	0.200	0.411	0.319	0.408	0.320
	(0.32)	(0.28)	(0.31)	(0.23)	(0.31)	(0.27)	(0.31)	(0.27)
Tertiary educ.	0.203	0.159	0.215	0.041	0.220	0.153	0.228	0.160
	(0.29)	(0.25)	(0.30)	(0.13)	(0.30)	(0.25)	(0.30)	(0.25)
Czech nationality	0.975	0.957	0.973	0.945	0.973	0.948	0.970	0.943
	(0.09)	(0.11)	(0.09)	(0.16)	(0.09)	(0.13)	(0.09)	(0.14)
N	46,867	5,110	51,491	1,559	47,715	5,507	47,391	6,011

* Standard deviations in parenthesis.
Note: The table reports characteristics of job cells that were / were not exposed to the NMW increases in each year. “Not affected" are the job cells in which all employees were paid above the NMW level effective in the next period. Similarly, “Affected" job cells contain at least one employee paid below the NMW level in the next period.

Table 1.8: Firm characteristics

	2012		2014		2015		2016	
	Not affected	Affected	Not affected	Affected	Not affected	Affected	Not affected	Affected
Share	-	0.0624	-	0.1154	-	0.0827	-	0.0920
		(0.112)		(0.189)		(0.152)		(0.170)
Intensity measure	-	0.0016	-	0.0059	-	0.0027	-	0.0039
		(0.034)		(0.015)		(0.011)		(0.015)
Firm employment	200.9	459.2	297.4	396.2	208.6	441.4	211.1	440.3
	(328.6)	(1335.2)	(840.9)	(1288.3)	(394.6)	(1262.4)	(358.5)	(1250)
Net hrs worked	1659	1605	1634	1582	1624	1575	1622	1586
	(332)	(367)	(344)	(410)	(358)	(394)	(355)	(392)
Female	0.374	0.444	0.401	0.506	0.386	0.456	0.391	0.455
	(0.25)	(0.26)	(0.25)	(0.27)	(0.26)	(0.25)	(0.25)	(0.26)
Age	41.2	41.2	41.2	42.6	41.7	41.5	41.9	41.5
	(4.6)	(4.7)	(4.8)	(4.7)	(4.7)	(4.7)	(4.8)	(4.6)
Tenure in the job	8.1	7.8	8.2	7.5	8.3	7.6	8.3	7.7
	(5.7)	(5.2)	(5.1)	(6.5)	(4.7)	(4.7)	(4.9)	(5.4)
Primary educ.	0.053	0.071	0.048	0.103	0.048	0.070	0.045	0.070
	(0.097)	(0.105)	(0.079)	(0.154)	(0.086)	(0.113)	(0.081)	(0.112)
Apprenticeship educ.	0.403	0.393	0.363	0.435	0.372	0.378	0.354	0.363
	(0.254)	(0.239)	(0.251)	(0.240)	(0.260)	(0.242)	(0.258)	(0.239)
Secondary educ.	0.340	0.328	0.353	0.303	0.349	0.333	0.351	0.334
	(0.179)	(0.166)	(0.178)	(0.179)	(0.184)	(0.172)	(0.182)	(0.175)
Tertiary educ.	0.162	0.156	0.193	0.108	0.194	0.167	0.202	0.180
	(0.191)	(0.182)	(0.209)	(0.134)	(0.213)	(0.189)	(0.219)	(0.201)
Czech nationality	0.967	0.956	0.958	0.945	0.963	0.945	0.957	0.940
	(0.075)	(0.094)	(0.088)	(0.148)	(0.083)	(0.118)	(0.092)	(0.121)
N	1,775	1,880	3,253	662	1,903	1,968	1,717	2,100

Standard deviations in parentheses.

Note: The table reports characteristics of firms that were / were not exposed to the NMW increases in each year. “Not affected" are firms in which all employees were paid above the NMW level effective in the next period. Similarly, “Affected" firms contain at least one employee paid below the NMW level in the next period.

Table 1.9: Employment effects of NMW increase: 2012-2014 data, firm level

	All firms			Firms with > 100 employees		
	Share	Gap	Weighted	Share	Gap	Weighted
Share	-0.248**		-0.241*	-0.189		-0.240*
	(0.113)		(0.126)	(0.136)		(0.134)
Gap		0.0427			0.0506	
		(0.107)			(0.103)	
Age	0.0191	0.0194	0.0653***	0.0567*	0.0570*	0.0742***
	(0.0283)	(0.0284)	(0.0250)	(0.0302)	(0.0303)	(0.0278)
Age sq.	-0.0287	-0.0298	-0.0855***	-0.0718**	-0.0731**	-0.0951***
	(0.0349)	(0.0350)	(0.0296)	(0.0354)	(0.0357)	(0.0330)
Tenure in the job	-0.00422	-0.00388	-0.0120**	-0.0180**	-0.0171**	-0.0205***
	(0.00257)	(0.00257)	(0.00576)	(0.00735)	(0.00732)	(0.00637)
Tenure sq.	0.000138***	0.000140***	0.000508**	0.000761**	0.000743**	0.000866***
	(0.0000308)	(0.0000305)	(0.000220)	(0.000316)	(0.000316)	(0.000242)
Female	0.00733	0.000303	0.0502	0.0397	0.0361	0.0494
	(0.0464)	(0.0454)	(0.0436)	(0.0589)	(0.0583)	(0.0470)
Czech	0.0304	0.0640	0.117	0.0961	0.126	0.129
	(0.129)	(0.131)	(0.108)	(0.116)	(0.121)	(0.113)
Constant	-0.193	-0.237	-1.153**	-1.170*	-1.195*	-1.391**
	(0.629)	(0.628)	(0.507)	(0.612)	(0.617)	(0.550)
Educ cat	Yes	Yes	Yes	Yes	Yes	Yes
Firm size	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes
County	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,206	2,206	2,206	1,605	1,605	1,605
Adjusted R^2	0.082	0.077	0.198	0.079	0.076	0.187

Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.010

Note: The table reports the employment effects (Dep. var: Δ logEmpl) of the NMW increase in 2013 using firm-level data from 2012 and 2014. Columns represent different exposure measures. Columns (1)-(3) consider all firms, columns (4)-(6) consider only firms that include more than 100 employees.

Table 1.10: Employment effects of NMW increases: 2012-2014 data, all job cells from the Manufacturing industry

	All cells			Cells > 10		
	Share	Gap	Weighted	Share	Gap	Weighted
Share	-0.269***		-0.112	-0.212		-0.0631
	(0.103)		(0.163)	(0.179)		(0.180)
Gap		-1.252***			-0.942***	
		(0.128)			(0.166)	
Age	0.0269***	0.0270***	0.0749***	0.0573***	0.0578***	0.0976***
	(0.00757)	(0.00755)	(0.0213)	(0.0170)	(0.0169)	(0.0288)
Age sq.	-0.0326***	-0.0327***	-0.0977***	-0.0765***	-0.0771***	-0.129***
	(0.00875)	(0.00873)	(0.0259)	(0.0204)	(0.0203)	(0.0355)
Tenure in the job	-0.0144***	-0.0142***	-0.0140**	-0.0144***	-0.0141***	-0.0155**
	(0.00266)	(0.00266)	(0.00582)	(0.00523)	(0.00525)	(0.00749)
Tenure sq.	0.000446***	0.000443***	0.000704***	0.000611***	0.000603***	0.000848***
	(0.0000757)	(0.0000756)	(0.000194)	(0.000175)	(0.000175)	(0.000266)
Female	-0.0252	-0.0253	-0.138**	-0.0216	-0.0218	-0.137**
	(0.0184)	(0.0185)	(0.0574)	(0.0327)	(0.0326)	(0.0623)
Czech	-0.0340	-0.0312	-0.0521	0.161*	0.164*	-0.0343
	(0.0666)	(0.0668)	(0.143)	(0.0948)	(0.0946)	(0.161)
Constant	-0.543***	-0.550***	-2.180***	-1.815***	-1.826***	-3.213***
	(0.173)	(0.173)	(0.649)	(0.364)	(0.364)	(0.820)
Educ cat	Yes	Yes	Yes	Yes	Yes	Yes
Firm size	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes
County	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,819	10,819	10,819	5,438	5,438	5,438
Adjusted R^2	0.039	0.038	0.212	0.061	0.061	0.224

Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.010.

Note: The table reports employment effects (Dep. var: Δ logEmpl) of the 2013 NMW increase estimated on the job cells that belong to the Manufacturing industry. We use data from 2012 and 2014. Columns represent different exposure measures. Columns (1)-(3) consider all job cells, columns (4)-(6) consider only job cells that include more than 10 employees.

Table 1.11: Employment effects - interactions with local unemployment rates

	All cells		Cells > 10	
	Share	Gap	Share	Gap
2012				
Share	-0.181		-0.558	
	(0.225)		(0.373)	
Shr*Unempl	0.00803		0.0456	
	(0.0218)		(0.0385)	
Gap		0.127		-1.846
		(1.691)		(2.694)
Gap * Unempl		-0.0391		0.217
		(0.180)		(0.287)
Unemployment	0.00311	0.00377	0.00580	0.00778
	(0.00898)	(0.00909)	(0.0128)	(0.0130)
2014				
Share	-0.0798		0.00313	
	(0.152)		(0.224)	
Shr*Unempl	0.0121		0.00743	
	(0.0143)		(0.0261)	
Gap		-0.505		5.281**
		(1.170)		(2.157)
Gap * Unempl		0.165		-0.479*
		(0.171)		(0.278)
Unemployment	-0.0000338	-0.0000157	-0.0103	-0.00668
	(0.00610)	(0.00590)	(0.0136)	(0.00744)
2015				
Share	-0.0519		-0.0490	
	(0.0876)		(0.122)	
Shr*Unempl	0.00620		0.00211	
	(0.0107)		(0.0151)	
Gap		-0.402		1.107
		(0.344)		(2.120)
Gap * Unempl		0.0642		-0.0491
		(0.0598)		(0.265)
Unemployment	0.0128***	0.0130***	0.0200***	0.0203***
	(0.00423)	(0.00419)	(0.00692)	(0.00683)
2016				
Share	-0.0290		-0.276**	
	(0.0644)		(0.122)	
Shr*Unempl	-0.00136		0.0212	
	(0.00881)		(0.0137)	
Gap		0.901		-5.264**
		(0.585)		(2.509)
Gap * Unempl		-0.253*		0.407*
		(0.154)		(0.233)
Unemployment	-0.000322	0.000557	-0.00168	-0.00157
	(0.00440)	(0.00430)	(0.00723)	(0.00717)
Controls	Yes	Yes	Yes	Yes

Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.010.

Note: The table reports employment effects (Dep. var: Δ logEmpl) using the interactions of exposure measures and county-specific unemployment rates in each year, i.e., we ask whether the employment effect is stronger in counties experiencing high unemployment rates. Columns represent different exposure measures. Columns (1)-(2) consider all job cells from the 1st quartile of job-cell wage distribution, columns (3)-(4) consider only job cells from the 1st quartile that consist of more than 10 employees. Controls include age, age sq., tenure in the job, tenure sq., county, firm size, and shares of females, Czechs, workers with the highest education attained.

Table 1.12: Employment effects with job-cell fixed effects

	Δ logEmpl		Δ logWage	
	(1)	(2)	(1)	(2)
Share	-0.0572***		0.105***	
	(0.0194)		(0.00825)	
Gap		-0.471**		0.584***
		(0.228)		(0.0969)
Age	-0.0277***	-0.0276***	0.00000390	-0.000328
	(0.00487)	(0.00487)	(0.00167)	(0.00167)
Age sq.	0.0285***	0.0282***	0.00575***	0.00626***
	(0.00589)	(0.00589)	(0.00200)	(0.00201)
Tenure in the job	-0.0307***	-0.0307***	0.00627***	0.00635***
	(0.00206)	(0.00206)	(0.000747)	(0.000750)
Tenure sq.	0.000297***	0.000297***	-0.0000689***	-0.0000688***
	(0.0000237)	(0.0000237)	(0.0000144)	(0.0000142)
Female	0.0635*	0.0640*	-0.0738***	-0.0743***
	(0.0375)	(0.0375)	(0.0106)	(0.0106)
Czech nat.	-0.0984	-0.0976	-0.146***	-0.146***
	(0.0607)	(0.0608)	(0.0198)	(0.0199)
Constant	2.542***	2.530***	9.527***	9.647***
	(0.153)	(0.154)	(0.0445)	(0.0785)
Educ cat	Yes	Yes	Yes	Yes
Firm size	Yes	Yes	Yes	Yes
Industry	Yes	Yes	No	Yes
Observations	60,441	60,441	60,441	60,441
Adjusted R^2	0.053	0.053	0.067	0.062

Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.010.

Note: The table reports the employment and wage effects (Dep. var: Δ logEmpl, Δ logWage) of the NMW increases. We estimate linear regressions on job-cell panel data for 2012-2017. The job-cell fixed effects are included. The first two rows are our coefficients of interest. We use all job cells.

Table 1.13: Employment effects: one large artificial NMW increase between 2012-2017

	All cells			Cells > 10		
	Share	Gap	Weighted	Share	Gap	Weighted
Share	-0.0414		-0.0740	-0.0529		-0.0480
	(0.0300)		(0.0869)	(0.0602)		(0.103)
Gap		0.116			0.121	
		(0.132)			(0.165)	
Age	0.0185***	0.0185***	0.0668***	0.0346***	0.0346***	0.0881***
	(0.00466)	(0.00465)	(0.0210)	(0.0125)	(0.0125)	(0.0292)
Age sq.	-0.0238***	-0.0239***	-0.0875***	-0.0526***	-0.0528***	-0.119***
	(0.00538)	(0.00538)	(0.0244)	(0.0147)	(0.0147)	(0.0346)
Tenure in the job	-0.0114***	-0.0113***	0.00249	0.00200	0.00228	0.0105
	(0.00280)	(0.00278)	(0.00509)	(0.00401)	(0.00401)	(0.00695)
Tenure sq.	0.000341***	0.000340***	0.000104	0.0000567	0.0000507	-0.000101
	(0.0000993)	(0.0000985)	(0.000160)	(0.000131)	(0.000131)	(0.000232)
Female	-0.0562***	-0.0569***	-0.226***	-0.0693***	-0.0708***	-0.235***
	(0.0121)	(0.0121)	(0.0573)	(0.0263)	(0.0263)	(0.0661)
Czech	0.0345	0.0368	-0.0335	0.0181	0.0247	-0.0523
	(0.0466)	(0.0466)	(0.132)	(0.0930)	(0.0934)	(0.153)
Constant	-0.383***	-0.390***	-1.347***	-0.884***	-0.892***	-1.759***
	(0.112)	(0.112)	(0.479)	(0.330)	(0.330)	(0.665)
Educ cat	Yes	Yes	Yes	Yes	Yes	Yes
Firm size	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes
County	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes	Yes	Yes
Observations	34,002	34,002	34,002	14,401	14,401	14,401
Adjusted R^2	0.030	0.030	0.147	0.056	0.055	0.166

Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.010.

Note: The table reports the employment effects (Dep. var: Δ logEmpl) of an artificial increase in the NMW between 2012 and 2017. We use all job cells that survived from 2012 to 2017. Columns represent different exposure measures. Columns (1)-(3) consider all job cells, columns (4)-(6) consider only job cells that consist more than 10 employees.

Table 1.14: Wage effects: one large artificial NMW increase between 2012-2017

	All cells			Cells > 10		
	Share	Gap	Weighted	Share	Gap	Weighted
Share	0.0893***		0.0644***	0.0744***		0.0566**
	(0.0108)		(0.0239)	(0.0183)		(0.0274)
Gap		0.286			0.111	
		(0.179)			(0.165)	
Age	-0.0140***	-0.0142***	-0.0301***	-0.0197***	-0.0200***	-0.0385***
	(0.00166)	(0.00166)	(0.00685)	(0.00322)	(0.00322)	(0.00931)
Age sq.	0.0142***	0.0143***	0.0332***	0.0215***	0.0219***	0.0435***
	(0.00197)	(0.00197)	(0.00760)	(0.00382)	(0.00382)	(0.0105)
Tenure in the job	-0.00247***	-0.00252***	-0.00315***	-0.00237**	-0.00259***	-0.00369**
	(0.000487)	(0.000491)	(0.00112)	(0.000983)	(0.000985)	(0.00147)
Tenure sq.	0.0000323**	0.0000323**	0.0000786**	0.0000499	0.0000538*	0.000110**
	(0.0000133)	(0.0000135)	(0.0000327)	(0.0000309)	(0.0000310)	(0.0000453)
Female	0.0464***	0.0473***	0.0622***	0.0524***	0.0540***	0.0661***
	(0.00427)	(0.00427)	(0.00800)	(0.00619)	(0.00617)	(0.00946)
Czech	0.0348*	0.0362**	-0.0146	-0.0128	-0.0134	-0.0175
	(0.0181)	(0.0182)	(0.0259)	(0.0246)	(0.0248)	(0.0294)
Constant	0.495***	0.498***	0.878***	0.543***	0.548***	0.985***
	(0.0489)	(0.0489)	(0.150)	(0.152)	(0.152)	(0.229)
Educ cat	Yes	Yes	Yes	Yes	Yes	Yes
Firm size	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes
County	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes	Yes	Yes
Observations	34,002	34,002	34,002	14,401	14,401	14,401
Adjusted R^2	0.085	0.084	0.186	0.124	0.122	0.215

Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.010.

Note: The table reports the wage effects (Dep. var: Δ logCellWage) of an artificial increase in the NMW between 2012 and 2017. We use only job cells that survived from 2012 to 2017. Columns represent different exposure measures. Columns (1)-(3) consider all job cells, columns (4)-(6) consider only job cells that consist of more than 10 employees.

Table 1.15: Employment effects for low- and high-wage employees: firm-level exposure

	Δ log Employment	
	low-paid employees	high-paid employees
2013 NMW increase	0.007	0.002
	(0.025)	*(0.007)*
2015 NMW increase	0.007	0.010*
	(0.005)	*(0.005)*
2016 NMW increase	-0.002	0.004
	(0.003)	*(0.004)*
2017 NMW increase	0.011	0.004
	(0.012)	*(0.004)*

Note: The table reports the Share estimates from regression equations where the dependent variables are $\Delta logEmpl$ (firm level) and independent variables are the Share measures at the level of firms during 2013-2017. The employment changes include only employees who belong to the first quartile of the job-cell wage distribution in the first column and employees from the 2nd-4th quartile in the second column. Rows show estimated effects for different NMW increases. Controls at the firm level are age, age sq., tenure in the job, tenure sq., county, firm size, industry, and shares of females, Czechs, and the highest education attained, and they are the same for both subsamples; Standard errors in parenthesis, p-values * $p<0.10$, ** $p<0.05$, *** $p<0.010$.

Table 1.16: Wage effects for different subsamples of job cells

	All observations	Below median	1st quartile	10 most affected occ
2012				
Share	0.0359**	0.0646***	0.0733***	0.0496**
	(0.0174)	(0.0190)	(0.0218)	(0.0204)
2014				
Share	0.0906***	0.0916***	0.0874***	0.0714***
	(0.0284)	(0.0286)	(0.0293)	(0.0205)
2015				
Share	0.0215**	0.0245***	0.0280***	0.0464***
	(0.00935)	(0.00859)	(0.00792)	(0.0108)
2016				
Share	0.0481***	0.0577***	0.0491***	0.0443***
	(0.00874)	(0.00852)	(0.00748)	(0.00895)
Controls	Yes	Yes	Yes	Yes

Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.010.

Note: The table reports β_3 coefficients from Eq. 4 for various subsamples of job cells based on the job-cell wage distribution in each year. Controls include age, age sq., tenure in the job, tenure sq., county, firm size, and shares of females, Czechs, and highest education attained.

Table 1.17: Wage cyclicality in firms with low-paid employees

	Low-paid: 1st decile	Low-paid: 1st quartile	Low-paid: below median
Pcnt. change in ind. prod.	0.000798***	0.000501**	-0.000386
	(0.000177)	(0.000234)	(0.000386)
Share of low-paid	0.602***	0.579***	0.632***
	(0.0298)	(0.0204)	(0.0204)
Ind. growth*shr of low-paid	-0.00215***	-0.000625	0.000829
	(0.000757)	(0.000503)	(0.000539)
Constant	-0.243	-0.336	-0.641***
	(0.228)	(0.226)	(0.246)
Observations	11,130	11,130	11,130
Adjusted R^2	0.170	0.238	0.278

Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.010.

Note: The table reports results on wage cyclicality of firms with varying proportions of low-paid employees across the economic cycle. The dependent variable is $\Delta log Wage$. The percentage change in industrial production approximates the phase of the economic cycle. We are primarily interested in the estimated coefficients *"Ind. growth*shr of low-paid"*, which are the interactions of the industry growth and the share of low-paid employees. These estimates allow us to determine whether the wage growth in firms with/without higher shares of low-paid employees systematically differs during economic booms and busts.. Columns represent different measures of low-paid employees. Controls include age, age sq., tenure in the job, tenure sq., county, firm size, and shares of females, Czechs, and highest education attained. We use semiannual data from SES 2007-2012, all firms. The regressions contains firm fixed effects.

Table 1.18: Employment cyclicality in firms with low-paid employees

	Low-paid: 1st decile	Low-paid: 1st quartile	Low-paid: below median
Pcnt. change in ind. prod.	0.00117***	0.00113**	0.00182**
	(0.000403)	(0.000489)	(0.000711)
Share of low-paid	0.0662	-0.0896	-0.129***
	(0.0801)	(0.0602)	(0.0492)
Ind. growth*shr of low-paid	0.00113	0.000967	-0.000499
	(0.00273)	(0.00142)	(0.00114)
Constant	-1.349**	-1.392**	-1.327*
	(0.683)	(0.677)	(0.687)
Controls	Yes	Yes	Yes
Observations	11,130	11,130	11,130
Adjusted R^2	0.101	0.101	0.102

Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.010.

Note: The table reports results on employment cyclicality of firms with varying proportions of low-paid employees across the economic cycle. The dependent variable is $\Delta logEmpl$. The percentage change in industrial production approximates the phase of the economic cycle. We are primarily interested in the estimated coefficients *"Ind. growth*shr of low-paid"*, which are the interactions of the industry growth and the share of low-paid employees. These estimates allow us to determine whether the employment patterns in firms with/without higher shares of low-paid employees systematically differ during economic booms and busts. Columns represent different measures of low-paid employees. Controls include age, age sq., tenure in the job, tenure sq., county, firm size, and shares of females, Czechs, and highest education attained. We use semiannual data from SES 2007-2012. The regressions contains firm fixed effects.

Chapter 2

Sick Pay and Absence from Work: Evidence from Flu Exposure

2.1 Introduction

Most European countries provide workers with sick-pay benefits if they experience a temporary sickness. Such insurance allows sick employees to pay their regular expenses when they are not able to work. In the optimal scenario, employees stay home when they are ill and avoid spreading their disease to coworkers, and return to work once they are healthy, without shirking. They also do not experience adverse long-term health effects from not treating their diseases.[1] The design of sick-pay programs have a substantial influence on employees' decisions to stay at home/go to work. Decreases in sick-pay benefits may either reduce shirking or result in employees going to work when they are sick, a typical moral hazard problem. Thus, it is important to understand and measure the effects of sick-pay programs, as they are associated with significant financial and health costs.[2]

[1] For example, literature finds a strong association between the recent incidence of respiratory infections and major cardiovascular events (Clayton et al., 2008).

[2] For example, money spent on sick pay policies amounted to 1.13 billion EUR (for simplicity, we use a conversion rate 25 CZK/ 1 EUR through this paper) - 28,3 billion Czech crowns (CZK)

The literature on workplace absences has so far focused on how individual characteristics affect the probability of being absent from work (e.g., Barmby, 2002; Scoppa, 2010; De Paola, 2010; Amuri, 2011), and analyzed the effects of changes in sick-pay programs (e.g., Ziebarth and Karlsson, 2010; Pettersson-Lidbom and Thoursie, 2013; De Paola et al., 2014; Pertold, 2019). Still, little is known about the relationship between actual sickness rates and absences from work. While a few studies consider health status of workers, (for example, see Ichino and Moretti, 2009; Herrmann and Rockoff, 2010) no study has explicitly linked the relationship between health status, rates of absence from work and how this is affected by changes in sick-pay programs.

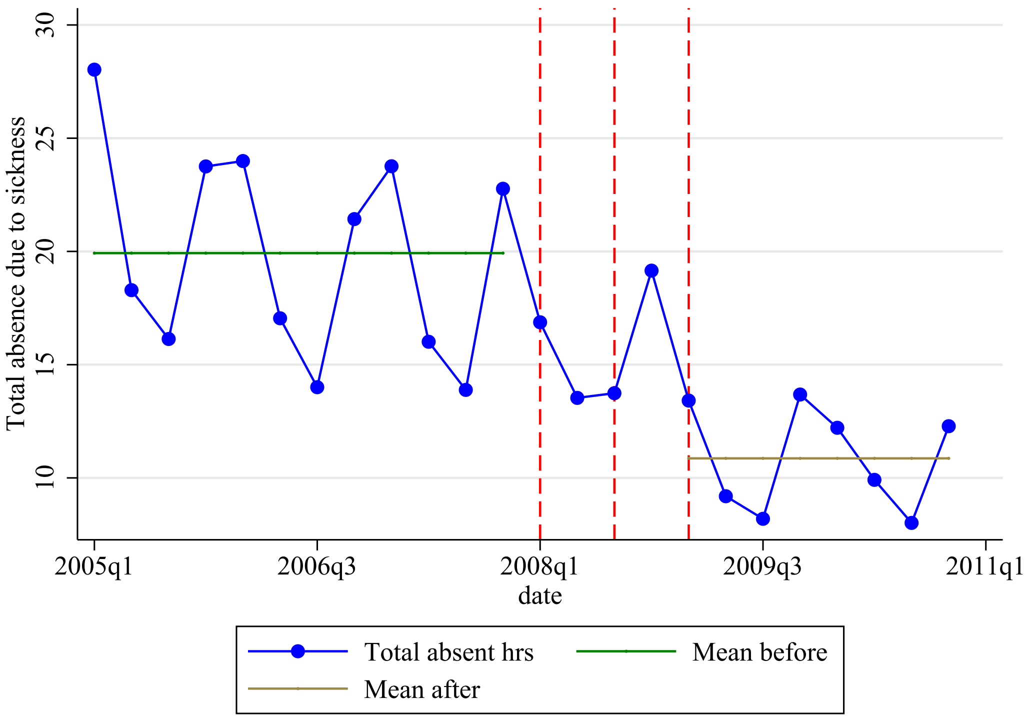
In this paper, we link local sickness rates to employees' records from the Czech Structure of Earnings Survey (SES), in order to study the impacts of two legislative changes in the Czech sick-pay program during 2008-2009 on absence rates of employees exposed to local infection outbreaks. These policy changes canceled the benefits that participants of the Czech sick-pay program previously received during the first three days of sick-leave.[3] Our analysis measures how the policy changes affected hours absent of employees differently exposed to infectious diseases. A significant decrease in the average sickness-related absences of private-sector employees in Graph 2.1 (page 46) indicates that the changes did affect the behavior of employees. Our results show that employees took more sickness-related absences when exposed to outbreaks of infectious diseases before 2008 than after the reforms. However, the abolition of sick-pay benefits motivated them to use paid and unpaid leave instead of sick-leave, and resulted in the total hours of absence remaining at their initial levels. To the best of our knowledge, this is the first study that shows how sickness rates relate to absence from work and how changes in the sick-pay system affect the behavior of employees exposed to outbreaks of infectious diseases.

We begin our analysis by showing that there is a positive relationship between

in the Czech Republic in 2017, which is approximately a 3.6 times higher number than was spent on unemployment policies in that year (Ministry of Labour and Social Affairs, 2017).

[3]The legislative changes are described in Section 2.2. The first three days of sickness are called a 'quarantine period' or 'waiting period'.

Figure 2.1: Sickness absence hours



Note: The graph shows the average quarterly hours of absence per employee before and after the reforms. The red vertical lines indicate the timing of the legislative changes.

the number of hours of absence and local sickness rates. The omission of variables that control for local sickness levels does not necessarily change the explanatory power of the estimated models. However, it may change the size of other estimated coefficients of interest, e.g., the controls for periods after the reforms, which are widely interpreted as the effects of policy reforms in a before-after comparison framework (e.g., De Paola et al., 2014; Pertold, 2019).[4] We first estimate the overall effects of policy reforms using a 'before-after' comparison. A disadvantage of this technique is that the estimates may capture other effects that are unrelated to the policy changes. Second, we focus on a specific part of the policy reforms' overall effect, and apply an intensity treatment / difference-in-differences identification strategy to estimate how the legislative changes affected absences of employees exposed to specific infectious diseases.[5,6] The *intensity treatment*

[4] For a demonstration see Table 2.10.

[5] We cannot apply the standard difference-in-differences estimator as the policy changes affected all employees in the Czech economy.

[6] In most of the estimated specifications, we use a measure of sickness where we count the

here refers to an intensity of the need for sickness insurance, i.e., those who were exposed to influenza outbreaks needed the sickness insurance more than those with no exposure. In this setup, the policy reforms did not change the need for sickness insurance among those who were shirking or suffered different diseases than influenza; hence, we strictly focus on changes in hours of absence caused by the lowering of sickness benefits for employees suffering from a specific infectious disease.[7] Using this approach, we estimate very specific local average treatment effects, which bring important insights into the spreading of disease in the workplace. Provided that we observe complete records of different types of hours of absence for each employee, we also bring qualitative evidence on the mechanism behind exposure to influenza and policy reforms.

Our baseline results suggest that one extra week of influenza outbreak among adults prior to 2008 caused an increase in *sickness-related absences* by 5 working days and a decrease in *unpaid leave* by 2 working days. Compared to the situation before 2008, the legislative changes resulted in one extra week of influenza outbreak among adults, decreasing the *sickness-related absences* by 6.7 days (the overall effect becomes negative but statistically insignificant)[8] and increasing *paid* and *unpaid leave* by 1.6 and 2.3 days, respectively (the overall effect became positive). This pattern suggests that as a result of the policy changes, employees exposed to outbreaks of influenza almost perfectly substituted different types of hours of absence (sickness-related absences, un/paid leave). However, the total effect of exposure to sickness among adults remained economically insignificant. We observe a similar substitution pattern when we use a sickness measure that counts the number of weeks with influenza epidemic status among children.[9]

These are important findings regarding moral hazard. One concern could be

number of weeks with influenza epidemic status among children, assuming that children can infect adults but not vice versa.

[7]Depending on what diseases we use as the exposure measure (influenza, other infectious diseases).

[8]By the “overall effect" we mean the effect of disease exposure in periods after 2008, i.e, the combination of β and γ coefficients from Equation 2.1 presented in Section 2.4.

[9]The size of these effects is smaller in the case of *sickness-related absences* but of similar magnitude for *paid* and *unpaid leave.* One extra week of influenza outbreak in periods before 2008 caused an increase in *sickness-related absences* by 1 day and a decrease in *paid* and *unpaid leave* by 1.8 and 1.6 days respectively. The legislative changes enacted caused *sickness-related absences* to decrease by 1.2 days and *paid* and *unpaid leave* to increase by 2.3 and 2.1 days respectively, compared to the initial situation.

that canceling sickness benefits would encourage workers to come to work even when they are sick, which would increase the spread of disease at the workplace. Our findings do not support this scenario. However, it is not exactly clear why the moral hazard behavior is not present; it may be that being sick at work is so uncomfortable that workers prefer to take un/paid leave instead; they are responsible and want to prevent disease spread when they feel unwell. There could also be other reasons behind the substitution pattern.

Using the same identification strategy, based on the varying need for sickness insurance, we examine the relationship between hours of absence and the incidence of other "non-respiratory" infectious diseases, including intestinal infectious diseases, bacterial diseases, etc. For many diseases we find similar substitution effects as in our baseline results; however, the results are economically negligible. Third, we focus on heterogenous effects. Specifically, we estimate the effects separately for mothers and fathers, shift-work occupations, occupations exposed to disease or infections, occupations with high interactions with co-workers, and occupations with high social interactions. In general, the substitution patterns associated with the legislative changes are in the same direction as in our baseline results, though the sizes of the effects differ. Our results suggest that mothers had more sickness-related absences during influenza outbreaks among children prior to 2008, and their response to the reforms was stronger than those of fathers and employees without children. Fathers, however, partially substituted this gap by increasing *paid* and *unpaid leave*. Shift-workers took more *sickness-related absences* and decreased the use of *paid* and *unpaid leave* less when exposed to influenza outbreaks. Furthermore, the effects of the reforms on *paid* and *unpaid leave* were not so strong among shift-workers. We find almost identical results when we classify occupations based on how easily infections spread in specific occupations, how much employees interact with others, and how much social interaction is needed to carry out necessary tasks. We find that occupations in the highest quartile, i.e., those who are the most exposed to infection, used more *sickness-related absences* and more *paid* and *unpaid leave* than the rest of the sample when exposed to influenza outbreaks prior to 2008. We also find that employees highly exposed to the infection decreased *sickness-related absences* more and increased *paid* and *unpaid leave* less as a reaction to the enacted legislative changes. Finally, we ask whether the effects differ by the size of organizational units. We see higher numbers of *sickness-related absences* during influenza outbreaks in organizational units with more employees, which suggests that influenza does indeed spread at the workplace. However, we did not find evidence that the policy reforms reduced or contributed to this spread.

Our paper contributes to several strands of the literature. First, there is only scarce literature on the effects of sick-pay reforms on workplace absences. One explanation could be that it is difficult to find a comparison group of employees who were not affected, which would make it possible to estimate causal effects.[10] Some scholars apply the 'before and after' identification strategy; however, such estimates also capture effects that are not necessarily attributable to the policy reforms. Exploiting the quasi-natural experiment setting, the difference-in-differences identification strategy is used most often (e.g., De Paola et al. (2014); Pettersson-Lidbom and Thoursie (2013); Ziebarth and Karlsson (2010)). Ziebarth and Karlsson (2010) studied the situation in Germany in 1996, when local authorities decreased the level of sickness benefits. This change fully impacted only private sector employees, allowing for the application of the standard difference-in-differences estimation strategy. Their results suggest that the legislative change increased the share of employees with zero sick leave days. A similar methodological approach was applied in Pettersson-Lidbom and Thoursie (2013), who study a legislative change in Sweden in 1987. This reform had two effects: the waiting period of one day was abolished and an income replacement rate for short-term illnesses increased. Pettersson-Lidbom and Thoursie (2013) found that the policy reform increased the share of workers who took additional sick-leave. De Paola et al. (2014) used a modified 'before and after' comparison, when they compared differences in days absent in the two periods before and two after the legislative change. They found that a partial wage reduction during the first ten days of sickness decreased the probability of absence. Our study complements preceding work by employing an alternative estimation strategy based on the intensity of exposure to sickness. Instead of focusing on comparing average absences of

[10] Sick pay programs are mainly in effect in small developed countries where the reforms are nation wide.

treated and control groups, or on the periods before and after the reform, we examine the local exposure to sickness. This allows us to compare the absences of employees who were / not exposed to influenza outbreaks. The advantage of this method is that it can be used even if the policy reforms are nation wide.
Second, we supply evidence on the relationship between health status and workplace absences. It is surprising that there is no research that would link sickness rates to absence from the workplace. Studies that consider the health status of employees focus solely on biological gender differences (e.g., Ichino and Moretti (2009); Paringer (1983)).
Third, our paper extends the evidence on changes in sick-pay programs in Central and Eastern Europe. There is limited literature from the Central and Eastern European region. Csillag (2017) studies the effects of a policy change in Hungary in 2011 that caused a 50% decrease in sickness benefits for the top 5% high-earning workers and a 25% decrease for a further 17% of workers. However, the benefits for low-earning workers remained the same. Applying the difference-in-differences estimation strategy, Csillag (2017) finds that the legislative change caused a small reduction in the incidence of absence form the workplace and a significant decrease in the number of days absent among low-earning employees. Pertold (2019) studies the effects of policy reforms in the Czech sick-pay system during 2008-2009, i.e., the same changes that are addressed in this paper. Working with the Czech the SES data, Pertold (2019) uses the 'before and after' comparison to find that these legislative changes significantly reduced the number of total days absent. Moreover, Pertold (2019) claims that the effects of the policy changes are extremely heterogenous across industries and occupations, and most affect employees working in manufacturing, hotels, and restaurants. Furthermore, Pertold (2019) shows that employees with more routine tasks and lower job-flexibility are much more likely to reduce their hours of absence.

The paper is structured as follows. Section 2.2 describes the 2008/09 changes in the sick-pay program. Section 2.3 describes the main data sources used in our analysis. Section 2.4 outlines our empirical strategy. Section 2.5 presents our main results. Section 2.6 concludes. The majority of graphs and tables can be found in appendix 2.7.

2.2 Institutional Context

The system of sick-pay insurance in the Czech Republic covers all salaried employees. The system is obligatory for employees (employers pay the insurance, which is 2.3% of the base salary), and voluntary for the self-employed.[11] The employee contribution is calculated as a share of gross wage with a floor that changes over time. Employees are obliged to inform employers about any obstacles to work, including sickness, immediately after they occur. In case of sickness absences, employees must deliver to their employer a sick note issued by a physician; this is aimed to restrict unnecessary absences from work. The physician decides whether the employee qualifies for a sick note. The Czech Social Security Administration can randomly check whether employees on sick-leave stay at home.[12].

There were two legislative changes during 2008-2009 that affected the sick pay program in the Czech Republic[13], the nature of which offer an interesting setting to study. The first legislation was enacted at the beginning of 2008, lasted only for approximately six months, and was then abolished by the Constitutional Court. At the beginning of 2009, a slightly modified version of the previous Act, which satisfied the objections of the Constitutional Court, was implemented. Table 2.1 summarizes the different stages of legislative changes in the Czech sick pay-program. To reduce the complexity of the sick-pay insurance system in the Czech Republic, we present only information about private sector workers who do not have a signed collective agreement. Extensive information regarding specific cases can be found on the web page of the Ministry of Labour and Social Affairs (2017).

Table 2.1: Timing of legislative changes enacted

comes into force	January 1, 2004	January 1, 2008	June 30, 2008	January 1, 2009
Stage number	stage 1	stage 2	stage 3	stage 4

Stage 1. Employees received sick-pay benefits of 25% of their wage (computed

[11] More information on rules and tariffs regarding the sick pay insurance can be found in Act no. 589/1992 Sb..

[12] These controls can be carried out also by employers if they pay the wage replacement, i.e. during the first 14 days of sickness in years after 2009.

[13] Currently, the Czech sick-pay system does not include a quarantine period (abolished in July 2019). During an employee's sick-leave, employers pay a contribution that amounts to 2.1% of the base salary. For more information see Act no. 32/2019 Sb..

based on the wage records from the past 12 calendar months) during the first three days of sickness i.e., during a so 'called waiting period'. After the first three days of sickness, employees were entitled to standard sick pay provided by the state. Employees who were sick usually received around 50-70% of their base wage, mainly based on the reason for the absence. The maximum duration for collecting sick-pay benefits was one calendar year. Employees are motivated to go to work as their salary is higher than what they receive during sickness. However, if they decide to stay home, they collect more money on sick-leave than they would on unpaid leave (assuming that they are able to obtain a sick note). Employers prefer to have their employees at work, as fewer workers limits production; however, absences are not associated with extra pay out of their pockets.

Stage 2. Employees did not receive any benefits in the waiting period, i.e., for the first three days of absence. They were also obliged to pay sick-pay insurance during the waiting period. After the 3rd day of their absence, employees were entitled to the same remuneration as in stage 1. Since employees do not receive any benefits during the first three day, they are motivated to limit their short-term absences. Depending on their preferences, they can prolong their sickness absence (the average cost of staying home is decreasing with days of absence) or they do not adjust the duration of the absence (the first three days are considered to be a sunk cost). The same incentives as in the preceding stage hold for employers.

Stage 3. The constitutional court decided to abolish the changes that became effective in stage 2. Stage 1 conditions were reinstated.

Stage 4. Employees were not entitled to any sick pay benefits in the waiting period; but they did not have to pay sick-pay insurance in the first three days of their absence. However, the employer became responsible for providing sick-pay benefits in the first 14 days of sickness. From the 15th day, the government provided sick-pay benefits from the sick pay insurance program. The maximum duration for collecting sick-pay benefits was 380 calendar days, counted from the day a sick note was issued. Employees have similar incentives to go to work as in the second stage - the only difference is that they do not have to pay sickness insurance. Employers are motivated to keep their employees at work because production is restricted if the workforce is diminished and they have to pay out sickness benefits during the first 14 days of an employee's sickness. It is not clear whether it is profitable for employers to force employees work when they are sick in terms of the cost of extra pay (cost of extra pay vs. economic costs of spreading diseases).

A detailed overview of the current situation across Europe can be found in the EU´s Mutual Information System on Social Protection (MISSOC) (European Commission's DG for Employment Social Affairs & Inclusion, 2017). The data show that the income replacement rate among European countries varies, but is not lower than 50 percent. Similarly, in the majority of cases, there is a quarantine period, sometimes also called a waiting period, during which an employee does not receive sickness benefits. The most common period in which sickness benefits can be collected is 52 weeks, but this varies substantially between countries. Governments pay out significant amounts of money for sickness benefits; therefore, it is not only in the best interest of employers to have an appropriate sick-pay setting, but all other interested parties, i.e, employees and the public. Our comparison shows that the Czech sick-pay system is similar to other European countries, and thus our results are also relevant to their sickness benefit programs. However, the generalization of our findings to countries with different sickness benefit programs is limited (e.g. developing countries or the US).

2.3 Data

This section introduces the three main data-sets used in our analysis: the ARI (Acute Respiratory Infections) data on the incidence of influenza and similar respiratory diseases, the EPIDAT (currently ISIN - "Information System on Infectious Diseases") data-set that collects the incidence of infectious diseases except respiratory diseases and HIV, and the ISPV ("Average Earnings Information System") that is the Czech Structure of Earnings Survey (SES). We convert all data to a county-quarter level. We create three age groups (children: 0-14 years old; adults: 15-59; elderly: 60+) that we use consistently through our analysis.

2.3.1 ARI

We use information on all reported incidences of acute respiratory diseases in the Czech Republic during 2005-2012, which amounts to approximately 1/2 of the total number of sickness spells per year.[14,15] The data-set contains the counts of weekly incidence of respiratory diseases by age group, gender, and the county where a sickness spell was reported. The type of data does not allow us to rule out the possibility that we might observe the same person several times in different weeks during the same quarter.

Graph 2.2 shows the evolution of incidences of influenza in Czech counties. Influenza is highly seasonal and children and teenagers are most affected. The incidence of respiratory diseases was approximately constant during 2005-2012 and there were no major drops in sickness around the time when the legislative changes affecting the sick-pay program were implemented. Table 2.4 compares the average incidence of acute respiratory diseases across counties in periods before and after the legislative changes happened. The incidence is higher for all age groups in the periods before the legislative changes, which suggests that the decrease is not caused by lower reporting rates among people who are economically active, but is more likely a general trend.

To confirm this, we regress the incidence rates on a dummy indicating periods after the policy changes, quarters, and a polynomial time trend. The results in Table 2.5 show that the indicator for periods after the legislative changes does not explain the decrease in respiratory infections when controlling for the time trend and seasonality. This is true for all age groups. Therefore, we assume that the counts of reported acute respiratory infections are not endogenous to the legislative changes under consideration. We use two measures of local sickness levels in our analysis. Apart from the normalized incidence of acute respiratory diseases (expressed as the number of influenza sickness spells per 100,000 people), we count the number of weeks with a flu epidemic status. The threshold for

[14]More information on the ARI database can be found on the website of the National Institute of Public Health `http://www.szu.cz/publikace/data/popis-systemu-ari?lang=1`.

[15]Sickness incidence is reported by physicians, hospitals, and other health centers. Every sickness with a clinical diagnosis of acute infection of the nasopharynx, acute infection of larynx and trachea, and flu must be reported to the ARI system.

influenza epidemics is 1,800 normalized incidences.

2.3.2 EPIDAT

The EPIDAT data contain all reported cases of infectious diseases except acute respiratory diseases and HIV.[16] The database contains a detailed classification of a reported disease, the county of report, and information about a patient's gender and age group. We use quarterly data during 2005-2012. The coverage is similar to the ARI database, i.e., only reported sickness spells are included, which is supposed to be around one half of the total incidence in the population. Graph 2.3 (in the Appendix) shows the incidence evolution of selected groups of infectious diseases from the EPIDAT database.[17] Similarly to the respiratory infections, the data show clear seasonal patterns and the incidence is the highest for the youngest patients. Table 2.6 shows that the incidence of infectious diseases in the period after 2008 was higher for children and young people, lower for the elderly, and approximately the same for adults. The diagnosis groups with the highest incidence for adults, who are most likely to be economically active, are *Intestinal infectious disease*, *Viral disease affecting skin*, and *Other virus diseases*.

2.3.3 SES

The Czech Structure of Earnings Survey (SES) allows us to observe the sickness-absence patterns of Czech employees. We use information on approximately 1.5 million private-sector employees from the Czech SES for each quarter during 2005-2012[18]. The data include repeated cross-section observations of all employed workers within surveyed firms. Firm-level data provide information about a firm's location (NUTS4 specification), 6-digit industry code, and the presence

[16]A complete list of diseases and their classification can be found, e.g., at `https://icd.who.int/browse10/2010/en#/I`. More information about the database itself can be found here: `http://www.szu.cz/publikace/data/infekce-v-cr?lang=1`.

[17]In line with the official classification, we aggregate the infectious diseases into following groups: *"Other"*, *"Intestinal infectious disease"*, *"Other bacterial diseases"*, *"Sexually transmitted diseases"*, *"Other spirochetes bacterias"*, *"Viruses affecting nervous system"*, *"Viral disease affecting skin"*, *"Viral hepatitis"*, *"Other viral diseases"*, *"Mykosis"*, *"Helminthiasis"*, *"Louses and similar"*. We omitted the groups of infectious diseases that included only a small number of observations.

[18]More information about the Czech SES can be found on the web page `https://ispv.cz/en/homepage.aspx`.

of a collective agreement. Data on employees contain: gender, age, place of work – county NUTS4 levels, hours worked, salary, occupation, education, tenure, and importantly, total hours absent, sickness-related absences, paid and unpaid leave.[19] Graph 2.4 shows a small decrease in *total hours absent* that was caused by a significant decrease in *sickness-related absences* and a small increase in *paid* and *unpaid leave* on average. The initial number of observations vary around 1.25 million per quarter during the period studied. After dropping observations with missing values and keeping only full time workers we are left with approximately 95% of the original data.

2.3.4 Other data

We use several other data-sets in our analysis. First, we use information on the number of employees by age groups, gender, and industry who work under the shift-work regime. The data come from the "Work organization and working time arrangements" survey compiled by Eurostat in 2004.[20] We implicitly assume that the structure of employees remained the same during subsequent years. Second, we use information on occupation-specific characteristics from the O*NET database.[21] Specifically, we use measures of how often (to what extent) specific occupations are exposed to other coworkers, disease spread, social interactions, etc., to determine which groups of workers are more likely to be exposed to sickness. Third, we use data from the Czech Statistical Office on the age profiles of parents, to assess whether an employee is a mother or father.

[19]We use four types of hours absent collected in the Czech SES. The variable *total hours absent* reports the total number of hours absent in a quarter. *Paid leave* gives information on the total amount of hours absent for which an employee received a wage, e.g., vacations, state holidays, etc. *Unpaid leave* states the number of hours an employee officially took vacation but was not paid for this, i.e., it is vacation that is taken on the top of the settled amount. *sickness-related absences* include all hours absent when an employee reported himself ill; however, the variable does not distinguish the sickness of employees from absences that employees took to take care of sick relatives.

[20]Description of the data can be found at https://ec.europa.eu/eurostat/cache/metadata/en/lfso_04_esms.htm.

[21]https://www.onetonline.org/

2.4 Empirical Design

We begin by establishing the relationship between hours absent and sickness rates. We approximate local sickness levels by the incidence of acute respiratory diseases and reported cases of infectious diseases.[22] We find positive correlations, conditional on seasonal and regional effects, between sickness-related absences and the incidence of acute respiratory infections in periods before and after the legislative changes were enacted (Table 2.7).[23] We argue that it is important to control for sickness rates when estimating the 'before-after' type of models, because it may affect the size of the estimated coefficients, e.g., the coefficient that is often reported as the effect of the policy reforms. Table 2.10 reports results when (not) controlling for the sickness rates in the estimated model.

We first estimate the effects of policy reforms using the 'before-after' estimation strategy (the γ coefficient in Equation 2.1). This estimate is based on the comparison of two conditional means with little identification variation involved. Possibly, the estimate also captures the effects of events timed simultaneously with the policy reforms but otherwise unrelated. The estimate captures the overall effect of the policy reforms including absence adjustments for those who were shirking, suffered a disease, or were absent for another reason. Second, we use the prevalence of selected infectious diseases to provide variation in the need for sickness insurance. We use this variation to apply the *intensity treatment* identification strategy.[24] The intensity treatment here is the need for sickness insurance, which is high in counties with high sickness rates and is affected by the policy reforms; i.e., our estimation strategy relies on a quasi-random assignment to different levels of treatment. We focus on a narrow group of employees exposed to a specific disease, e.g., those who were exposed to influenza, and compare absences of employees differently exposed to influenza (i.e. with different needs

[22] Data-sets and the construction of variables used in our analysis are described in Section 2.3.

[23] We use a normalized incidence of acute respiratory infections and the number of weeks with influenza epidemic status in a quarter (1,800 cases per 100,000 employees and more) to measure local sickness levels. The data allow us to measure the sickness levels among children, adults, and the elderly. We prefer to use the counts of weeks with influenza epidemic status on the right hand side of the regression equations to avoid circular measurement.

[24] By construction, the *intensity treatment* is similar to the *difference-in-differences* estimator (for earlier applications see, e.g., Card, 1992b; Machin et al., 2003).

for sickness insurance because of influenza) in the periods before and after the policy changes. In this framework, the policy changes did not affect the need for sickness insurance among those who were shirking or suffered from diseases other than influenza. The advantage of this strategy is that our estimates are based on much more variation than a simple 'before-after' comparison.[25] Our local average treatment estimates are important because they show how the policy reforms affected the spread of the disease in the workplace. Canceling sickness benefits during the first three days of sickness could incentivize employees to go to work sick, where they may infect other coworkers. The underlying identifying assumptions are that influenza outbreaks were not a consequence of sickness transmission in the workplace (it can make the situation worse but does not initiate the outbreaks) and that shirking is not affected by the epidemic situation, i.e, employees do not shirk more during an outbreak. Given that we observe complete absence records of each employee decomposed by the type of absences (*sickness-related absences*, *paid* and *unpaid leave*), we bring qualitative evidence on the mechanism behind policy reforms and influenza outbreaks. This allows us to explain why the changes in hours absent happened in such a manner, and why these changes differ across selected socio-demographic groups of employees. We estimate the following equation.

$$\text{absence hours}_{c,i,t} = \alpha + \beta \text{ sickness}_{c,t} + \gamma \text{ after} + \delta \text{ sickness}_{c,t} * \text{after} + \eta \text{ X}_{c,i,t} + \epsilon_{c,i,t} \tag{2.1}$$

where the term *absence hours* denotes various types of hours absent, *sickness* represents measures of sickness incidence (normalized numbers and no. of weeks with epidemic status in counties; see Section 2.3 for more details). A possible concern would be that adult sickness rates (or total incidence in population) are endogenous to absences.[26] To address this issue, we use the measure of local influenza outbreaks among children, implicitly assuming that they can infect adults

[25]We use information about sickness incidence in 76 Czech counties. For more details, see section 2.3.

[26]There might be another problem with using the normalized incidence measure among adults. If sick employees go to work, spread the flu among coworkers, and eventually report themselves sick, there would be one-to-one matching between local sickness-related absences and officially reported sickness incidences.

but not vice versa. We prefer to use the counts of weeks with epidemic status, i.e., we identify the effects using the variation that is based on whether or not the incidence of infections was significant during a specific week in a county. The corresponding threshold is the official definition of an infection outbreak - epidemic; for influenza it corresponds to 1,800 reported cases per 100,000 inhabitants. The announcement of epidemic status is not solely a formal declaration, but it allows government to use specific measures to fight the fast spreading disease (see Government decree no. 258/2000 Sb.)[27]. Thus, our identification strategy assumes that the threshold for epidemic status correctly assesses the seriousness of the then current epidemic situation.

The variable *after* is an indicator for periods after the legislative changes, and hence the γ coefficient captures an effect of periods after the legislative changes were introduced (irrespective of our sickness exposure measures and other controls), X stands for other control variables, and ϵ are cluster-robust standard errors. Subscript c stands for county, i individuals, t time. The coefficient β shows how *sickness* affects hours absent in periods before the policy changes. We expect this coefficient to be positive for sickness absence hours as employees most likely stay home when they are sick (the adults' sickness measure) or they have to take care of their sick children (the sickness incidence among children; which also captures a possibility that children infect their parents). Unfortunately, the nature of the data does not allow us to distinguish whether an employee is sick or taking care of his/her sick children. It is less clear what sign to expect in the case of paid and unpaid hours absent; our expectation is that both coefficients have negative signs as employees take less vacation when sick/taking care of their sick children. The δ coefficient represents the adjustment/change in absences for employees in locations with incidence of sickness caused by the policy changes. Given that the β and δ coefficients capture the effect at the county-quarter level, a positive sign for the δ coefficient in the equation with the sickness-related absences as dependent variables would mean that the policy changes contributed to the spread of the disease among coworkers. In that case, the policy changes would cause negative externalities in the form of moral hazard, which may be

[27] For example, the Government can put restrictions on production, transport, distribution of food etc.

financially more damaging than spending more on sickness benefits.

2.5 Results

We begin our analysis by establishing the relationship between our sickness measures and hours absent. Table 2.7 shows that there is a statistically significant relationship between normalized sickness-related absences and exposure to sickness, i.e., employees take more sick-leave during the flu season. Table 2.8 shows strong positive correlations between the normalized incidence of respiratory diseases and our exposure measure, defined as the number of weeks with influenza epidemic status per quarter. We further explore the relationship between sickness and absence rates on the level of occupations (for details see Table 2.9 in the Appendix). It could be that employees in some occupations take more sickness-related absences after exposure to influenza, e.g., they are more likely to get infected. In particular, we find persistent patterns of positive correlations among *Technicians and associate professionals* and *Clerical support workers* in the periods before and after 2008.

Baseline results

We continue by exploring how hours absent vary by differing exposure to sickness. We estimate Equation 2.1 using both measures of local sickness levels, i.e., the normalized incidence of acute respiratory diseases and the number of weeks with influenza epidemic status in a county per quarter. Our main results are presented in Table 2.2. The first three columns show results for when we employ the influenza outbreak among adults, the last three show results when the children' outbreak is employed. Both sets of estimated coefficients are similar in their signs.

For simplicity, we discuss our results below in terms of days absent for employees who were exposed to sickness, i.e., we interpret the effects for those who were sick.[28] We first focus on the results where we use the outbreak among adults.

[28]We recalculate the average effects (the β and δ coefficients) to the proportion of employees

Table 2.2: Hours absent - Respiratory infections outbreak

	Adults' outbreak			Children's outbreak		
	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave
After	-9.003***	7.416***	2.243***	-7.726***	3.948***	-0.684
	(0.280)	(1.007)	(0.482)	(0.295)	(1.020)	(0.661)
Sickness	0.719***	-0.0574	-0.292**	0.148***	-0.259***	-0.228***
	(0.0629)	(0.142)	(0.124)	(0.0106)	(0.0405)	(0.0301)
After*Sickness	-0.961***	0.232	0.337	-0.176***	0.331***	0.302***
	(0.270)	(0.620)	(0.238)	(0.0112)	(0.0316)	(0.0256)
Observations	15,327,196	15,326,330	15,318,628	15,327,196	15,326,330	15,318,628
Adjusted R2	0.031	0.299	0.355	0.031	0.300	0.356

Notes: The table shows two sets of regression results (Equation 2.1). We use counts of weeks with epidemic in a quarter for adults in the first three columns, and for children in the last three columns. Dependent variables are: sickness-related absences (sickness absences), paid leave, and unpaid leave. Controls were: age, tenure, gender, collective agreement, quarter, year, county, industry, occupation, firm size cat., educ. cat., nationality and constant term. Cluster-robust errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

sickness-related absences are positively related to sickness incidence (coefficient β from Equation 2.1). Back-of-the-envelope calculations reveal that one extra week of epidemic status causes an increase in sickness-related absences of 5 working days. This is in line with the fact that flu symptoms last from 5 to 7 days.[29] Therefore, it appears that employees do not take advantage of influenza outbreaks to shirk; however, we cannot rule this out completely as some employees may go to work when sick and a similar number stay home during flu season when they are healthy. The opposite relationship holds for *unpaid leave*, where one week of influenza epidemics causes a 2-day decrease in hours absent, which means that employees took less unpaid leave when there was a higher incidence of influenza. Employees might not need to use unpaid leave when they are sick at home. The coefficients δ show how the relationship between hours absent and the influenza outbreak changed after the legislative changes were introduced. We find that an extra week of influenza epidemic decreased *sickness-related absences* (by 6.7

who were actually sick. The definition of influenza outbreak is 1,800 infected per 100,000 inhabitants, i.e., we multiply the estimated coefficients by 100,000/1,800 to obtain hours absent, and further divide by 8, assuming that an average working day has 8 working hours. We use the same calculation procedure consistently throughout.

[29]For example, see `https://www.health.harvard.edu/staying-healthy/how-long-does-the-flu-last`.

days), resulting in the overall effect of the influenza outbreak being negative (-1.7 days) in the periods after 2008. However, the legislative changes also caused increases in *paid* (1.6 days) and *unpaid leave* (2.3 days) compared to the situation before 2008, making the overall effect of an influenza outbreak on *paid* and *unpaid leave* positive in the periods after 2008 (1.5 days in total).[30] Our results indicate that the legislative changes led employees to almost perfectly substitute *sickness-related absences* by *paid* and *unpaid leave* when exposed to influenza outbreaks.

We observe a similar pattern when we employ the influenza outbreak among children measure (the last three columns of Table 2.2). The estimated β coefficients suggest that one extra week of influenza outbreak causes an increase in *sickness-related absences* of 1 day, which is a significantly lower effect compared to the estimate when the influenza outbreak among adults is employed.[31] It could be that employees are either sick because their children infected them (not every child infects its parents) or because they have to take care of their sick children (and they can share the responsibility to stay home within the family). However, they do not need to take as many *sickness-related absences* as for the adults' outbreak. Our findings further suggest that one extra week of influenza outbreak decreases *paid absence hours* by 1.8 days (employees spend less time on vacation when they or their children are sick) and *unpaid leave* by 1.6 days (they do not need to take extra unpaid leave when they are home sick or babysitting their children). Similarly to the results for the adults' outbreak, the δ coefficients indicate that the legislative changes caused a decrease in *sickness-related absences* by 1.2 days and an increase in *paid* (2.3 days) and *unpaid leave* (2.1 days) compared to the situation before 2008. The composite effects in periods after 2008 are that one extra week of influenza outbreak among children causes a small decrease in *sickness-related absences* (0.2 days) and a small increase in *paid* and *unpaid leave* (1 day). Overall, our results show that the legislative changes induced significant adjustments in employees' absences. However, though the substitution effects among types of hours absent are substantial, the overall effect on total hours absent is small.

[30]However, the δ coefficients for paid and unpaid leave are imprecisely estimated.

[31]The coefficients could be smaller because the sickness-related absences and adults' sickness rates are endogenous, there is a measurement error that produces downward bias, or a combination of both.

Table 2.11 shows similar results to Table 2.2, where the normalized measure of influenza incidence is used on the right hand side of the regression equations. Both sets of results are similar in terms of their sign and size. We observe that *sickness-related absences* are positively related to sickness levels (one st.dev. increase in sickness exposure causes an increase of 1.03 hours). The opposite relationship holds for *paid leave* (one st.dev. increase in sickness exposure causes a decrease of 1.48 hours) and *unpaid leave* (one st.dev. increase in sickness exposure causes a decrease of 0.76 hours). A one standard deviation increase in sickness incidence after the legislative changes were introduced (i.e., the coefficient *delta*) would cause *sickness-related absences* to decrease by 1.01 hours, *paid leave* to increase by 1.82 hours, and *unpaid leave* by 1.53 hours. Similarly to our previous results, this pattern indicates that employees substituted *sickness-related absences* with *paid* and *unpaid leave*.

We complement our main results by studying the effects of other “non-respiratory" infectious diseases on hours absent. Specifically, we use the normalized incidence of selected infectious diseases except acute respiratory diseases and HIV among children from the EPIDAT data-set and estimate Equation 2.1.[32] The results are in Table 2.12. For many diagnosis groups, we find similar substitution effects to those our main results. However, all these effects are small and have almost no economic impact (the effects of the highest magnitude vary around 1 hour of absence). Therefore, in the following text we continue with our analysis of exposure to influenza.

Heterogenous effects

We extend our baseline analysis by identifying groups of employees who may have different reasons for being absent, for example because they have to stay with sick children at home, are frequently in contact with other people that may infect them, or their work conditions make their absence more costly. Specifically, we estimate the effects for likely mothers and fathers separately, identify the occupations with high shares of employees working in the shift-work regime, and use the O*NET database to select employees with i) a high probability of a disease

[32] We use only selected groups of diseases with sufficient numbers of local incidence and variation. The data from the EPIDAT data-set are described in Section 2.3.

spread, ii) frequent contact with other coworkers, and iii) high social interactions. In each case we split our sample and compare whether the estimated coefficients from Equation 2.1 differ for observations with high/low values. For each chosen characteristic, we divide the observations into quartiles and present the results (Tables 2.14, 2.15, 2.16, 2.17) for all observations in the first three columns, those in the top quartile (columns 4-6), and these up to the 75th percentile (columns 7-9) separately.

First, we assess whether an employee is likely to be a mother or father. Since our data do not include information about the number of employees' children, we approximate that every female between 23-44 years old is likely to be a mother and every male between 26-47 is likely to be a father. Our assessment is based on the age profile of parents in the Czech Republic.[33] The estimates from Equation 2.1, with the adults' sickness exposure in Table 2.3 for mothers and Table 2.13 for fathers, show that coefficients β and δ are not statistically different from the estimated population-wide effects, nor do they differ across the subsamples of mothers and fathers. However, there are statistically significant differences in the case of influenza outbreaks among children. Mothers had more sickness-related absences during influenza outbreaks before 2008 (the β coefficient). It is likely that they spent more time at home with their sick children than their partners. The decrease in sickness-related absences caused by legislative changes (the δ coefficient) is statistically larger for mothers, who adjusted their behavior more. Our results suggest that fathers partially compensated for this by increasing paid leave (the coefficient δ in Table 2.13).[34]

Second, we use two-digit-occupation-specific information on the share of employ-

[33]The fertility distribution among Czech women shows that the most common age to give birth is between 23-33 years (the mean age is 29 years). We use 2005 data assuming that women stay home with a child for 3 years, so the data correspond to the timing of the policy changes that happened during 2008-2009. Therefore, given that children's needs are most time-consuming up to the age of 11 (Milkie et al., 2015), we approximate that a woman is a mother of a young child who needs to be taken care of when she is 23-44. Similarly, we define fathers as on average, 3 years older, i.e. we say that a man is father if he is 26-47 years old. Source `https://www.czso.cz/csu/xb/vek-rodicu-v-jihomoravskem-kraji-v-roce-2017`.

[34]Our findings suggest that prior to 2008, it was mothers who stayed home with their sick children. This could implicitly disadvantage them (it is likely that they will have more absences compared to their male coworkers) and contribute to a larger gender-wage gap. The situation improved when fathers started to take care of sick children more after 2008, which could steer perceived differences and make the situation more equal, but not completely.

Table 2.3: Subsample of mothers - Influenza outbreak

	Adults' outbreak			Children'n outbreak		
	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave
After	-11.05***	6.775***	1.475***	-9.364***	3.794***	-1.188***
	(0.439)	(0.763)	(0.349)	(0.443)	(0.760)	(0.410)
Sickness	0.873***	-0.259*	-0.508***	0.205***	-0.274***	-0.237***
	(0.0927)	(0.136)	(0.118)	(0.0154)	(0.0340)	(0.0255)
After*Sickness	-1.080***	0.0731	0.679***	-0.224***	0.291***	0.293***
	(0.207)	(0.595)	(0.225)	(0.0151)	(0.0290)	(0.0184)
Observations	3,292,272	3,292,206	3,290,553	3,292,272	3,292,206	3,290,553
Adjusted R2	0.033	0.347	0.409	0.033	0.348	0.411

Notes: The table shows two sets of regression results (Equation 2.1) for the subsample of mothers(females 23-44 y.o.). We use counts of weeks with epidemics in a quarter for adults in the first three columns, and for children in the last three columns. Dependent variables are sickness-related absences (sickness absences), paid leave, and unpaid leave. Controls included are age, tenure, gender, collective agreement, quarter, year, county, industry, occupation, firm size cat., educ. cat., nationality and constant term. Cluster-robust errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

ees who work under the shift-work regime (for details see Section 2.3). We assume that shift-work employees have fewer opportunities to shirk, since the production is heavily dependent on their presence at the workplace, and thus employers do not tolerate unjustified absences. Table 2.14 shows that there are small differences between occupations with “high" and “low" shares of shift-work employees but the substitution pattern is similar to our main results. We find that employees in the “high" occupations classification show a slightly larger β estimate in the *sickness-related absences* regression and smaller estimates (half the size) in absolute values for *paid* and *unpaid leave*.[35] This indicates that employees in occupations with high shares of shift-workers took more sick leave, but also took more paid and unpaid leave than the rest of the sample during flu outbreaks. The estimated δ coefficients are significantly lower for *paid* and *unpaid absence hours* among occupations with high shares of employees working shifts, which suggests that the substitution effect of the legislative changes was much lower for these occupations. This is in line with our expectation that shift-working employees have less flexibility.

[35]The regressions control for occupation fixed effects.

Third, we use the O*NET database to classify occupations based on how much they are exposed to diseases or infections, how intensive their contact with others is, and how socially oriented their occupation.[36] We expect a stronger reaction to influenza outbreaks among occupations with high scores.[37] Similarly to previous classifications, we divide observations into quartiles based on the above defined O*NET scores and estimate Equation 2.1 for all observations, for the top quartile, and those that belong to the first three quartiles. The results are presented in Tables 2.15, 2.16, and 2.17. We find that the estimated coefficients for subsamples based on all three classifications yield similar results. Similarly to our baseline results, there are positive effects of influenza outbreaks on *sickness-related absences* and negative effects on *paid* and *unpaid leave* in the periods before 2008. The legislative changes (δ estimates) caused the opposite effects, i.e., employees substituted *sickness-related absences* by *paid* and *unpaid leave*. However, the sizes of estimated coefficients differ for employees who belong to the highest quartiles, based on our classifications. On average, we observe a higher decrease in *sickness-related absences* and a smaller increase in *paid* and *unpaid absence hours* in the periods after 2009 (the γ coefficients). Those employees also took more *sickness-related absences* and more *paid* and *unpaid absence hours* compared to the rest of the sample when exposed to influenza outbreaks prior to 2008. We further observe that their *sickness-related absences* decrease more and *paid* and *unpaid leave* increase less as a reaction to the enacted legislative changes (the δ estimates from Equation 2.1). The results suggest that the substitution effect for the "high" group was not as large as for the rest of the employees.

Finally, we ask whether absence behavior differs by the size of job cells. These are firm-county-occupation specific organizational units with employees who have similar characteristics and, thus similar absence behavior (for determinants of

[36]The classification is carried out based on the following exact formulations. Exposed to Disease or Infections: *"How often does this job require exposure to disease/infections?"*, Contact With Others: *"How much does this job require the worker to be in contact with others (face-to-face, by telephone, or otherwise) in order to perform it?"*, Social Orientation: *"Job requires preferring to work with others rather than alone, and being personally connected with others on the job"*. Each occupation is assessed on a scale from 0 (minimum) to 100 (maximum).

[37]We expect that influenza spreads with higher intensity in occupations that are more exposed to disease or infections, (employees are more likely to be exposed to disease), with more intense contact with others (employees are more likely to meet someone who is infectious), and in more "pro-social" occupations (more frequent interaction with people is associated with a higher probability catching a disease from someone who is infectious).

absence behavior see, e.g., Barmby, 2002; Scoppa, 2010).[38] The larger the job cell is, the more likely is that sick employees infect more of their coworkers. We employ the regression equation similar to Equation 2.1 but we estimate it on the level of job cells. The dependent variables are the average absence hours in the job cell and the independent variables of interest are mutual interactions of a dummy indicating periods after the change, job-cell size (the natural logarithm of number of employees in a job cell), and our sickness measures. We use standard control variables and include the job-cell fixed effects to control for unobserved differences between organizational units.

The results from the first two rows in Table 2.18 are similar to our β and δ baseline estimates and follow the story that employees substitute sickness-related absences by paid and unpaid leave. The positive coefficient associated with the interaction term of the sickness measure and job-cell size shows that employees of larger job cells took more sickness-related absences when exposed to influenza outbreaks.[39] This suggests that employees spread the flu at work in both periods before and after the legislative changes to a similar extent, which is natural as they interact. However, we do not find evidence that the legislative changes had either a positive or negative effect on the spread of disease at the workplace (the triple interaction term in Table 2.18). If the reforms caused more intense disease spread, we would observe coefficients with positive signs.

2.6 Conclusion

This paper studies the effects of decreases in sickness benefits during 2008/9 in the Czech Republic on hours absent from work of private sector employees. We use local exposure to sickness -influenza outbreaks to provide variation in the need for sickness insurance, which allows us to apply an intensity treatment estimator on the county level. Compared to the 'before-after' approach, our local average treatment estimates are more conducive to causal interpretation. Assuming that adults' sickness rates and absences from work may be endogenous,

[38]We prefer to use job cells to firms because employees in the same occupations, tend to meet each others more frequently within a firm e.g., manual assembly workers vs. managers.

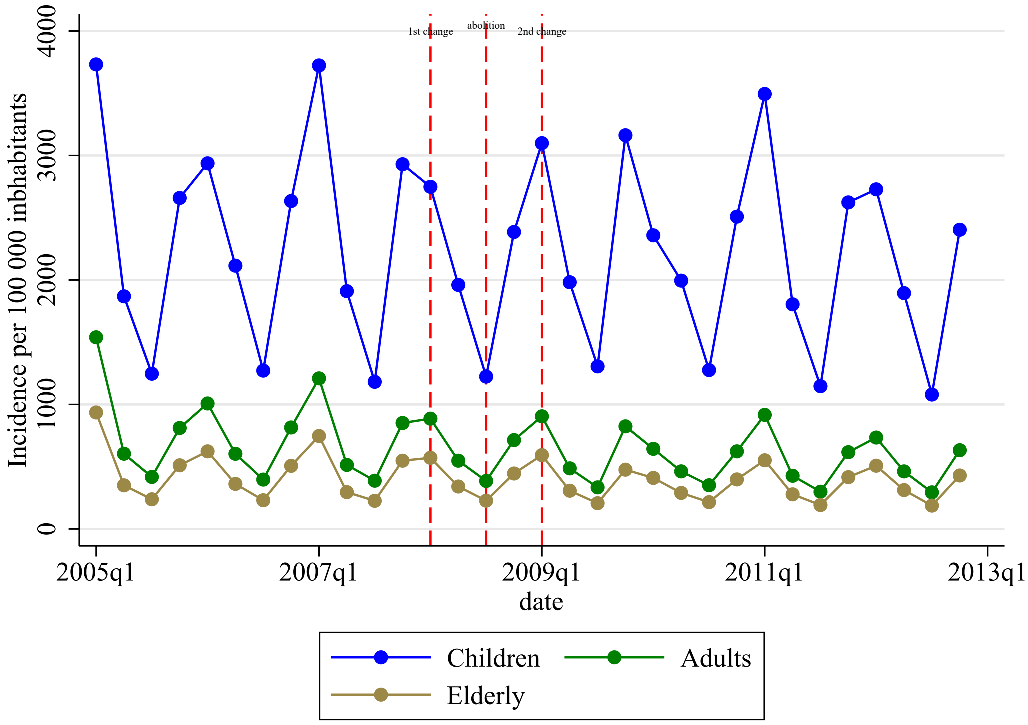
[39]The effect of the outbreak among adults on sickness-related absences is imprecisely estimated.

we use children's sickness rates to measure the exposure to sickness. Our results suggest that as a reaction to the decrease in sickness benefits, employees exposed to influenza outbreaks reduced their sickness-related absences but almost perfectly compensated this drop by taking paid and unpaid leave, leaving total absences at the same level. Hence it is unlikely that the reforms led to more employees working while sick. We do not find evidence that employees spread influenza among their coworkers more in periods when sickness benefits were reduced or that employees took advantage of influenza outbreaks to shirk; although we cannot rule this out completely. The size of the substitution effect differs by occupational and sociodemographic characteristics of employees. Mothers took more sickness-related absences before the policy changes (probably in order to take care for sick children), while this group took fewer sickness-related absences than other employees in periods after the reform. Fathers compensated for this reduction in sickness-related absences taken by mothers by increasing their paid and unpaid leave. We find a smaller substitution effect for employees who work in occupations that are more exposed to diseases or where social interaction is more frequent.

2.7 Appendix

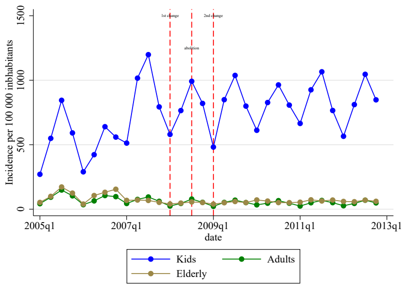
Graphs

Figure 2.2: Incidence of respiratory diseases by age groups (ARI database)

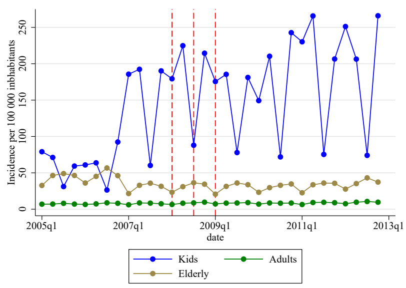


Note: Average incidence across 77 counties for age groups (0-14; 15-59; 60+) in the Czech Republic. The red bars indicate the timing of the legislative changes in sick-pay policy. The red vertical lines indicate the timing of the legislative changes.

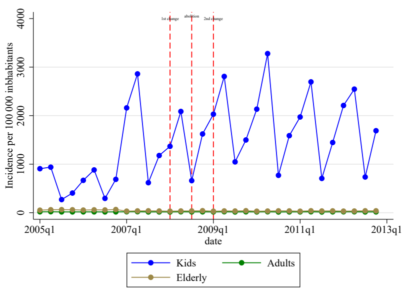
Figure 2.3: Incidence of selected groups of infectious diseases (EPIDAT database)



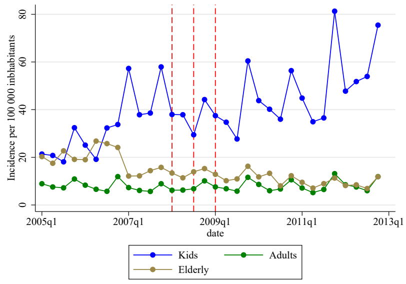
(a) Diagnosis group: Intestinal infectious diseases



(b) Diagnosis group: Other bacterial diseases



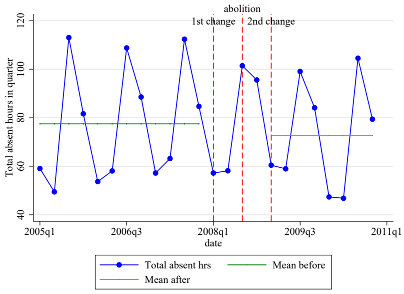
(c) Diagnosis group: Viral diseases affecting skin



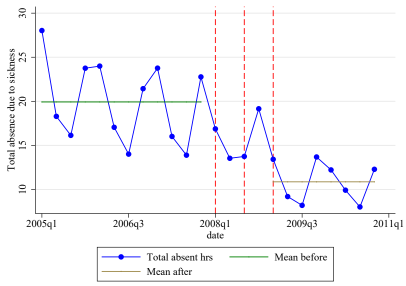
(d) Diagnosis group: Louses and similar

Note: The graphs show the incidence of selected infectious diseases from the EPIDAT database. The red vertical lines indicate the timing of the legislative changes.

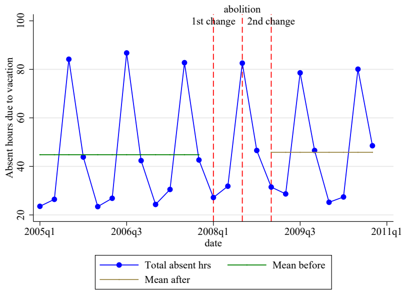
Figure 2.4: Hours absent before and after the reform



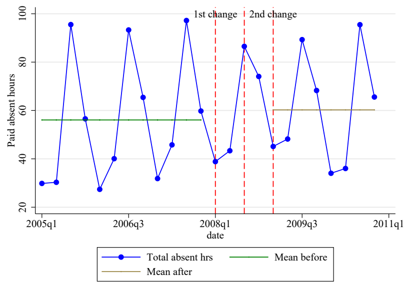
(a) Total hours absent



(b) Sickness-related absences



(c) Paid leave (hrs)



(d) Unpaid leave (hrs)

Note: The graphs show average quartal hours absent (by category) with the means for periods before and after the reforms. The red vertical lines indicate the timing of the legislative changes.

Tables - Descriptive

Table 2.4: Comparison of acute respiratory diseases by age groups, before and after the policy changes

	Children			Adults			Elderly			Total		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
After reform	-171.9***	320.8***	18.90	-199.9***	134.4***	11.27	-103.9***	72.92***	-5.477	-204.5***	163.6***	8.386
	(42.89)	(43.41)	(42.45)	(13.18)	(18.76)	(17.76)	(12.97)	(13.82)	(13.23)	(16.22)	(19.11)	(17.84)
Constant	2350.9***	2534.2***	3252.6***	763.4***	971.7***	1233.8***	464.7***	586.9***	756.1***	1102.6***	1311.7***	1664.1***
	(70.58)	(64.79)	(72.08)	(19.49)	(25.46)	(29.59)	(20.32)	(24.72)	(31.43)	(26.33)	(28.73)	(32.50)
Time trend	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Quartal	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492
Adjusted R^2	0.006	0.017	0.509	0.084	0.129	0.603	0.035	0.056	0.369	0.049	0.078	0.619

Notes: Average incidence of acute respiratory diseases per 100,000 inhabitants across 77 counties by age groups. Standard deviations in italics.

Table 2.5: Incidence of acute respiratory diseases

	Children			Adults			Elderly			Total		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
After reform	-171.9***	320.8***	18.90	-199.9***	134.4***	11.27	-103.9***	72.92***	-5.477	-204.5***	163.6***	8.386
	(42.89)	(43.41)	(42.45)	(13.18)	(18.76)	(17.76)	(12.97)	(13.82)	(13.23)	(16.22)	(19.11)	(17.84)
Constant	2350.9***	2534.2***	3252.6***	763.4***	971.7***	1233.8***	464.7***	586.9***	756.1***	1102.6***	1311.7***	1664.1***
	(70.58)	(64.79)	(72.08)	(19.49)	(25.46)	(29.59)	(20.32)	(24.72)	(31.43)	(26.33)	(28.73)	(32.50)
Time trend	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Quartal	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492
Adjusted R^2	0.006	0.017	0.509	0.084	0.129	0.603	0.035	0.056	0.369	0.049	0.078	0.619

Notes: We regress a county level incidence of acute respiratory infections on a dummy variable indicating periods after the policy change, conditional on polynomial time trends and indicators of quarters. We use clustered errors on the county level. Significance levels: *** 0.01, ** 0.05, * 0.1. Standard errors in parenthesis.

Table 2.6: Comparison of infectious diseases by age groups before and after the legislative changes

		before 2008				after 2009			
		kids	adults	elderly	total	kids	adults	elderly	total
	Diagnosis group								
0	Other	9.8	2.8	6.1	2.4	15.2	2.6	3.9	2.2
		6.3	*2.2*	*4.3*	*1.9*	*9.9*	*2.2*	*2.3*	*1.8*
1	Intestinal infectious disease	639.9	80.6	95.4	145.3	817.0	48.4	59.3	102.5
		398.3	*48.0*	*84.9*	*76.4*	*407.5*	*30.2*	*55.5*	*49.2*
2	Other bacterial diseases	93.9	7.5	40.0	20.3	181.1	8.6	32.2	25.1
		107.6	*5.3*	*34.4*	*12.8*	*160.0*	*6.5*	*26.6*	*15.4*
3	Sexually transmitted diseases	5.8	5.9	5.0	4.2	15.8	7.5	4.4	5.2
		2.8	*9.5*	*4.1*	*6.7*	*7.0*	*8.6*	*3.3*	*6.1*
4	Other spirochetes bacterias	21.7	10.1	22.8	11.0	38.8	10.5	16.5	11.7
		20.2	*13.8*	*31.3*	*14.9*	*35.9*	*12.9*	*19.8*	*14.4*
5	Viruses affecting nervous system	19.2	5.1	10.2	5.2	23.4	4.7	6.6	4.4
		23.3	*5.3*	*11.4*	*5.7*	*18.8*	*4.7*	*6.6*	*4.5*
6	Viral disease affecting skin	990.1	17.7	51.3	125.5	1822.8	17.2	35.6	143.3
		1148.7	*10.7*	*35.3*	*104.7*	*1610.9*	*10.7*	*23.5*	*114.0*
7	Viral hepatitis	19.2	5.7	6.8	4.6	62.1	5.9	4.5	5.1
		35.2	*7.2*	*5.3*	*6.0*	*106.9*	*6.8*	*3.1*	*7.1*
8	Other viral diseases	44.2	11.4	8.3	11.9	70.9	11.1	3.9	11.3
		65.8	*16.5*	*11.1*	*16.5*	*141.1*	*26.6*	*2.4*	*24.7*
9	Mykosis	12.4	6.1	20.8	5.7	23.9	6.0	11.3	5.5
		9.8	*7.0*	*20.8*	*7.3*	*19.6*	*6.3*	*9.3*	*6.4*
10	Helminthiasis	24.3	1.8	6.0	3.2	30.9	1.6	3.2	2.4
		21.8	*1.4*	*3.6*	*2.8*	*25.6*	*1.0*	*1.7*	*1.9*
11	Louses and similar	32.9	8.0	19.0	10.3	49.3	8.2	10.7	9.2
		39.0	*8.2*	*25.2*	*10.2*	*55.0*	*9.1*	*12.0*	*9.9*

Note: Average incidence of infectious diseases per 100,000 inhabitants across 77 counties by age groups. Standard deviations in italics.

Tables - Results

Table 2.7: Hours absent and respiratory infections correlations

	Before			After		
	Total	Children	Adults	Total	Children	Adults
Normalized sickness	0.00133**	0.00007	0.00397***	0.00101***	0.000101	0.00254***
Sickness outbreak	0.241***	0.102***	0.705***	-0.168***	-0.0556***	-0.392*
Observations	9,385,668	9,385,668	9,385,668	7,660,028	7,660,028	7,660,028

Notes: The table presents correlations between sickness-related absences and county-level exposure to influenza-like diseases in periods before and after the legislative changes were enacted. Correlations are net of seasonal and regional effects. The first line counts the sickness exposure expressed as a normalized incidence of influenza. The second line shows results where the sickness exposure is measured as number of weeks with epidemics status. Standard age groups are used. Significance levels: *** 0.01, ** 0.05, * 0.1.

Table 2.8: Correlation matrix of influenza measures

		Normalized incidence			
		Children	Adults	Elderly	Total
Influenza outbr.	Children	0.8625*	0.6886*	0.4504*	0.8083*
	Adults	0.3013*	0.4688*	0.2854*	0.4062*
	Elderly	0.0904	0.0716	0.5901*	0.1349*
	Total	0.6882*	0.7228*	0.5401*	0.7461*

Notes: The table presents a correlation matrix for normalized incidence of influenza and the influenza outbreak measure computed as the number of weeks with influenza epidemic status per quarter. The correlations are shown for three age groups as well as aggregates. An observational unit is quarter-year-county specific. Standard age groups are used. Significance level: * 0.01

Table 2.9: Hours absent and respiratory infections correlations (by occupations)

			Sickness exposure					
			Before			After		
Group	CZ ISCO	Occupation	Children	Adults	Total	Children	Adults	Total
Professionals								
	21	Science and engineering professionals	0	0.00206***	0.00119***	0.000190**	0.00169***	0.000900***
			292,996	*292,996*	*292,996*	*310,888*	*310,888*	*310,888*
	23	Teaching professionals	0	0.00101**	0	0.000262**	0.00106*	0.000893**
			131,712	*131,712*	*131,712*	*115,572*	*115,572*	*115,572*
	24	Business and administration professionals	0	0.00316***	0.00147***	0	0.00316***	0.000936**
			413,076	*413,076*	*413,076*	*380,332*	*380,332*	*380,332*
Technicians and associate professionals								
	31	Science and engineering associate professionals	0.000231**	0.00252***	0.00134***	0.000299**	0.00285***	0.00139***
			904,060	*904,060*	*904,060*	*737,004*	*737,004*	*737,004*
	32	Health associate professionals	0.00127***	0.00236**	0.00373***	0	0	0
			127,976	*127,976*	*127,976*	*204,236*	*204,236*	*204,236*
	34	Legal, social, cultural and related associate professionals	0.000557***	0.00257***	0.00194***	0.000220*	0.00254***	0.00120***
			727,981	*727,981*	*727,981*	*634,021*	*634,021*	*634,021*
Clerical support workers								
	41	General and keyboard clerks	0.000983*	0.00458***	0.00357**	0.000329*	0.00385***	0.00175***
			514,144	*514,144*	*514,144*	*445,024*	*445,024*	*445,024*
	42	Customer services clerks	0.00134***	0.00625***	0.00523***	0	0.00457***	0.00225***
			319,064	*319,064*	*319,064*	*256,236*	*256,236*	*256,236*
Service and sales workers								
	51	Personal service workers	0.000723*	0.00456***	0.00336***	0	0.00302**	0
			251,712	*251,712*	*251,712*	*260,092*	*260,092*	*260,092*
	52	Sales workers	0.000876**	0.00457***	0.00392***	0	0.00323***	0.00157**
			230,688	*230,688*	*230,688*	*275,344*	*275,344*	*275,344*
Skilled agricultural, forestry and fishery workers								
	61	Market-oriented skilled agricultural workers	0	0.00657***	0.00486**	0.00135*	0	0.00508**
			73,248	*73,248*	*73,248*	*32,696*	*32,696*	*32,696*
Craft and related trades workers								
	71	Building and related trades workers, excluding electricians	0	0.00378**	0	0	0.00528***	0.00231**
			278,940	*278,940*	*278,940*	*228,740*	*228,740*	*228,740*
	72	Metal, machinery and related trades workers	0	0.00482***	0.00219**	0.000437**	0.00495***	0.00235***
			1,476,708	*1,476,708*	*1,476,708*	*1,013,580*	*1,013,580*	*1,013,580*
Plant and machine operators, and assemblers								
	81	Stationary plant and machine operators	0	0	0	0	0.00319**	0
			583,536	*583,536*	*583,536*	*400,828*	*400,828*	*400,828*
	82	Assemblers	0	0.00365***	0	0	0.00400***	0.00171*
			895,104	*895,104*	*895,104*	*639,996*	*639,996*	*639,996*
	83	Drivers and mobile plant operators	0.00103**	0.00510***	0.00392***	0	0.00424***	0.00180**
			697,768	*697,768*	*697,768*	*555,728*	*555,728*	*555,728*
Elementary occupations								
	91	Cleaners and helpers	0	0	0	0	0	0
			198,544	*198,544*	*198,544*	*145,008*	*145,008*	*145,008*
	92	Agricultural, forestry and fishery labourers	0	0	0	0	0	0
			5,240	*5,240*	*5,240*	*2,848*	*2,848*	*2,848*
	93	Labourers in mining, construction, manufacturing and transport	0	0.00290***	0	0.000716*	0.00369**	0.00250**
			338,324	*338,324*	*338,324*	*237,380*	*237,380*	*237,380*

Notes: The table shows correlations between sickness-related absences and county-level exposure to influenza-like diseases by chosen occupations. Correlations are net of seasonal and regional effects. The exposure to sickness is measured as a normalized incidence of influenza. The number of observations are in italics. For clarity, the insignificant correlations are in the table substituted by 0. Standard age groups are used. Significance levels: *** 0.01, ** 0.05, * 0.1.

Table 2.10: Sickness-related absence - controlling for sickness rates

	Controlling for sickness	Not controlling for sickness
After	-9.003***	-9.632***
	(0.280)	(0.271)
Observations	15,327,196	15,327,196
Adjusted R2	0.031	0.031

Notes: The table shows two regression results from Equation 2.1. The dependent variables are sickness-related absences. The results in the first column control for sickness exposure (influenza outbreak among adults) whilst the results in the second column control only for periods after the change. Other controls included: age, tenure, gender, collective agreement, quarter, year, county, industry, occupation, firm size cat., educ. cat., nationality. Cluster-robust standard errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

Table 2.11: Hours Absent - Respiratory infections exposure

	Adults' exposure			Children's exposure		
	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave
After	-4.813***	-0.0771	-3.080***	-6.901***	2.729**	-1.271*
	(0.343)	(1.666)	(0.867)	(0.400)	(1.227)	(0.750)
Sickness	0.00570***	-0.00276***	-0.00253***	0.000967***	-0.00138***	-0.000713***
	(0.000296)	(0.000940)	(0.000633)	(0.0000979)	(0.000379)	(0.000203)
After*Sickness	-0.00606***	0.0135***	0.00964***	-0.00106***	0.00190***	0.00159***
	(0.000470)	(0.00152)	(0.000945)	(0.000140)	(0.000270)	(0.000190)
Observations	15,327,196	15,326,330	15,318,628	15,327,196	15,326,330	15,318,628
Adjusted R2	0.031	0.301	0.357	0.031	0.300	0.356

Notes: The table shows two sets of regression results (Equation 2.1). We use normalized adult sickness rates in the first three columns and normalized children's sickness rates in the last three columns. The dependent variables are: total absences, sickness-related absences, paid and unpaid leave. Controls include: age, tenure, gender, collective agreement, quarter, year, county, industry, occupation, firm size cat., educ. cat., nationality. Cluster-robust standard errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

Table 2.12: Hours absent - Infectious diseases other than respiratory (EPIDAT)

	Sickness absences	Paid leave	Unpaid leave
After	-9.269***	9.515***	4.237***
	(0.349)	(1.447)	(0.644)
Group 1	-0.00275***	0.00371***	0.00121**
	(0.000245)	(0.000944)	(0.000511)
After * Group 1	0.00127***	-0.00566***	-0.00466***
	(0.000307)	(0.000809)	(0.000485)
Group 2	0.00285***	-0.00330**	-0.00253**
	(0.000650)	(0.00148)	(0.00123)
After * Group 2	-0.00350***	0.00816***	0.00945***
	(0.000760)	(0.00216)	(0.00129)
Group 6	0.000132**	0.000404**	0.000379***
	(0.0000550)	(0.000169)	(0.000115)
After * Group 6	-0.000200***	-0.000198	-0.000260**
	(0.0000609)	(0.000214)	(0.000131)
Group 11	0.0103***	-0.0186**	-0.0104*
	(0.00200)	(0.00913)	(0.00592)
After * Group 11	-0.0162***	0.0285***	0.0233***
	(0.00256)	(0.00833)	(0.00526)
Observations	15,327,196	15,326,330	15,318,628
Adjusted R^2	0.031	0.300	0.356

Notes: The table shows regression results from Equation 2.1. We use the normalized incidence of selected infectious diseases from the EPIDAT database as defined in 2.3 (Group 1: *Intestinal infectious diseases*; Group 2: *Other bacterial diseases*; Group 6: *Viral diseases affecting skin*; Group 11: *Louses and similar*). The dependent variables are: sickness-related absences, paid and unpaid leave. Controls include: age, tenure, gender, collective agreement, quarter, year, county, industry, occupation, firm size cat., educ. cat., nationality and constant term. Cluster-robust errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

Table 2.13: Subsample of Fathers - Influenza outbreak

	Adults' outbreak			Children's outbreak		
	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave
After	-7.383***	6.996***	2.310***	-6.488***	3.131**	-0.940
	(0.293)	(1.346)	(0.534)	(0.321)	(1.285)	(0.754)
Sickness	0.814***	-0.0815	-0.282*	0.126***	-0.286***	-0.263***
	(0.0793)	(0.164)	(0.145)	(0.0113)	(0.0462)	(0.0326)
After*Sickness	-1.030***	0.448	0.182	-0.148***	0.368***	0.322***
	(0.333)	(0.591)	(0.271)	(0.0118)	(0.0385)	(0.0305)
Observations	5,329,632	5,329,325	5,327,098	5,329,632	5,329,325	5,327,098
Adjusted R2	0.031	0.268	0.330	0.031	0.270	0.331

Notes: The table shows two sets of regression results (Equation 2.1) for the fathers' subsample (males 26-47 y.o.). We use counts of weeks with epidemics in a quarter for adults in the first three columns and for children in the last three columns. The dependent variables are: sickness-related absences, paid leave, and unpaid leave. Controls include: age, tenure, gender, collective agreement, quarter, year, county, industry, occupation, firm size cat., educ. cat., nationality and constant term. Cluster-robust errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

Table 2.14: Hours Absent: shift-work classification

	All			High			Low		
	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave
After	-7.922***	3.583***	-0.732	-7.375***	5.809***	1.100*	-8.070***	2.769**	-1.376*
	(0.310)	(1.079)	(0.688)	(0.447)	(0.855)	(0.615)	(0.335)	(1.178)	(0.762)
Sickness	0.149***	-0.269***	-0.235***	0.182***	-0.140***	-0.127***	0.143***	-0.302***	-0.265***
	(0.0110)	(0.0380)	(0.0251)	(0.0132)	(0.0366)	(0.0234)	(0.0121)	(0.0414)	(0.0286)
After × Sickness	-0.183***	0.376***	0.323***	-0.208***	0.164***	0.189***	-0.179***	0.440***	0.370***
	(0.0111)	(0.0336)	(0.0262)	(0.0141)	(0.0308)	(0.0180)	(0.0124)	(0.0391)	(0.0306)
Observations	14,253,940	14,253,076	14,245,396	3,144,660	3,144,514	3,141,398	11,109,280	11,108,562	11,103,998
Adjusted R^2	0.031	0.295	0.354	0.024	0.305	0.348	0.033	0.294	0.359

Notes: The table shows three sets of regression results (Equation 2.1). We divide the observations into quartiles (based on the shift-work classification) and present results for all observations in the 1st three columns, those in the top quartile (columns 4-6), and those up to the 75th percentile (columns 7-9) separately. We measure the sickness by counting the number of weeks with epidemic status in a quarter, using the incidence of influenza among children. The dependent variables are: sickness-related absences (sickness absence), paid and unpaid leave. Controls include: age, tenure, gender, collective agreement, quarter, year, county, industry, occupation, firm size cat., educ. cat., nationality. Cluster-robust errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

Table 2.15: Hours Absent (by occupations differently Exposed to Disease or Infections)

	All			High			Low		
	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave
After	-7.215***	3.311***	-1.434***	-5.877***	7.908***	1.084*	-7.551***	2.232*	-2.012***
	(0.334)	(1.181)	(0.521)	(0.676)	(1.027)	(0.584)	(0.362)	(1.347)	(0.567)
Sickness	0.131***	-0.277***	-0.236***	0.183***	-0.183***	-0.163***	0.120***	-0.289***	-0.242***
	(0.0119)	(0.0459)	(0.0281)	(0.0197)	(0.0503)	(0.0373)	(0.0127)	(0.0530)	(0.0300)
After $\times$ Sickness	-0.168***	0.369***	0.331***	-0.208***	0.139***	0.147***	-0.156***	0.424***	0.374***
	(0.0129)	(0.0359)	(0.0244)	(0.0233)	(0.0514)	(0.0279)	(0.0140)	(0.0409)	(0.0276)
Observations	6,027,004	6,026,560	6,024,372	1,309,800	1,309,596	1,309,434	4,717,204	4,716,964	4,714,938
Adjusted R^2	0.032	0.298	0.360	0.029	0.232	0.259	0.034	0.319	0.392

Notes: The table shows three sets of regression results (Equation 2.1). We divide the observations into quartiles (based on the O*NET classification) and present results for all observations in the 1st three columns, those in the top quartile (columns 4-6), and those up to the 75th percentile (columns 7-9) separately. We measure the sickness by counting the number of weeks with epidemic status in a quarter, using the incidence of influenza among children. The dependent variables are: sickness-related absences (sickness absence), paid and unpaid leave. Controls include: age, tenure, gender, collective agreement, quarter, year, county, industry, occupation, firm size cat., educ. cat., nationality. Cluster-robust errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

Table 2.16: Hours Absent (by occupations with different Contact With Others)

	All			High			Low		
	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave
After	-7.215***	3.311***	-1.434***	-4.291***	6.484***	0.213	-7.794***	2.398*	-1.923***
	(0.334)	(1.181)	(0.521)	(0.491)	(1.034)	(0.511)	(0.376)	(1.266)	(0.559)
Sickness	0.131***	-0.277***	-0.236***	0.169***	-0.176***	-0.166***	0.122***	-0.293***	-0.246***
	(0.0119)	(0.0459)	(0.0281)	(0.0182)	(0.0441)	(0.0343)	(0.0132)	(0.0507)	(0.0303)
After × Sickness	-0.168***	0.369***	0.331***	-0.169***	0.0742*	0.156***	-0.166***	0.443***	0.374***
	(0.0129)	(0.0359)	(0.0244)	(0.0194)	(0.0399)	(0.0231)	(0.0146)	(0.0394)	(0.0274)
Observations	6,027,004	6,026,560	6,024,372	1,251,616	1,251,533	1,251,324	4,775,388	4,775,027	4,773,048
Adjusted R^2	0.032	0.298	0.360	0.025	0.261	0.282	0.034	0.310	0.384

Notes: The table shows three sets of regression results (Equation 2.1). We divide the observations into quartiles (based on the O*NET classification) and present results for all observations in the 1st three columns, those in the top quartile (columns 4-6), and those up to the 75th percentile (columns 7-9) separately. We measure the sickness by counting the number of weeks with epidemic status in a quarter, using the incidence of influenza among children. The dependent variables are: sickness-related absences (sickness absence), paid and unpaid leave. Controls include: age, tenure, gender, collective agreement, quarter, year, county, industry, occupation, firm size cat., educ. cat., nationality. Cluster-robust errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

Table 2.17: Hours Absent (by occupations with different Social Orientation)

	All			High			Low		
	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave
After	-7.215***	3.311***	-1.434***	-4.907***	5.351***	-0.537	-7.725***	2.654**	-1.738***
	(0.334)	(1.181)	(0.521)	(0.520)	(1.043)	(0.495)	(0.369)	(1.275)	(0.573)
Sickness	0.131***	-0.277***	-0.236***	0.170***	-0.211***	-0.187***	0.122***	-0.284***	-0.241***
	(0.0119)	(0.0459)	(0.0281)	(0.0176)	(0.0452)	(0.0320)	(0.0131)	(0.0507)	(0.0307)
After × Sickness	-0.168***	0.369***	0.331***	-0.185***	0.116***	0.185***	-0.163***	0.435***	0.368***
	(0.0129)	(0.0359)	(0.0244)	(0.0194)	(0.0420)	(0.0234)	(0.0145)	(0.0393)	(0.0277)
Observations	6,027,004	6,026,560	6,024,372	1,310,676	1,310,587	1,310,395	4,716,328	4,715,973	4,713,977
Adjusted R^2	0.032	0.298	0.360	0.028	0.266	0.290	0.034	0.309	0.382

Notes: The table shows three sets of regression results (Equation 2.1). We divide the observations into quartiles (based on the O*NET classification) and present results for all observations in the 1st three columns, those in the top quartile (columns 4-6), and thosev up to the 75th percentile (columns 7-9) separately. We measure the sickness by counting the number of weeks with epidemic status in a quarter, using the incidence of influenza among children. The dependent variables are: sickness-related absences (sickness absence), paid and unpaid leave. Controls include: age, tenure, gender, collective agreement, quarter, year, county, industry, occupation, firm size cat., educ. cat., nationality. Cluster-robust errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

Table 2.18: Hours Absent: job-cell size interaction

	Adults' outbreak			Children's outbreak		
	Sickness absences	Paid leave	Unpaid leave	Sickness absences	Paid leave	Unpaid leave
After	-4.599***	9.215***	1.083***	-3.946***	3.914***	-2.758***
	(0.390)	(0.280)	(0.155)	(0.469)	(0.382)	(0.255)
Size	1.276***	-0.0449	0.0233	1.072***	0.336***	0.533***
	(0.100)	(0.0712)	(0.0313)	(0.119)	(0.0966)	(0.0594)
Sickness	0.478***	0.167**	-0.283***	0.0563**	-0.229***	-0.153***
	(0.151)	(0.0818)	(0.0721)	(0.0230)	(0.0195)	(0.0161)
After*Sickness	-0.00352	0.706*	0.908***	-0.105***	0.523***	0.417***
	(0.555)	(0.407)	(0.280)	(0.0287)	(0.0268)	(0.0203)
Sickness*Size	0.0441	-0.104***	-0.0586***	0.0213***	-0.0404***	-0.0512***
	(0.0390)	(0.0252)	(0.0219)	(0.00659)	(0.00574)	(0.00469)
After* Sickness*Size	-0.129	-0.0172	-0.0514	-0.0124	0.00911	0.0146**
	(0.164)	(0.128)	(0.0867)	(0.00888)	(0.00912)	(0.00647)
Observations	310,945	310,945	310,945	310,945	310,945	310,945
Adjusted R2	0.101	0.486	0.557	0.101	0.490	0.560
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variables are average absence hours (sickness-related absences, paid and unpaid leave) in job cells. The sickness variable counts the number of weeks with epidemic status in quarter using the incidence of influenza among children. Controls include: indicator of periods after the reform and its interaction with the size of job-cells, age, tenure, gender, collective agreement, quarter, year, county, occupation, firm size cat., educ. cat., nationality, job-cell fixed effects. Cluster-robust errors in parentheses. Significance levels: *** 0.01, ** 0.05, * 0.1.

Chapter 3

Forced migration, staying minorities, and new societies: Evidence from post-war Czechoslovakia[1]

3.1 Introduction

The global number of displaced people is at new record highs, with violent conflicts and wars at the root of most forced migration and ethnic cleansing.[2] Forced migration has immediate dramatic consequences for the displaced and for the communities that become their new homes. There are also long-term effects on the displaced and on their descendants, documented by a large literature (for surveys, see Ruiz and Vargas-Silva, 2013; Becker and Ferrara, 2019). However, ethnic cleansing is never complete, as some members of the displaced ethnicity always manage to evade expulsion and become members of newly created societies (for examples, see Bell-Fialkoff, 1993; Kaufmann, 1996). Little is known about

[1] **Co-authored with Štěpán Jurajda (CERGE-EI) and Felix Roesel (ifo Institute Dresden).**

[2] Of the 70 million displaced people worldwide today, over 20 million were forced to leave their country (UNHCR data as of March 2020).

such 'stayers' and the way they integrate into their re-settled communities after ethnic cleansing—communities in which they become a minority without moving from the homes of their ancestors.[3] Are the consequences of ethnic cleansing for stayers as long-lasting and multi-generational as for the displaced? Do those who escape forced migration integrate into the new majority or do they segregate and cultivate their own ethnic identity? Can stayers act as a 'small seed' of development and take an active role in forming the identity of their new re-settled communities, the way that migrants entering established societies sometimes do?[4] Answering these questions is important for understanding ethnic cleansing. It can also shed light on community-identity formation, since stayers are more strongly rooted locally than the new incoming majority settlers, but, similarly to migrants, they are a minority in their new societies.

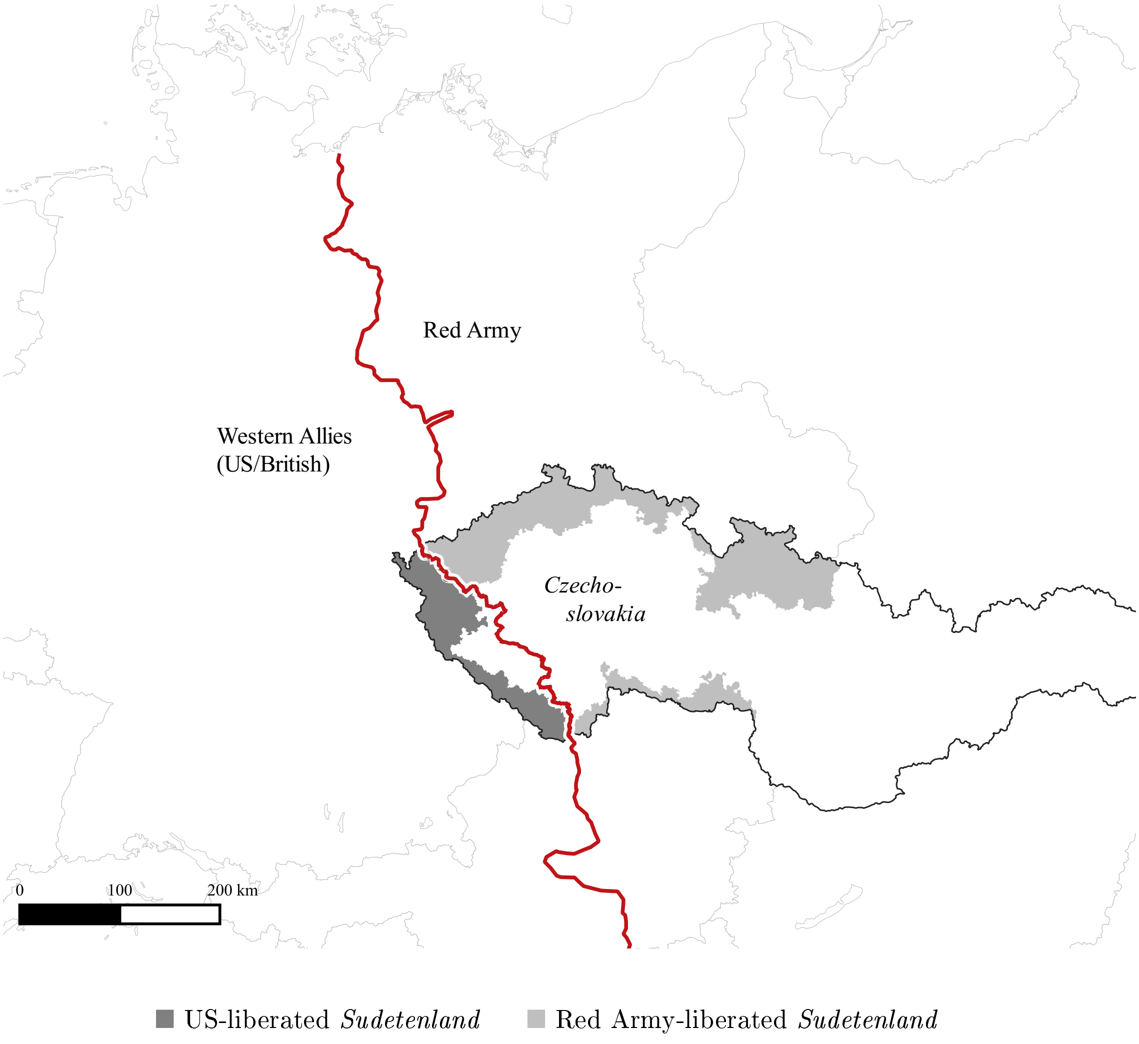
In this paper, we study the footprint of the staying German minority that evaded Czechoslovakia's expulsions after World War Two. Based on the Beneš Decrees, three million ethnic Germans were forced to leave *Sudetenland*—a region in the Czech borderlands that was predominantly populated by ethnic Germans prior to the war (see the gray shaded region in Figure 3.1).[5] However, some 200,000, mainly anti-fascists and industrial workers, avoided deportation. We exploit quasi-experimental local variation in the extent and structure of deportations that allowed more anti-fascist Germans to stay in some areas. This variation was the result of the US Army liberating parts of Czechoslovakia, which in turn was the consequence of the unexpected military progress of the US Army through Germany in the spring of 1945. The line of contact with the Red Army (Figure 3.1), which divided *Sudetenland* between May and December 1945, did not coincide with any pre-existing geographic, administrative, or ethnic boundaries. The almost straight line was drawn to connect US troops in Germany and Austria.

[3]A handful of studies shows lasting differences between ethnically cleansed areas and neighboring regions with no ethnic cleansing (Acemoglu et al., 2011; Chaney and Hornbeck, 2016; Arbatli and Gokmen, 2018; Becker et al., 2020; Testa, 2020). There is also evidence on the local economic impacts of the expulsion of Jews on Nazi Germany (Waldinger, 2010, 2012; Akbulut-Yuksel and Yuksel, 2015; Huber et al., 2020) and of slave trade on affected African countries (Nunn, 2008; Nunn and Wantchekon, 2011).

[4]Ochsner and Roesel (2020) and Giuliano and Tabellini (2020) show that migrants can affect the long-term political identity of their new residence communities.

[5]Ethnic cleansing in post-war Europe uprooted a total of 20 million Belarusians, Germans, Hungarians, Poles, Ukrainians, and others (Schechtman, 1953).

Figure 3.1: Line of contact in the final days of World War Two in Europe (May 1945)



Notes: The red line is the line of contact where the Western Allies (mainly British and US forces) and the Red Army met in May 1945. The gray lines correspond to national boundaries as of 1930. The gray shaded area in Czechoslovakia represents *Sudetenland*—a region settled by around three million Germans, which was annexed by Nazi Germany in October 1938. The US-liberated part of *Sudetenland* is in dark gray, the Red Army-liberated part in light gray.

The US Army immediately locked its Czechoslovak zone in May 1945 and prevented early ('wild') expulsions of ethnic Germans. On the other side of the demarcation line, Czech officials began to expel Germans immediately after liberation, supported by the Red Army, which also recruited thousands of anti-fascist *Sudeten* Germans to help build the Communist party in the Soviet occupation zone in Germany, as anti-fascists were typically strongly aligned with the Communist party (Pecka, 1995; Gerlach, 2007; Řeháček, 2011). This opened a gap across the demarcation line in the share of deported Germans, and anti-fascist Germans in particular. When mass organized deportations started in early 1946, anti-fascists became entitled to stay in Czechoslovakia. At that time, the Red Army had already cleared its zone of a large number of anti-fascist Germans. Thus, the 1945 demarcation line in *Sudetenland* amounts to a natural experiment varying the local presence of anti-fascist Germans staying in post-war Czechoslovakia. This natural experiment occurred in the only region of post-war Europe where forced migration was at least temporarily controlled by the US Army, rather than by the Red Army.[6]

Sudetenland was quickly re-settled by about two million Czechs, Slovaks, and other nationals. The quasi-random variation in the presence of left-leaning German stayers in post-war *Sudetenland* allows us to ask novel questions: Do stayers who escape forced migration influence their re-settled communities, and do they assimilate into the new majority or do they maintain their minority identity? We investigate these questions by contrasting neighboring regions within *Sudetenland*, separated by the 1945 demarcation line between the US and the Red Army. We use a spatial regression discontinuity (RD) framework and study ethnic identity, political attitudes, social policies, and election outcomes using both individual-level data and new community-level data hand-collected from German and Czech archives.

Our results imply a lasting political legacy of staying anti-fascist Germans. Today's Communist party vote shares, density of local Communist party cells, and Communist party membership rates are higher where the presence of US forces led to more anti-fascist Germans avoiding deportation. The effects are sizable.

[6] Our analysis is thus the first to directly contrast the consequences of ethnic cleansing in areas under US as opposed to Red Army control.

Ten anti-fascist German stayers after World War Two account for three to four votes for the Communist party in Czech national elections today. The Czech Communist party is one of the least reformed of the formerly ruling Communist parties of Central and Eastern Europe.[7] Therefore, our main findings, together with the absence of any effects on central-left parties, signal long-term persistence of far-left political preferences. Geocoded survey data eliciting political values corroborate our main findings and show stronger preferences for redistribution, planned economies, and authoritarianism in places where more anti-fascist Germans stayed. German surnames among local Communist elites in the 1950s and among local-election Communist-party candidates today allow us to trace our main findings to the post-war presence of anti-fascist German stayers. We also rule out other potential mechanisms behind our main findings, including post-war resettlement, changes in industrial structure, selective mobility, and direct effects of liberation by the US or the Red Army.

While we uncover strong evidence of the political legacy of stayers, we do not find any spatial discontinuity across the demarcation line in self-declared German ethnicity. Post-war Czechoslovakia eliminated the use of German in public life (in schools, administration, and employment) and, according to our findings, the outcome of this forced assimilation did not interact with the size of the stayer community.[8] Our findings thus imply that staying anti-fascist Germans transmitted their political identity across three generations, but not their German identity, and their far-left political identity may have supplanted their German ethnic identity. The expression of political identity by the offspring of stayers is not merely an opportunistic survival strategy within the Czechoslovak communist regime, because the far-left political values we measure correspond to free and democratic elections in the modern Czech Republic up to 2018, long after the fall of the Iron Curtain. Stayer parents deciding on which of the two main

[7] Along with the Moldovan Communist party, it is the only former ruling party in post-Communist Europe, which has not dropped 'Communism' from its name. It has never been part of a governing coalition in the Czech Republic. The party's platform remains close to its original agenda, its youth organisation was banned from 2006 to 2010, and there have been repeated calls from other parties to outlaw the party.

[8] Such interactions are a feature of models of cultural identity (e.g., Bisin and Verdier, 2001), in which parental and peer socialization are substitutes. Language restrictions can heighten the sense of cultural identity, as observed by Fouka (2020) for the German minority in the US after World War One.

identities (German or far-left) to inculcate in their children reflected an environment that supported one, but suppressed the other identity. This is consistent with Egan (2020), who shows that ethnic identity can be adjusted in response to political identity, and, more generally, with the growing literature suggesting that integration decisions by minorities respond to incentives (Algan et al., 2020; Fouka, 2019; Atkin et al., 2020). The existing literature, however, studies how immigrants integrate into an existing majority (Bisin et al., 2011, 2016; Verdier and Zenou, 2017), while our setting offers a view of an ethnic group that does not re-locate, but becomes a minority in a re-settled new society.

To the best of our knowledge, we provide the first evidence implying that a small minority of stayers can affect attitudes and values of societies after ethnic cleansing.[9] Only a handful of studies exploit local variation in the intensity of ethnic cleansing. Arbatli and Gomtsyan (2019) uncover ethnic-cleansing origins of a current nationalist party identification in Armenia—origins that survived seven decades of Soviet rule. In Poland, preferences for public goods and redistribution increase in cultural diversity measured as the share of staying Germans not expelled after World War Two, a finding similar to ours (Charnysh, 2019).[10] In our study, we are able to trace today's place-based political outcomes to the small group of stayers exempted from displacement over 70 years ago. Furthermore, while the extent of forced displacement analyzed in existing studies may be endogenous, a key feature of our research design is the exogenous variation in the local intensity of forced migration induced by the quasi-random line of contact between US and Red Army forces in 1945 Czechoslovakia. This enables us to ask whether non-displaced individuals from an ethnic minority can have causal long-term effects on the political identity of their newly resettled communities. Our findings provide support for the 'small seed' theory of political development (Giuliano and Tabellini, 2020).

Our results complement recent related work on migrants and political values.

[9] A related literature investigates the effects of voluntary emigration on family members left behind (for example, Beine et al., 2008; Antman, 2011, 2012; Ivlevs et al., 2019). For a survey see Antman (2013). In related research, it has been shown that traumatic war experiences have lasting effects on the political identity of local communities (for example, Blattman, 2009; Rozenas et al., 2017; Fontana et al., 2017).

[10] Becker et al. (2020) also study Poland, but focus on values of forced migrants, not on stayers and sending regions.

Ochsner and Roesel (2020) find that far-right voting is more pronounced today in Austrian regions that have absorbed more Nazis fleeing the Red Army, i.e., that a small number of arriving migrants with radical political values can shape long-term local political equilibria in *established* communities. In comparison, our evidence suggests that a small group of stayers, i.e., non-migrants, with strong political values, is also sufficiently powerful to influence political outcomes in *newly formed* societies. The findings by Ochsner and Roesel (2020) and by Arbatli and Gomtsyan (2019) are consistent with the transmission of far-right and nationalist political values, respectively, across several generations, in line with a growing body of research highlighting the persistence of far-right political values (for example, Voigtländer and Voth, 2012; Cantoni et al., 2020; Jurajda and Kovač, 2021).[11] Our study supports the notion that *far-left* political values are similarly strongly transmitted across generations, and can survive transitions across political and economic systems as well as ethnic cleansing episodes. This is a new insight in the growing literature discussing the historical roots of populism and extremism (e.g., Grosfeld and Zhuravskaya, 2015; Ochsner and Roesel, 2017; Avdeenko, 2018).

Although we primarily contribute to the literature on the political and ethnic identity consequences of forced migration,[12] our analysis also brings novel findings to the research exploring various effects of the line of contact between Red Army troops and US and British forces in 1945 Europe (Fontana et al., 2017; Eder and Halla, 2016, 2018; Ochsner, 2017; Martinez et al., 2020). While the demarcation line in Austria and Germany divided homogeneous societies, the line of contact in Czechoslovakia cut through both the Czech-populated lands of the Nazi-occupied *Protectorate of Bohemia and Moravia* (hereafter, the 'Czech main lands') and *Sudetenland*—the German-populated region of Czechoslovakia incorporated into Nazi Germany between 1938 and 1945. Our analysis is the first to investigate the

[11] Other papers have documented persistence in socioeconomic outcomes beyond political values, for example, Acemoglu et al. (2001); Alesina and Fuchs-Schündeln (2007); Nunn (2008); Dell (2010); Brosig-Koch et al. (2011); Nunn and Wantchekon (2011); Becker et al. (2016); Valencia Caicedo (2018).

[12] We study the effects on sending regions of *Sudetenland* while Bauer et al. (2013) and Braun and Dwenger (2020) explore the economic and political impacts of arriving displaced Germans on their destinations in Germany; Semrad (2015) studies similar questions and focuses on expellees from Czechoslovakia.

demarcation line in Czechoslovakia, which was divided between US and Red Army forces between May and December 1945.[13] This allows us to contrast the effects of US versus Red Army liberation across two qualitatively different settings. We find short-term population declines in German-inhabited regions liberated by the Red Army (similar to findings from Austria and Germany, where such declines were long-term, Ochsner, 2017; Eder and Halla, 2018), but no population declines in the Czech-populated regions initially under Red Army control. This is in line with anecdotal evidence that Red Army soldiers treated Slavic people and Germans differently (Řeháček, 2011; Glassheim, 2016, among others) and suggests that the faster progress of US and British forces in 1944/1945 may have reduced post-war violence and acts of revenge.

3.2 Historical background

3.2.1 *Sudeten* Germans in the Czech lands

Prior to World War Two, Czechoslovakia hosted one of the largest German-speaking minorities outside Germany. The borderlands of Czechoslovakia, *Sudetenland*, were home to three million ethnic Germans representing about 30% of the population of the Czech lands (Bohemia, Moravia, and Silesia) in 1930.[14] Ethnic Germans began settling in *Sudetenland* during the rule of Ottokar I of Bohemia at around 1200. By 1930, German and Czech communities were sharply divided: in three of four counties of the Czech lands in 1930, either self-declared German or Czech ethnicity accounted for more than 90% of the population.[15] Tensions between Czechs and Germans surfaced after Czechoslovakia broke away from the Habsburg Empire in 1918. There were separate political parties for both ethnic communities along the entire political spectrum, with the exception of the ethnicity-bridging Czechoslovak Communist Party (*Komunistická strana Československa*, KSČ). Nationalism among *Sudeten* Germans accelerated after

[13]Guzi et al. (2019) and Testa (2020) compare the evolution of social capital, population, and economic outcomes across the border *between* the former *Sudetenland* and the neighboring Czech main lands. We study differences in outcomes *within* the formerly German-populated part of Czechoslovakia as well as within the Czech-populated main lands.

[14]Figure 3.4 in the Online Appendix shows the population of the Czech lands between 1921 and 2011.

[15]Section 3.10.2 in the Online Appendix reports our sources for census statistics.

Adolf Hitler seized power in Germany in 1933. The Sudeten German Party (*Sudetendeutsche Partei*) supported the annexation of *Sudetenland* to Germany and won two thirds of the *Sudeten* German vote in the 1935 Czechoslovak election.

Nazi Germany annexed *Sudetenland* in September 1938 as a result of the Munich Agreement, followed by a first wave of ethnic cleansing. About 175,000 Czechs, including 25,000 Jews, were forced to leave *Sudetenland* (Němeček, 2002). When Nazi Germany unleashed World War Two in September 1939, *Sudetenland* was fully incorporated into the *Reich* and the remaining Czech lands became the Nazi-administered territory of the 'Protectorate of Bohemia and Moravia'. After Germany's surrender in May 1945, national boundaries as of 1937 were restored immediately, and *Sudetenland* returned to Czechoslovakia. In a second, reversed wave of ethnic cleansing, almost the entire German population was expelled from *Sudetenland* during 1945 and 1946 and replaced by about two million Czechs, Slovaks, and other nationals. However, some 200,000 Germans stayed, corresponding to about 6% of the pre-war population. After decades of continuous assimilation, some 39,000 citizens—less than 0.4% of present-day Czech Republic's 10 million population—declared German ethnicity in 2001.[16]

3.2.2 Demarcation line in 1945 Czechoslovakia

It was neither intended nor foreseeable that US forces and the Red Army would meet in Czechoslovakia in May 1945. The Yalta Conference in February 1945 had already informally allocated Czechoslovakia to the Soviet post-war sphere of influence. However, military developments in the final weeks of World War Two altered the original plan. The German Western front collapsed after British and American forces crossed the Rhine river in March 1945. In the East, by contrast, the German resistance against the Red Army was still substantial. During March and April, the Soviets gradually agreed to the further eastward progress of the US forces, but they stressed their ambition to liberate the Vltava valley including the Czech capital of Prague. In the heavy battles of April 1945, the Red Army prioritized Germany's and Austria's symbolic capitals of Berlin and Vienna, and did not make significant progress into the Czech lands in between. The US

[16] In 2001, 31,000 (1.0%) of the 3.1 million residents in *Sudetenland* declare German ethnicity.

Army, by contrast, had already liberated large parts of Germany and Austria, and demanded to connect their troops standing at the German Elbe and Mulde rivers with US troops along the Danube river in Austria (see, Franzel, 1967, and Figure 3.1). The Soviets accepted General Eisenhower's proposal for a more or less straight demarcation line formed by the Czech cities of Karlovy Vary (Carlsbad), Plzeň (Pilsen), and České Budějovice (Budweis).

US troops approached the Czech part of the demarcation line on May 5 and stopped there.[17] When Nazi Germany ultimately surrendered on May 8, the US Army controlled a strip of around 10,000 square kilometers in western Czechoslovakia and was waiting for the Red Army, which stood some 200 kilometers east of Prague and arrived a few days later. The red line in Figure 3.1 shows the final position of the demarcation line as reported by Pecka (1995). The line cut through *Sudetenland* as well as the Czech-populated former 'Protectorate'. It followed roads and railways[18] and it did not coincide with any pre-existing geographic, administrative, or ethnic boundaries. The exception was its southernmost part (south of the village of Žernovice, see Figure 3.6 in the Online Appendix), where the line somewhat overlapped with the border of *Sudetenland*, i.e., with ethnic divisions. In all of our analysis, we thus omit this southernmost part of the line. Both the Red Army and the US Army locked up their zone's borders as of May 1945 (Pogue, 1954; Dickerson, 2006). *Sudeten* Germans thus had a very limited opportunity to self-select into fleeing either zone.[19]

[17] Eisenhower attempted to shift the line of contact eastward to include Prague. This time, however, Soviet General Antonov rejected the plan. General Patton, who commanded the US forces in the region, was then not allowed to progress towards Prague in early May (Mendelsohn, 2010, p. 14).

[18] The line overlaps with main roads and railways, 27% and 45% respectively in a 500 meter buffer. See Figure 3.5 in the Online Appendix.

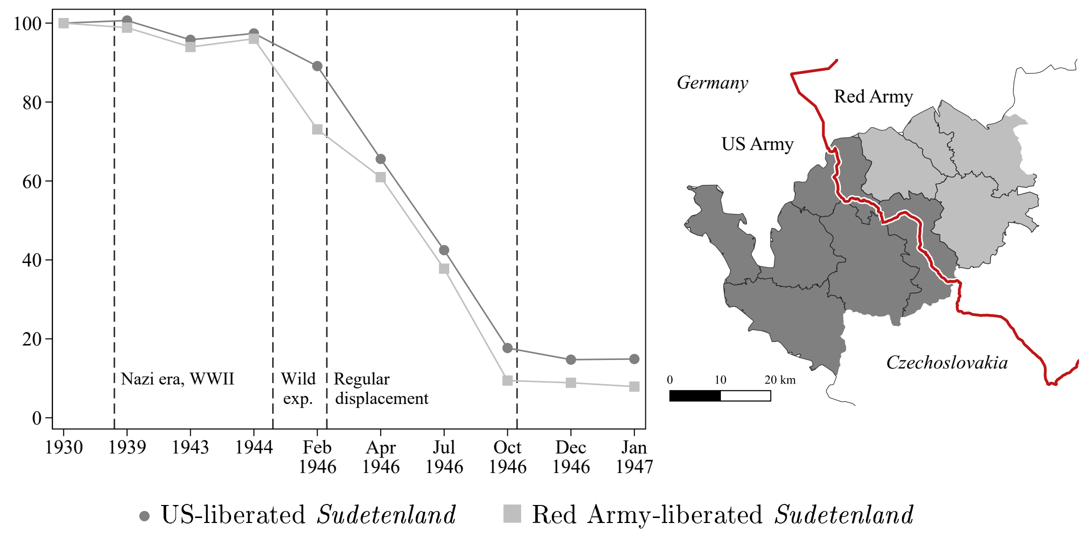
[19] Crossing the demarcation line was possible only with permits from both Soviets and Americans and one had to return by the end of the day (Fischer and Kodet, 2013) The Red Army frequently opened fire on those crossing the line illegally (Řeháček, 2011). The US Army as well as the Red Army implemented similar restrictions to the re-installed Czech-German border. US soldiers burnt all belongings of illegal migrants from *Sudetenland* at the German border and sent them back (Brandes, 2001). After December 1945, all borders to Germany and Austria were under strict Czechoslovak control.

3.2.3 Expulsion of Germans from Czechoslovakia

In regions controlled by the Red Army, the expulsion of *Sudeten* Germans from Czechoslovakia began immediately after Germany's surrender (Brandes, 2001). At least 700,000 *Sudeten* Germans were displaced in 'wild expulsions' in the Red-Army zone between May and July 1945, and thousands were killed (Suppan, 2006; Glassheim, 2016). The US forces, by contrast, prevented any displacement of Germans at this stage (Slapnicka, 2000). Therefore, the number of staying Germans was substantially larger in the US zone by December 1945 when both US and Red Army forces left Czechoslovakia. Figure 3.2 traces the German population in % of the 1930 population in US and Red Army-liberated counties along the northern half of the demarcation line in *Sudetenland*, where we have collected rare monthly population data during the expulsions. There is no difference in population dynamics before 1945. At the end of 1945, around 90% of the German population as of 1930 was still living on the US side, while in the Red Army-controlled areas approximately one of three Germans had already been expelled.

The second stage of expulsions occurred between February and October 1946. These organized (regular) mass deportations covered two million *Sudeten* Germans from both the formerly US and Red Army zone (Řeháček, 2011; Bundesministerium für Vertriebene, Flüchtlinge und Kriegsgeschädigte, 1957). Figure 3.2 shows that these organized expulsions never fully closed the initial gap across the demarcation line in the extent of displacement. A total of around 240,000 Germans lived in Czechoslovakia when the last mass transports left in October 1946 (Luža, 1964), though another few thousand Germans left during 1947 and 1948. In post-war Czechoslovakia, the remaining 200,000 Germans were not allowed to practice their language, their movement was restricted, and inter-ethnicity marriages required government approval (Kučera, 1992). German identity faded. The 1950 Czech census counted 160,000 self-reported Germans (Reindl-Mommsen, 1967), a substantial decrease despite very little out-migration. After decades of assimilation, less than 40,000 Czech citizens reported German ethnicity by 2001.

Figure 3.2: Germans in US- and Red Army-liberated regions (in % of 1930 population)



Notes: The graph on the left compares the share of staying Germans in % of the 1930 population in the US and Red Army-liberated counties corresponding to the northern half of the *Sudetenland* demarcation line. The corresponding map on the right shows the primarily US-liberated counties in dark gray, while the Red Army-liberated counties are in light gray. The 1947 counties of Aš, Cheb, Kraslice, Loket, Sokolov, and Vildštejn sum up to the US region, the Red Army-liberated region is the sum of the counties of Horní Blatná, Jáchymov, Karlovy Vary and Nejdek. The red line in the map represents the demarcation line between US and Red Army forces between May 1945 and December 1945. The first two dashed vertical lines in the graph bracket the period from the annexation of *Sudetenland* by Nazi Germany in October 1938 to Germany's surrender in May 1945. The second set of vertical lines corresponds to the presence of US forces in western Czechoslovakia (April/May 1945 to December 1945) and 'wild expulsions' in Red Army-liberated *Sudetenland*. The period of organized mass displacement of Germans from *Sudetenland* (February to October 1946) corresponds to the third bracketed period. For sources, see Section 3.10.2 in the Online Appendix.

3.2.4 Anti-fascist Germans

The German stayer community in post-war Czechoslovakia consisted primarily of indispensable industrial workers and anti-fascists.[20] *Sudetenland* was a highly industrialized region with mining, heavy industries, and manufacturing. About 100,000 indispensable German specialists and their families were allowed (often forced) to stay where significant industries were present. The second main group of German stayers consisted of about 100,000 anti-fascists (Kučera, 1992), who were certified by local authorities (national commitees, *národní výbory*). German elite anti-fascists, the Communist party (KSČ), and the Social Democratic party (ČSSD) were typically involved in the certification process (Foitzik, 1983; Schneider, 1995).[21] Certified anti-fascists chiefly consisted of (pre-war) members of the Czechoslovak Communist party and the Social Democratic party, as well as Germans active in the anti-Nazi resistance.

Three mechanisms gave rise to local over-representation of anti-fascist German stayers in regions liberated by US forces. First, in the 'wild expulsions' that occurred in the Red Army zone in the summer of 1945, ethnicity was often the only selection criterion and so Nazi Germans and anti-fascist Germans were often treated equally and expelled together (Turnwald, 1951; Schneider, 1995; Klepsch, 2013). The absence of 'wild expulsions' in the US zone thus opened a gap in the number of Nazi Germans and also anti-fascist Germans across the demarcation line. Second, an agreement between the Soviet administration in Germany and the Czechoslovak government increased this gap for anti-fascist Germans.[22] The Soviets aimed to roll out Communist party cells in its East German zone as fast as possible. Communist party membership was high in many parts of *Sudetenland*, but almost no party structures existed in the rural north of the Soviet zone in Germany. As a result, some 30,000 anti-fascist Germans left Czechoslovakia for East Germany in prioritized transfers in 1945 (Foitzik, 1983), and these early leavers

[20] A small number of German Jews, Germans married to Czechs, and individuals granted mercy were also allowed to stay.

[21] For a detailed description of the certification process, see, for example, 'Směrnice pro ověřování antifašistů', published in newspapers in Liberec on 25 July 1945 (Hoffmann et al., 2010, p. 673–674).

[22] See, the documents in Bundesministerium für Vertriebene, Flüchtlinge und Kriegsgeschädigte (1957, p. 343-355) and Schneider (1995).

came from the Red Army-controlled part of *Sudetenland*.[23] Third, when organized mass displacement started in 1946, anti-fascist Germans became entitled to stay. Because of the two processes discussed above, more anti-fascist Germans were still present at this point (and thus could stay) in the US-liberated parts of *Sudetenland*. Wilde (2015) notices a remarkably high number of anti-fascist Germans in the county of Sokolov located on the US side of the demarcation line.

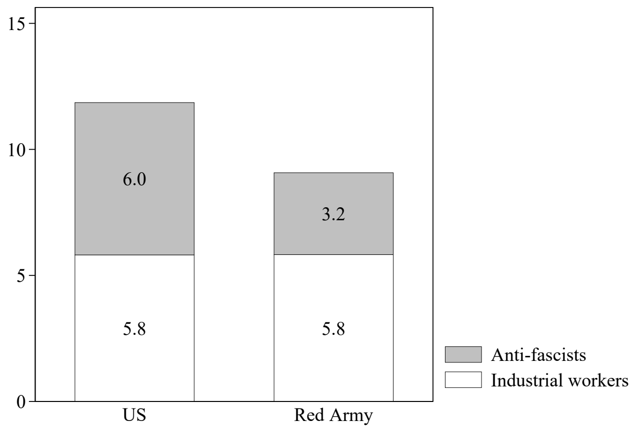
To directly explore the nature of the gap in staying Germans, we went to local archives on both sides of the demarcation line, and collected data from handwritten lists at the municipality level on the total number of Germans in late 1946 when mass transfers were completed (Figure 3.7 in the Online Appendix provides samples). These lists count Germans by the reason they were allowed to stay. We were able to gather data for three counties divided by or in close proximity to the demarcation line (Karlovy Vary, Kraslice, and Loket). The lists distinguish anti-fascists and industrial specialists.[24] We relate these counts to the 1930 local German population and compute averages for 76 US-liberated and Red Army-liberated municipalities. Figure 3.3 shows the results. Corroborating Figure 3.2, we find that more Germans stayed on the US side (12% of the 1930 population) than on the Red Army side (9%).[25] A similar share of 6% of the former German population stayed as industrial specialists on either side of the demarcation line. By contrast, we observe a higher share of German certified anti-fascists on the US side of the demarcation line: 6% in terms of the 1930 population as opposed to 3% on the Red Army side. Thus, the entire gap in the share of the staying German population between US and Red Army-liberated regions can be explained by the numbers of anti-fascists. This evidence supports the notion that the initial presence of US and Red Army forces created different local trajectories of German displacement, particularly for the anti-fascists.

[23] Schneider (1995) reports that in the Red Army-liberated county of Ústí nad Labem (formerly Außig) all Communists had already departed for East Germany by May 1946.

[24] We add the small number of Germans in mixed marriages, German Jews, and other exceptions to industrial specialists. Anti-fascists include Germans subject to potential later deportation and Germans receiving 'special treatment' or who were granted citizenship, as these are likely to be anti-fascists as of late 1946.

[25] Figure 3.2 reports 15% and 9% of the German population staying in December 1946 in the US and the Red Army sections of our North *Sudetenland* sub-sample, respectively, consistent with the municipalities covered in Figure 3.3 being representative of the entire North sub-sample.

Figure 3.3: Staying Germans after expulsions by entitlement (in % of 1930 population)



Notes: The figure shows how the staying German population in neighboring US- and Red Army-liberated regions of *Sudetenland* after the end of organized mass transports in late 1946 (in % of 1930 population) breaks down into different legal entitlements. Data were hand collected from local archives in Karlovy Vary and Sokolov. The sample consists of 76 municipalities (US Army: 22, Red Army: 54) in the counties of Karlovy Vary, Kraslice and Loket. Industrial workers also include the few Germans exempt from displacement based on Jewish origin, high age, and mixed marriage. The anti-fascist group includes certified anti-fascists and Germans subject to potential future deportation, who are likely to be anti-fascists as of late 1946).

Anecdotal evidence suggests that the staying anti-fascist Germans were powerful and prominent actors in the Communist regime. Urban (1964, p. 36) reports that 'a considerable share of the Germans who are allowed to stay are senior Communists', some of them being 'even more fanatic Communists than Czechs'.[26] In 1948, the Czechoslovak Communist party (KSČ) took control of the government of Czechoslovakia and introduced a Stalin-style regime lasting until 1989. Anti-fascist Germans, such as the violin maker Josef Pötzl living in US-liberated *Sudetenland*, made it to the Czech parliament in the 1950s as Communist MPs.[27] Table 3.9 in the Online Appendix compares the names of around 550 Communist county-level party leaders in 1959 on both sides of the demarcation line, hand-collected from local archives. We find that the share of German surnames among these leaders on the US side of the line is about 3 percentage points higher than on the Red-Army side.[28] This is consistent with the gap in the share of staying

[26] Original in German, translation by the authors.

[27] Other examples of KSČ MPs of German ethnicity are Jan Jungbauer and Rudolf Müller.

[28] The methodology for identifying German as opposed to Slavic names is discussed in Section 3.6.

anti-fascists reported above. Staying anti-fascist Germans actively contributed to building Communism in Czechoslovakia. Below, we investigate how deep and lasting their impact has been.

Overall, both the presence of US forces in Czechoslovakia and the location of the demarcation line were the result of unexpected military events. The line of contact did not follow any previous boundaries and it corresponded to separate governance of the two zones until the end of 1945. It induced a quasi-experimental difference in ethnic cleansing and, specifically, in the presence of left-leaning German stayers in post-war Czechoslovakia.

3.3 Data

We compile a new dataset of Czech municipalities covering the interwar period and the era after World War Two. It includes information on the last national election in the interwar period (1935) and in the Czech Republic (1996 to 2017). We also collect data on democratic national elections in Czechoslovakia (1946, 1990, 1992) which, however, are not directly comparable to other elections because Germans were not eligible to vote and deportations and resettlement were still ongoing in May 1946 or because municipalities were consolidated into large units during the Communist regime, affecting the 1990 and 1992 data. This information is then translated to the territorial status of the present-day 6,244 Czech municipalities. After excluding the capital city of Prague, the average Czech municipality has a population of about 1,500. As some of the municipality-level information is not available prior to World War Two, we rely on pre-war information at the level of the 330 Czech counties as of 1947 with an average population of about 25,000.[29] We also use the 2010 and 2016 waves of the Life in Transition Survey (LITS), for which we are able to geo-code the residence of the respondent. The LITS asks respondents in Central and Eastern European countries about their political values and attitudes. We combine all data with information on the location of the 1945 demarcation line, which we reconstruct based on the report by Pecka (1995) (see, Section 3.10.4 in the Online data appendix).

[29]We use historical GIS information on boundaries of former Czech counties and regions, and on the national boundaries of 1930 Europe.

The Online Appendix 3.10 describes how we retrieved and processed data from digitized hardcover copies, local and national archives, and both hand-collected and administrative sources. Election data are obtained from the Czech Statistical Office, including local (municipal) election outcomes between 1994 and 2018 with the corresponding candidate names.[30] We digitize population data from 1930 and 1950 census hardcover publications. In addition, we collect data on the German population from local archives in Sokolov and Karlovy Vary, from the archives of the Czech Ministry of Foreign Affairs, and from various monographs. Further population data come from the Czech Statistical Office and from the German Statistical Office for *Sudetenland* counties annexed by Nazi Germany between 1938 and 1945. Data on local monuments and memorials and on German names are retrieved from various websites listed in the appendix. Finally, we rely on several publications for information on the deportation of Germans after the war, the names of local Communist party elites in the 1950s, and the bombings during World War Two.

3.4 Identification

Differences in expulsion policies across the demarcation line in *Sudetenland* (discussed in Sections 3.2.3 and 3.2.4) led to quasi-experimental variation in the local presence of staying anti-fascist Germans. We rely on this variation within a regression discontinuity design to estimate its causal effects on political identity and ethnicity. In this section, we outline our econometric approach and analyze the exogeneity of the demarcation line location. Our two main outcomes of interest are the extent of self-declared German ethnicity and the vote share of the Czech Communist party (KSČ, KSČM since 1990). The latter is a natural choice of a political identity measure since anti-fascist German stayers were closely aligned with the Communist party and generally likely to support left-wing values (see Section 3.2.4). The Communist party was the ruling party between 1948 and 1989 and its direct successor is the leading far-left party in the Czech Republic.[31]

[30]The exception are data for the 1946 election which we retrieve from hardcover copies.

[31]Figure 3.8 in the Online Appendix depicts Communist national vote shares separately for the (former-Protectorate) Czech main lands and for *Sudetenland*; since 1990 they vary between 10% and 20% in both parts of the Czech Republic.

3.4.1 Regression discontinuity design

Our identification strategy is to exploit the natural experiment of the demarcation line and to compare areas close to the line, assuming that neighboring US and Red Army-liberated areas share similar trends and unobserved characteristics prior to the mass expulsion of Germans. We test this assumption in the next section. Adjacent areas under Red Army control thus provide a counterfactual for US-liberated regions where displacement took place later, was less extensive, and displaced fewer anti-fascist Germans.[32]

We apply a spatial regression discontinuity (RD) design (Lee and Lemieux, 2010) to the most granular data available—municipalities. Our preferred specification corresponds to a local-linear RD strategy (Calonico et al., 2017), but we use a parsimonious polynomial RD regression model as a reference and a starting point (Gelman and Imbens, 2019). This model is estimated with OLS and allows for standard errors robust to spatial correlation (Conley, 1999, 2010):

$$\begin{aligned} Communist_i = \alpha + \beta_1 US_i + \beta_2 Distance_i + \beta_3 Distance_i^2 + \\ \beta_3 Distance_i \times US_i + \beta_4 Distance_i^2 \times US_i + X_i'\gamma + \epsilon_i. \end{aligned} \tag{3.1}$$

Here, $Communist_i$ denotes the vote share for the Communist party in a national election in Czech municipality i. We also use other political outcomes as dependent variables later. The vector of β coefficients refers to a quadratic RD polynomial interacted with a dummy variable US_i taking on the value one if a municipality was liberated by US forces in 1945 (zero otherwise). $Distance_i$ measures the great circle distance of a municipality to the demarcation line in kilometers. Distances are positive on the Red Army side and negative on the US side. X_i is a vector of municipality-level geography controls (distance to the German border, distance to the next main road, distance to the next railway line, mean altitude and slope as the difference between maximum and minimum altitude) and population controls (logged pre-war population and logged present-day population). We restrict this least-squares estimation to municipalities ± 25 kilometers around the demarcation line; the rationale for this bandwidth choice

[32] In our main analysis we focus on the demarcation line within *Sudetenland*, but we perform a similar analysis also for the demarcation line within the Czech main lands.

is provided in Section 3.4.2. We exclude the few municipalities divided by the demarcation line, so our dataset covers four types of municipalities: *Sudetenland* and former-Protectorate (Czech main lands) municipalities which were allocated either to the US or the Red Army zone in 1945.[33]

Most of our RD analysis is then based on flexible RD specifications corresponding to to the local-linear procedure with a data-driven optimal bandwidth choice proposed by Calonico et al. (2017). We report RD standard errors robust to optimal bandwidth choice (Calonico et al., 2014; Hyytinen et al., 2018). In these specifications, we do not pre-define any maximum bandwidth around the demarcation line. However, the optimal bandwidth ends up being close to that used in our reference polynomial specification.

3.4.2 Exogeneity of the demarcation line

Geographical RD estimates have a meaningful causal interpretation only if the cut-off location is set quasi-randomly and if self-selection is ruled out. Self-selection of Germans into the US or the Red Army zone was prevented by the fact that the ultimate location of the line was not known to the public as it was the result of unforeseen military developments in the last few weeks of World War Two, and by the severe restrictions on individual mobility applied by both liberating forces upon their arrival (see Section 3.2.2 for details).

To provide statistical evidence on the absence of pre-war differences across the demarcation line formed in May 1945, we test for discontinuities using the local-linear RD method proposed by Calonico et al. (2017). In Table 3.1, we provide such a test for *Sudetenland* and the Czech main lands separately in columns (1) and (2), respectively, and then combining both areas in column (3). All pre-war characteristics balance well at the later demarcation line, including 1930 ethnicity, religion, population density and growth, including geographical features as well as the extent of bombing during the war. The only exception is the distance to the external border with Germany, which is somewhat higher on the US side

[33]We also exclude municipalities divided by the border between *Sudetenland* and the Czech main lands (former Protectorate) as well as municipalities south of the village of Žernovice, where the demarcation line corresponded with ethnic divisions. See the maps in Figures 3.6 and 3.9 in the Online Appendix.

within *Sudetenland* municipalities. The maximum optimal bandwidths across the three geographic areas (columns) in Table 3.1 are 14, 20, and 28 kilometers. We therefore set 25 kilometers on either side of the demarcation line as our bandwidth choice in the few specifications where the optimal bandwidth procedure is not available.

Table 3.10 in the Online Appendix further shows no significant pre-1930 differences across the demarcation line in municipality population and housing (relative to 1930 levels). However, we do find a discontinuity in total population directly after the expulsions (in 1950), which is in line with less extensive deportations, and thus less depopulation in the US zone. Finally, in Table 3.11 in the Online Appendix we use county-level data on Communist election outcomes in 1935. We compare Communist vote shares in counties with a maximum distance of 25 kilometers of the county capital to the eventual demarcation line. We find no significant differences in election outcomes before displacement; if anything, Communist vote shares were slightly lower in the later US zone. Given the empirical support for the quasi-random location of the RD line and the likely absence of self-selection, we conclude that our RD strategy allows for a causal interpretation.

3.5 Results

3.5.1 Communist party vote shares

Our baseline results in Table 3.2 provide robust evidence of long-run effects of the presence of US forces in 1945 *Sudetenland* on the electoral success of the Czech Communist party. Applying a quadratic-interacted RD polynomial in column (1), we find the vote share of the Czech Communist party in the 2017 national election to be about 9 percentage points higher as one steps across the demarcation line from the most western Red Army-liberated *Sudetenland* municipalities to adjacent municipalities under US control.[34] Point estimates do not change and effects become more precisely estimated when we control for local geography and for pre-war and present-day population in column (2). These findings are confirmed in our preferred RD specification, where we allow for flexible local-linear

[34]Figure 3.10 in the Online Appendix shows the corresponding RD plot.

Table 3.1: Balancing of pre-displacement covariates at the US–Red Army demarcation line

	Sudetenland	Czech main lands	Full line
	(1)	(2)	(3)
Census 1930			
Population (log)	0.103	0.090	-0.108
	(0.488)	(0.302)	(0.232)
Population growth 1921–1930	-0.578	0.328	-0.034
	(0.744)	(0.446)	(0.301)
Population density	-0.632	0.217	-0.061
	(0.679)	(0.237)	(0.189)
Czechs %	-0.024	-0.006	0.091
	(0.023)	(0.004)	(0.098)
Germans %	0.024	0.005	-0.093
	(0.028)	(0.003)	(0.099)
Foreigners %	0.006	0.002	0.002
	(0.011)	(0.002)	(0.002)
Catholics %	0.041	0.012	-0.011
	(0.056)	(0.071)	(0.065)
Protestants %	0.005	0.011	0.011
	(0.014)	(0.024)	(0.020)
Geography			
Distance to external border	10.875*	0.647	3.269
	(5.988)	(3.614)	(4.069)
Minimum altitude	21.250	3.557	-2.342
	(54.231)	(21.589)	(19.238)
Mean altitude	30.543	-7.369	-18.155
	(74.391)	(22.633)	(27.966)
Maximum altitude	-3.943	-24.277	-43.275
	(86.227)	(27.527)	(35.966)
Slope (altitude range)	-35.325	-22.746	-33.913
	(55.221)	(18.360)	(22.596)
Military events			
War bombings	0.061	0.051	0.047
	(0.046)	(0.073)	(0.059)
Controls	No	No	No
Max. bandwidth	28.412	14.839	20.393
Max. obs.	211	347	624

Notes: The table shows the effect for US-liberated regions (RD estimates) at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. We use a local-linear RD procedure including a data-driven optimal bandwidth choice (Calonico et al., 2017). The unit of observation are municipalities, the dependent variables are pre-war characteristics (1930 census), geographical characteristics, and military operations during World War Two. Column (1) shows estimates for *Sudetenland*, i.e., for the regions historically settled by ethnic Germans, column (2) refers to the Czech main lands, while column (3) pools both parts of Czechoslovakia. We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Population growth 1921–1930 refers to the average annual growth rate. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.2: Communist votes in national election

	Communist vote share 2017					
	Sudetenland			Czech main lands		
	Para-metr. RD	Para-metr. RD	Local-lin. RD	Para-metr. RD	Para-metr. RD	Local-lin. RD
	(1)	(2)	(3)	(4)	(5)	(6)
US zone 1945	0.094***	0.094***	0.079***	0.002	0.002	0.004
	(0.026)	(0.022)	(0.026)	(0.013)	(0.013)	(0.017)
Geography contr.	No	Yes	No	Yes	Yes	No
Population contr.	No	Yes	No	Yes	Yes	No
Mean dep. var.	0.107	0.108	0.107	0.107	0.107	0.105
RD bandwidth	25.000	25.000	17.739	25.000	25.000	13.346
Eff. obs.	186	185	125	572	572	313
R^2	0.798	0.832	–	0.800	0.814	–

Notes: The table shows the effect for US-liberated regions (RD estimates) at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia based on a parametric (quadratic-interacted) polynomial approach without/with control variables (columns (1), (2), (4), and (5), bandwidth: 25 km) and a local-linear RD specification including a data-driven optimal bandwidth choice (Calonico et al., 2017). The units of observation are municipalities, the dependent variable is the vote share of the Communist party (KSČM) in the 2017 Czech national elections. Columns (1) to (3) show estimates for regions originally settled by ethnic Germans (*Sudetenland*), columns (4) to (6) refer to the Czech main lands. We exclude municipalities south of Žernovice, where ethnicity divides corresponded with the demarcation line. Geography controls are the distance to the external (German) border, distance to the nearest main road, distance to the nearest railway line, mean altitude and slope (difference between maximum and minimum altitude). Population controls are logged population in 1930 and logged present-day population. Significance levels (Conley (2010) standard errors/robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

polynomials and rely on an optimal data-driven bandwidth: In column (3) of Table 3.2, we find a statistically significant effect of 8 percentage points in the Communist vote share at the demarcation line within *Sudetenland*. Since the local-linear RD specification is the most flexible of the four alternatives, we use it as a baseline in what follows.

Within *Sudetenland*, the different expulsion policies in the US and Red Army zones led to a higher share of anti-fascist Germans on the US side of the demarcation line. In the Czech main lands, however, there were almost no Germans as of 1947 and thus no meaningful difference in the share of staying Germans across the demarcation line.[35] If the presence of US forces affects present-day Commu-

[35]Figure 3.1 and Figures 3.6 and 3.9 in the Online Appendix show how the demarcation line cut through both the German-populated areas and the Czech main lands.

nist vote shares via the anti-fascist German channel, one would expect no effects within the Czech main lands. This is indeed born out in columns (4) to (6) of Table 3.2, where we uncover precisely estimated zero effects for the part of the demarcation line cutting through the Czech main lands, consistent with effects operating through German stayers. The Czech main lands here provide a placebo test of our interpretation of the *Sudetenland* effects. Point estimates for the Czech main lands are also significantly different from those for *Sudetenland*.[36] We return to discussing the mechanisms underlying our baseline *Sudetenland* findings in Section 3.6.

3.5.2 Other election outcomes

The baseline findings are fully robust to various sensitivity and sub-sample checks (all based on the local-linear RD approach). First, in Table 3.12 in the Online Appendix we split the *Sudetenland* municipalities near the demarcation line to a north and a south sub-sample (based on the latitude of the village of Bezvěrov, see Figure 3.6 in the Online Appendix). The two estimated effects are both similar to the baseline effect from column (3) of Table 3.2 and they are not statistically distinguishable.

Second, we estimate the effects of various pseudo treatments, for which we expect to find no effects if our identification and inference strategy is valid. Table 3.13 in the Appendix (columns (1) and (2) as well as (4) and (5)) show precisely estimated zero effects when we move the demarcation line 25 kilometers eastwards or westwards. We also use the Ohře river as a pseudo demarcation line. Unlike the North-South demarcation line, the river cuts *Sudetenland* from east to west. Again, we find no significant change in the Communist vote at this alternative pseudo cut-off.

Third, we extend our analysis from the 2017 Czech national election to all national elections since the Czech independence. Table 3.14 in the Online Appendix reveals that the 2017 effects are very similar to those in all other national elections since Czech independence in 1993. In columns (1) and (6), we uncover strong effects

[36]We estimate difference-in-discontinuities models pooling observations in columns (1) and (4) as well as (2) and (5). The differences are statistically significant at the 5% and 1% level, respectively (t-values 2.21 and 2.63).

on the Communist party vote shares within *Sudetenland* and precisely estimated zero effects in the Czech main lands. The only exception is the 1998 *Sudetenland* effect (p-value of 0.103). However, once we add other fringe far-left parties to account for the fragmented far-left camp in the 1990s, we find, in column (2), a highly significant 7-percentage-points effect of the US zone in *Sudetenland*. The Communist vote share effects are largest in 2002 and 2013 when the party received its best and second-best results after Czech independence. We have also attempted to study the three free Czechoslovak elections in 1946, 1990 and 1992. The elections in the early 1990s are, however, not comparable to post-1993 elections in democratic Czech Republic.[37] The 1946 election is a specific case in that the deportation of Germans was in full swing, Czech parties competed on an anti-German platform, and Germans including anti-fascists were not eligible to vote. We discuss the 1946 election in Section 3.6.3 in more detail.

In the remaining columns of Table 3.14, we extend our analysis beyond the Communist party. We divide the party spectrum into far-left, centrist parties (mainstream), and far-right. Column (3) implies that the higher Communist (far-left) vote share within *Sudetenland* comes at the cost of electoral success of mainstream parties, where we find mirrored decreases at the demarcation line. Far-right vote shares and voter turnout are not affected in most observed elections. We also do not find significant discontinuities for the centrist-populist ANO 2011 party as well as when we pool all votes cast for populist parties in the 2017 national election.[38] We conclude that the Communist vote share effects are related to far-left ideology, not to populism in general. We consistently obtain no statistically significant or sizeable estimates within the Czech main lands (columns (6) to (10)). We also zoom in on the election results of the Social Democratic party (ČSSD). Both Communist and Social Democratic Germans were certified as anti-fascists. Early transports of anti-fascist Germans to the Soviet zone, however, mainly tar-

[37]Municipality boundaries in 1990 and 1992 do not coincide with the territorial status of municipalities we use in our main analysis. This is due to heavy consolidation of municipalities during the Communist era, which obscures allocation of municipalities to either *Sudetenland* or the Czech main lands as well as allocation across the demarcation line. It took several years after the Velvet Revolution to dissolve and split thousands of municipalities again. Therefore, the 1996 election data are the first offering reliable municipality territorial status information. All of the estimated Czechoslovak-elections coefficients were statistically insignificant.

[38]The corresponding p-values of the RD estimates are $p = 0.786$ and $p = 0.401$, respectively.

geted Communists (see Section 3.2.4). We would therefore expect a difference in radical far-left but not in moderate left-wing votes across the former demarcation line. Column (1) of Table 3.15 in the Online Appendix confirms our expectation in that there are no effects of the presence of US versus Red Army troops in *Sudetenland* on the vote shares of the Social Democratic party.

Finally, we ask about the effect of the line on the presence of local Communist party cells. We collect data on all local (municipal) elections in the Czech Republic between 1994 and 2018 and code whether the Communist party stands in a given municipality. We pool all local elections to measure long-term Communist party structures. Table 3.16 in the Online Appendix reports the results of RD estimations. Municipalities on the US side of the demarcation line are about 12% more likely to host a local Communist party cell. Thus, we find not only more Communist voters but also more active Communist party structures where anti-fascist Germans stayed in larger numbers after 1945 thanks to the presence of the US Army.

In sum, vote share effects for the Communist party are persistent and robust, and they are related to the activity of local party structures. The presence of US troops does not *per se* increase far-left votes—we find no effects at the demarcation line in the Czech main lands. A prime explanation for the pattern of our findings is that the staying anti-fascist Germans transmitted their political identity across three generations. We discuss evidence supporting this hypothesis in Section 3.6, which is devoted to exploring possible mechanisms underpinning our main findings. At the end of Section 3.6 we also return to the issue of the overall magnitude and interpretation of the estimated vote share effects. But first, in the next section we extend our analysis beyond voting behavior as we study political values and party membership on either side of the demarcation line.

3.5.3 Communist party membership and political values

Given the absence of free elections during the Communist regime, our main analysis studies election outcomes after the Velvet Revolution. However, household surveys allow us to study also the Communist era before 1989. Specifically, we employ waves II (2010) and III (2016) of the Life in Transition Survey (LITS), which asks respondents in Eastern and Central Europe about their values and attitudes. Importantly for our analysis, respondents were also asked about their membership in the Communist party before 1989.[39] Both waves include information on the location of the respondents, which enables us to geo-code the data. However, the municipality-clustered sampling of respondents limits the extent of variation in the distance to the demarcation line. We therefore use a simplified RD approach. Instead of controlling for an RD polynomial, we control for latitude and longitude and again manually limit observations to a bandwidth of 25 kilometers around the demarcation line.[40] Of the 2,500 observations for the entire Czech Republic, we use 126 observations in *Sudetenland* and 197 in the Czech main lands. We control for age and gender of the respondents, and for survey years, and compare conditional outcome means across the line in probit and ordered probit specifications.

Table 3.3 shows the LITS results for *Sudetenland* in column (1) and for the Czech main lands in column (2). Respondents or their relatives living on the formerly US side of the line in *Sudetenland* were statistically significantly more likely to be members of the Communist party prior to 1989. During the Communist regime, party membership did not always imply full conviction. Mareš (2008) reports that ordinary Communist party members often joined the party for career rather than ideological reasons. However, our results imply not only higher Communist party membership on the US side of the demarcation line, but also stronger left-wing values. Respondents in US-liberated regions of *Sudetenland* are significantly more likely to be in favor of redistribution in order to close the gap between the rich and the poor, prefer planned economies over markets, and accept authoritarianism replacing democracy.[41] By contrast, we find no effects of the demarcation line within the Czech main lands on any of the LITS outcomes in line with our main findings, see column (2).[42] Again, the absence of any effects across the line in the Czech main lands is consistent with the *Sudetenland* effects being driven by the

[39] Present-day party membership is not available in the LITS data.

[40] Again, we use only observations north of the municipality of Žernovice.

[41] We have also tested for differences in trust towards institutions and groups. Table 3.17 in the Online Appendix reveals hardly any statistically significant effects. Trust towards the government and foreigners tends to be lower on the US side of the demarcation line in *Sudetenland*.

[42] The exception is a somewhat higher probability to prefer authoritarianism, statistically significant at the 10% level (p=0.08).

difference in expulsion policies and the presence of anti-fascist German stayers. In sum, survey-data evidence on party membership and values are fully in line with our baseline Communist-party vote share estimates.

3.5.4 Social policies

Locally embedded left-wing values and preferences are likely to give rise to stronger social and redistribution-related policies. To study the issue, we collected data on local public infrastructure in Czech municipalities. We take the presence of health facilities and kindergartens as a signal of stronger social policies. We also consider water mains and schools, which are perhaps less likely to be associated with a left-wing agenda. On average, only one of two Czech municipalities provides a health facility or a kindergarten. We use a dummy variable indicating the presence of a given type of public infrastructure and again apply our preferred local-linear RD approach. Table 3.4 shows the results. We find a large and statistically significant positive increase in the presence of local health facilities and kindergartens in US-liberated regions where anti-fascist Germans stayed in *Sudetenland*. Again, the estimated effects are smaller and at best marginally statistically significant in the Czech main lands. We find no effects on the presence of water mains or schools. Overall, these findings suggest that the legacy of US Army liberation manifests itself not only in stronger left-leaning political values, but also in real-world outcomes.

Table 3.3: Communist party membership and values (LITS micro data)

	Sudetenland	Czech main lands
	(1)	(2)
Were you or any member of your family a member of the Communist Party prior to 1989?		
Responent, parents or other family member	0.690**	-0.045
	(0.288)	(0.098)
Economic values		
Gap between rich and poor should be reduced	1.712**	-0.020
	(0.694)	(0.193)
Prefered economic system		
Market economy	-0.958***	-0.015
	(0.277)	(0.101)
Sometimes planned economies	0.870***	0.013
	(0.330)	(0.074)
Does not matter	0.251	0.015
	(0.326)	(0.097)
Prefered government system		
Democracy	-0.728**	-0.002
	(0.284)	(0.097)
Sometimes authoritarianism	0.479	0.137*
	(0.309)	(0.078)
Does not matter	0.265	-0.148
	(0.271)	(0.092)
Geography controls	Yes	Yes
Sociodemographic controls	Yes	Yes
Year fixed effects	Yes	Yes
Bandwidth	25.000	25.000
Max. obs.	126	197

Notes: The table shows the marginal effects for US-liberated regions from probit specifications estimated at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia (Exception: Gap between rich and poor should be reduced: ordered probit, table shows the estimated coefficient). The units of observation are individual respondents in the Life in Transition Survey, the dependent variables are answers to survey questions. We pool survey II (2010) and III (2016) and include year fixed effects. Geography controls are longitude and latitude of the respondent. Socio-demographic controls are age and gender. We impose a 25 km bandwidth around the demarcation line. Column (1) shows estimates for regions originally settled by ethnic Germans (*Sudetenland*), column (2) refers to the Czech main lands. We exclude residents from municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.4: Presence of public infrastructure

	Infrastructure (yes = 1)							
	Sudetenland				Czech main lands			
	Health facility	Kinder-garten	Water main	School	Health facility	Kinder-garten	Water main	School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
US zone 1945	0.516**	0.596**	0.027	0.118	0.312*	0.104	-0.089	0.068
	(0.246)	(0.271)	(0.185)	(0.224)	(0.169)	(0.168)	(0.148)	(0.104)
Geography controls	No	No	No	No	No	No	No	No
Population controls	No	No	No	No	No	No	No	No
Mean dep. var.	0.418	0.488	0.863	0.265	0.225	0.334	0.714	0.119
RD bandwidth	21.370	18.013	24.695	19.117	10.650	14.200	13.080	13.638
Eff. obs.	158	127	183	136	267	338	308	319

Notes: The table shows the effects for US-liberated regions (RD estimates) at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. We use a local-linear RD procedure with a data-driven optimal bandwidth choice (Calonico et al., 2017). The units of observation are municipalities, the dependent variable is a dummy variable indicating the presence of a given type of local public infrastructure. Health facilities and schools measured as of 2016, kindergartens as of 2017, and water mains as of 2018. Columns (1) to (4) show estimates for regions originally settled by ethnic Germans (*Sudetenland*), columns (5) to (8) refer to the Czech main lands. We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

3.6 Mechanisms

More Germans, anti-fascists in particular, stayed in post-war *Sudetenland* on the US side of the demarcation line (Section 3.2.3). The US side also features stronger Communist vote shares (Section 3.5.1), far-left political values (Section 3.5.3), and social policies (Section 3.5.4). And these effects are conditional on the presence of German stayers as we consistently find no effects in the Czech main lands. This suggests that these effects operate through anti-fascist German stayers. In this section, we present additional evidence supporting the importance of this channel and discuss its magnitude. We also explore five other potential mechanisms that may explain differences at the demarcation line in Communist vote shares today. We find that the legacy of anti-fascist Germans is the only compelling channel through which the events of 1945 impact far-left attitudes in the present-day Czech Republic.

3.6.1 Germans

To provide further evidence on the importance of the German-stayer channel for the Communist vote-share effects, one would ideally study family backgrounds and social linkages of Communist voters. Although such information is not available, we can check for the presence of descendants of German stayers among Communist-party candidates running for municipality-council seats. Standing in local elections indicates a strong party affiliation; Communist candidates can be considered leading local far-left politicians. Candidates are not asked to disclose their ethnicity, but we can rely on a unique feature of non-anonymized election data: family names of candidates. Germanic and Slavic languages (German and Czech in our case) are highly distinguishable in terms of family names. Further, in the Czech context, German surnames, which indicate German ancestry, were not dropped with German ethnic identity (Beneš, 1998).[43] We thus collect surnames, residence, and party affiliation of all 1.3 million candidates standing in Czech local elections between 1994 and 2018. We then consult the family his-

[43] Some of the German names on local-election candidate lists likely correspond to Czech post-war settlers of *Sudetenland* who also have German ancestors, but whose German identity had been abandoned long before World War Two. Given the evenly structured resettlement populations at the demarcation line (documented in Section 3.6.3), however, it is likely that there is no discontinuity at the line in the share of Czech settlers with German family names.

tory research website *Forebears.io* to identify German names among candidates. Names most frequent to Germany and Austria are coded as German.[44] Quality checks confirm that this simple algorithm correctly classifies 9 in 10 names, with no accuracy gap between Communist and other candidates.[45] A total of 16% of all candidate names in our sample are found to be of German origin. We distinguish Communist-party candidates from those of all other parties.

There were more anti-fascist German stayers on the US side of the demarcation line in post-war *Sudetenland*. If they and their offspring were not disproportionately geographically mobile (see, Section 3.24), and if far-left values were transferred across generations within their families, one would expect a higher share of German surnames on Communist-party election lists in the US-liberated municipalities. We therefore apply our local-linear RD procedure to test whether the frequency of German names differs across the demarcation line. Column (1) of Table 3.5 presents evidence, which is fully in line with our hypothesis. The share of German names among Communist party candidates is around 15% higher where US troops were located in 1945, compared to adjacent Red Army-liberated municipalities (within the set of municipalities where the Communist party ran in local elections). This difference across the line is unique to the Communist party. German names on candidate lists of all other parties (irrespective of whether they ran in municipalities with or without a Communist party cell) are equally distributed across the former demarcation line, see column (2). Again, we find no effects of the demarcation line in the Czech main lands (columns (3) and (4)). We present results based on the most recent 2018 local elections, but all results hold when we pool all elections between 1994 and 2018.

We conclude that the different expulsion policies across the demarcation line are a prime channel to explain why we observe stronger Communist voting preferences,

[44]The spelling of some German names changed. For example, *Fischer* often became the homophonous *Fišer*. We account for such changes and use both the 'Czechified' surname and its German version. Names are classified as German if either the original or its homophonous match appears in the *Forebears.io* list.

[45]Typical German names are *Schneider*, *Meier* or *Süßner*; Czech names are, for example, *Novák*, *Svoboda*, or *Černý*. The Online Appendix provides details of the coding procedure. Four Czech- and German-speaking research assistants independently double-checked the outcomes for a subsample of around 780,000 names, i.e., more than half of our candidate data-set. In 87.8% of all candidates and in 86.7% of Communist candidates, the majority of research assistants confirmed the coding of our algorithm.

Table 3.5: German names in local elections

	% German candidate names 2018			
	Sudetenland		Czech main lands	
	Communist party	Other parties	Communist party	Other parties
	(1)	(2)	(3)	(4)
US zone 1945	0.152**	0.024	-0.114	-0.002
	(0.077)	(0.077)	(0.126)	(0.036)
Mean dep. var.	0.158	0.155	0.160	0.168
RD bandwidth	27.400	14.691	19.271	17.152
Eff. obs.	49	95	43	400

Notes: The table shows the effect for US-liberated regions (RD estimates) at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. We use a local-linear RD procedure with a data-driven optimal bandwidth choice (Calonico et al., 2017). The units of observation are municipalities, the dependent variable is share of German names on candidate lists in the 2018 local elections. Columns (1) and (2) show estimates for regions originally settled by ethnic Germans (*Sudetenland*), columns (3) and (4) refers to the Czech main lands. We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

party cell presence, and left-wing values and policies where more left-leaning Germans stayed after the presence of the US Army. While we are not able to provide direct evidence on inter-generational transmission of political values,[46] our findings are strongly consistent with German stayers inculcating their political values in their offspring. It is also plausible that anti-fascist Germans were able to spread their values within the newly re-settled communities after ethnic cleansing was over. We return to the issue of spillovers within the discussion of the magnitude of the estimated effects in Section 3.7.2.

3.6.2 Ethnic legacy

One may argue that our results are driven by the *German* and not by the *anti-fascist* identity of anti-fascist German stayers. In this section, we ask whether the political legacy of the demarcation line that we have uncovered corresponds to an expression of ethnic identity. German ethnic identity was systematically suppressed in post-war Czechoslovakia, where staying Germans experienced vari-

[46] Table 3.9 in the Online Appendix provides suggestive evidence for Communist county-level party leaders in 1959.

ous types of discrimination. They were not allowed to practise their language and were initially subject to movement and inter-ethnicity marriage restrictions. At the aggregate level, homogenization policies during the Communist era resulted in low levels of self-reported German identity today (see Section 3.2.3). The share of German names in the Czech Republic is considerably above its share of citizens self-declaring German ethnicity. Perhaps families of German stayers kept their German name but discarded their German past. This would be consistent with a literature suggesting that integration decisions by minorities respond to incentives (Algan et al., 2020; Fouka, 2019; Atkin et al., 2020). On the other hand, there are also studies of assimilation policies suggesting that in the face of discrimination, immigrants may invest less in assimilation and retreat into their ethnic enclaves.[47] Ethnic polarization can in turn spur conflict, political polarization, and segregated voting (Montalvo and Reynal-Querol, 2005; Segura and Fraga, 2008).

We know that there were more Germans stayers on the US side and that the anti-fascist German stayers were more easily integrated into the post-war Czech Communist regime.[48] Fouka (2019) suggests that initially more integrated minority sub-groups assimilate faster when exposed to a wave of discrimination. More generally, outcomes of forced assimilation interact with the size of the minority community in models of cultural transmission (e.g., Bisin and Verdier, 2001). Our research design based on the quasi-random location of the demarcation line allows us to ask whether assimilation outcomes vary by the size of the German stayer community, where a larger community corresponds to higher ex ante integration potential. However, in Table 3.6, we find no discontinuity in self-declared German ethnicity or any other ethnicity across the demarcation line today, despite the differing initial share of German (anti-fascist) stayers after World War Two (columns (1) to (4)).

One possible explanation for the lack of German ethnic identity effects is that the Communist take-over in 1948 facilitated the expression of far-left political values,

[47]For example, Fouka (2020) finds that language restrictions at schools directed at second-generation German Americans strengthened their sense of ethnic identity. See also Edin et al. (2003) on the economic effects of enclaves.

[48]Section 3.2.4 discusses the cases of German Communist MPs in the Czechoslovakian parliament.

Table 3.6: Ethnicity in the 2001 census

	Population share declaring a ethnicity			
	German	Czech	Moravian	Slovak
	(1)	(2)	(3)	(4)
US zone 1945	-0.022	0.003	-0.001	0.024
	(0.027)	(0.045)	(0.002)	(0.021)
Mean of dep. var.	0.032	0.886	0.001	0.042
RD bandwidth	21.597	19.806	21.93	15.01
Eff. obs.	160	143	160	99

Notes: The table shows RD estimates at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. We use a local-linear RD procedure with a data-driven optimal bandwidth choice (Calonico et al., 2017). The units of observation are municipalities, the dependent variables are the population shares self-declaring a given ethnicity in the 2001 Czech census. We present evidence for regions historically settled by ethnic Germans (*Sudetenland*). We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

such that political identity, through all stages of inter-generational transmission, may have fully supplanted ethnic identity for the group of anti-fascist Germans. Our research design provides no information on the cultivation of ethnic identity among staying German industrial workers, as there was no discontinuity in their presence across the demarcation line (Figure 3.3). However, we can again rely on the candidate names employed in Table 3.5 and ask how many original German names were 'Czechified'—a process in which German characters in names were replaced by homophonous Czech characters (e.g., *Fischer* becomes *Fišer*). On average, 80% of all names classified as German in our data underwent such a transformation. There are no reasons to expect the share of German names among stayers or settlers that was 'Czechified' *before* World War Two (a common practise long before the expulsions of *Sudeten* Germans) differed between the US- and the Red Army-liberated regions. We find no statistically significant spatial discontinuity in such 'Czechifications' across the demarcation line.[49] Thus, we conclude that there is no evidence for ethnic assimilation differences across the demarcation line.

[49]The coefficient of the corresponding RD estimate is -0.164 with a p-value of 0.354.

3.6.3 Resettlement by Czechs

Selective re-settlement of *Sudetenland* on either side of the demarcation line provides another plausible explanation for our main findings. Were settlers more likely to be Communists on the US side of the line? Most of the resettlement process was centrally organized by the Czech government and the Czech Communist party, and it is not clear why the party would aim to strengthen the share of Communists in areas that already had a higher share of anti-fascists. If anything, an ex ante plausible settler selection strategy would operate against our findings. However, several pieces of evidence suggest that the resettlement process was evenly structured across the demarcation line, and thus speak against the selective re-settlement hypothesis. First, the resettlement process did not result in differently sized populations on either side of the line, and it distributed re-settler nationalities evenly as well. Resettlement quotas were applied to level out any initial local population differences.[50] This is confirmed in Table 3.10 in the Online Appendix, which shows no long-run population effects of the demarcation line. Returning Germans also play no role. Once expelled, basically no German returned to Czechoslovakia. Similarly, restitution of former German business and private property was limited to rare cases and cannot drive our results.[51] Further, in Table 3.6, we do not observe any significant discontinuity in self-declared ethnicity of re-settlers.

Second, and most importantly, we do not find that Czech settlers in US-liberated regions were more likely to come from pre-war Communist 'hotspots' within the Czech main lands. We combine information on the origins of the new settlers from 1947 county-level migration matrices with pre-war voting results from the 1935 Czechoslovak election and find equal pre-war Communist support for re-settler sending areas on either side of the line. We compute the predicted numbers of Communists among settlers as the sum of 1935 Communist vote shares in the

[50] The government aimed at a minimum of 75% of the pre-war population. See, Wiedemann (2016).

[51] 'After the collapse of the Communist rule in Czechoslovakia, lawmakers decided to return property ownership to anyone from whom the Communists had confiscated, provided the confiscation occurred after the February 1948 coup. This effectively barred former German and Hungarian minorities from qualifying.' Source: Jolyon Naegele, 'Czech Republic: The Beneš Decrees – How Did They Come To Be And What Do They Mandate?', Radio Free Europe/Radio Liberty, 01 March 2002, `https://www.rferl.org/a/1098965.html`.

118 counties of the Czech main lands weighted by the number of settlers from each county (see Table 3.18 in the Online Appendix). We find this predicted vote share for the Communist party among settlers from the Czech main lands to be equal (at 11%) for all *Sudetenland*, for Northern *Sudetenland* (Karlovy Vary region), and for the neighboring US-liberated Sokolov county and the Red Army-liberated Karlovy Vary county. All counties close to or divided by the demarcation line are very similar in this regard (at 10 to 11%). We thus find no evidence for a Communist bias among settlers on either side of the line. The outcomes of the 1946 national election underpin this finding. The election took place in May 1946 when displacement was in full swing and resettlement was not yet finished. Germans were not eligible to vote and all parties competed on an anti-German platform. We do not find any statistically significant spatial discontinuities in the Communist vote share in the 1946 election, when only non-German re-settlers were eligible to vote (see Table 3.19 in the Online Appendix). We conclude that settlers to *Sudetenland* are unlikely to drive the results.

3.6.4 Industrial structure

Sudeten Germans were well known for their crafts and industrial production (Semrad, 2015). The German displacement after World War Two thus could have led to substantial economic consequences, as not all specialized pre-war jobs could easily be filled by Czech workers. A stronger decline of formerly German-staffed industries on the Red Army side of the line, where fewer Germans were allowed to stay, could have lowered the attraction of Communist ideas. However, the share of stayers who are designated as industrial workers is equal across the demarcation line where we can measure it (Figure 3.3). Further, there is no evidence that labor shortages affected industrial structures differently across the line. Tables 3.20 and 3.21 in the Online Appendix show no significant discontinuity in sectoral employment shares as of 1950 and 2001 based on applying our RD strategy to census data. The only exception is the agricultural sector, which is somewhat more pronounced in the former US zone of *Sudetenland* in 2001, but not in 1950; the effect for 2001 is also not robust to other RD polynomials.[52] Thus, we find

[52] When we use a parametric RD approach similar to that used in Table 3.2, column (1) or (2), p-values are 0.237 and 0.129, respectively.

no robust evidence for shifts in sectoral shares. Long-run population and housing figures also do not diverge between the US and Red Army-liberated regions, as shown in Table 3.10. Bombing during the war, and hence, presumably, industrial destruction, also did not differ across the demarcation line (Table 3.1). Altogether, we see little reason to believe that changes in the structure of the economy drive our main results.

3.6.5 Memories of war and liberation

Thus far, our analyses of population, industry-structure, ethnicity, and political identity have not uncovered significant differences across the demarcation line within Czechoslovakia, with the exception of political identity discontinuities within *Sudetenland*. We focused on the presence of anti-fascist German stayers in *Sudetenland*, but local memories of violent acts of liberating troops against civilians are also likely to be limited to the historically German-settled regions. In particular, anecdotal evidence suggests that Red Army rapes and shootings were less extensive when liberating Slavic populations (Řeháček, 2011; Glassheim, 2016), implying limited differences in negative memories across the demarcation line within the Czech main lands. In *Sudetenland*, by contrast, many sources report that the liberating US Army forces treated Germans much less violently than Red Army forces did (Bundesministerium für Vertriebene, Flüchtlinge und Kriegsgeschädigte, 1957). Extensive Red Army violence towards Germans may have depressed the attraction of Communism among German stayers, and this could contribute to the voting and values pattern we uncover.[53] To shed more light on the issue, we employ the LITS micro-data previously used in Table 3.3. The survey includes questions about violence during World War Two. Table 3.22 in the Online Appendix does not show any significant differences in war violence memories across the demarcation line.[54]

This evidence is clearly limited by the small share of the German stayers in the population and the size of the LITS survey. We therefore additionally investi-

[53] A growing literature (e.g., Fontana et al., 2017) implies that traumatic war events can have long-term effects on political identity. Furthermore, Ochsner and Roesel (2017) shows that long-forgotten local historical events can be reactivated to affect voting preferences.

[54] Bombings during the war also do not vary across the line (Table 3.1).

gate collective memories. Liberation experiences may manifest in the presence of memorials, which are frequent all over Europe. We were able to collect data on local memorials commemorating World War Two, the liberating forces specifically, but also those related to the German history for the sub-sample of municipalities along the northern half of the *Sudetenland* demarcation line depicted in Figure 3.2. We employ the same strategy as for the LITS survey and compare mean differences within a 25 kilometer bandwidth on both sides of the demarcation line. Estimates listed in Table 3.23 in the Online Appendix show no statistically significant discontinuities in the presence of any of the memorial types we analyze.

Finally, the memories of the Allied forces could also have been shaped by Communist propaganda in the 1950s and 1960s, which downplayed the role of US troops in 1945 or demonize them.[55] It is not clear how such propaganda interacts with direct experiences of the liberating forces. Anecdotal evidence from the Czech main lands suggests the local population still remembers US forces fondly (see, for example, Mišterová, 2013).[56] However, if the memories of the US forces are fonder than those of the Red Army, or if anti-US propaganda back-fired in the former US zone, one may expect lower rather than higher Communist vote shares in US-liberated municipalities. Altogether, we find no evidence suggesting that different memories of the US or Red Army troops help explain our results.

3.6.6 Mobility

Selective mobility out of *Sudetenland* after the end of displacement may also be related to the (size of the) effects we uncover. We have already discussed the issue of selective re-settlement. For instance, if more fanatic Communists among the anti-fascist stayers move out of their ancestral homes, they may take their radical values to new places in Czechoslovakia (Ochsner and Roesel, 2020), which would imply our baseline estimates correspond to fewer stayers than we assume. Generally, the more mobility in and out of *Sudetenland*, the more dilution of political identity one might expect. However, our combined evidence on German names

[55]One famous example is the anti-US propaganda by Bartošek and Pichlík (1951). Some brochures and books show US soldiers aiming to shoot at Czech girls.

[56]For example, since 1990 the city of Plzeň (Pilsen), located just south of the demarcation line, celebrates an annual festival commemorating the liberation by the US Army.

among local Communist leaders in 1959 (Table 3.9 in the Online Appendix), on Communist party membership before 1989 (Table 3.3), and on the stable discontinuities across the demarcation line in the Communist vote share spanning almost two decades of Czech democratic elections (Table 3.14) suggest a continuous presence of German-ancestry Communist affiliation in the US-liberated regions, from post-war times to the Communist era, stretching to both the early 1990s after the Velvet Revolution and the present day. Finally, Table 3.24 in the Online Appendix corroborates the notion that mobility did not systematically vary across the former demarcation line. About 40% of *Sudetenland* residents as of 2001 are born in their residence municipality; the corresponding share is 10% for those born before 1945. Point estimates are positive, consistent with more stayers on the US side. However, RD estimates do not show any significant difference between US and Red Army liberated regions.[57] In sum, we do not find evidence that effects fade or that migration differed across the demarcation line.

3.7 Discussion

The evidence on mechanisms presented above implies that anti-fascist German stayers are the prime channel behind our baseline causal effects. To complete the interpretation of our main findings, we now discuss whether our local RD estimates speak to broader tendencies in post-war Czechoslovakia, and we ask whether our findings suggest that anti-fascist German stayers had a significant spillover effect in their newly re-settled local communities, beyond transmitting their values to their offspring.

3.7.1 Cross-sectional evidence

Our baseline estimates of the effect of staying anti-fascist Germans on present-day Communist vote shares are based on a well-defined identification strategy. However, as a consequence of the RD design we use, they correspond to local comparisons, which raises the question of whether they can be generalized. To provide a tentative insight into this issue, we present two pieces of descrip-

[57] The share of residents born prior to the war is about 2 percentage points higher on the US side in line with our historical evidence on German stayers.

tive cross-sectional evidence on the long-run relationship between the presence of staying German anti-fascists and election outcomes, one based on the entire Czech Republic, the other based on the entire *Sudetenland*.

We regress regional Communist party vote shares today on the corresponding population shares of staying anti-fascist Germans.[58] The most granular country-wide data on anti-fascist German stayers as of late 1946 covers 13 Czech regions. We also form estimates of staying anti-fascists for 67 *Sudetenland* counties; county-level data is not available for the Czech main lands.[59] This allows us to estimate cross-sectional least-squares specifications of the following form:

$$Communist_i = \alpha + \beta Antifascist_i + \gamma Industry_i + \epsilon_i, \tag{3.2}$$

where $Communist_i$ is the vote share of the Communist party (KSČM) in 2017 in region or county i. $Antifascist_i$ is the corresponding population share of anti-fascist Germans staying in Czechoslovakia, either directly measured or estimated. Finally, $Industry_i$ is the employment share in the industrial sector in 1930, which is related to Germans staying as specialized industrial workers. The coefficient β captures the cross-area association between the presence of anti-fascist German stayers and today's Communist vote shares, controlling for the pre-war industrial structure.

The estimates presented in Table 3.7 are in line with our baseline local causal estimates in that they confirm a positive relationship between staying anti-fascist Germans and Communist vote shares today. There are, of course, two major potential issues with the specifications corresponding to Equation 3.2. First, regional regressions are based on a small number of observations, and county-level regressions are affected by measurement error concerns. Thus, it is not surprising that the size of the estimates in column (3) differ from those in column (1). Sec-

[58] We use the same definition for anti-fascists as in Figure 3.3.

[59] Anti-fascist stayer population shares at the regional level are based on Luža (1964). The German stayer data sources at the county level do not distinguish between indispensable industrial workers and anti-fascists. We therefore proxy the anti-fascist county shares as residuals from a regression of the population share of all staying Germans from Urban (1964) on the employment share of industry in 1930; these residuals correspond to the part of the variation in staying Germans unexplained by industry, and hence should reflect the share of anti-fascist stayers.

Table 3.7: Anti-fascist Germans and Communist vote shares: Cross-sectional evidence

	Czech Republic (13 regions)	*Sudetenland* (67 counties)	
	Communist vote share 2017	Germans %	Communist vote share 2017
	(1)	(2)	(3)
Anti-fascist Germans %	0.540*		0.027**
	(0.271)		(0.011)
Industry %	-0.115**	0.509***	-0.076***
	(0.044)	(0.186)	(0.027)
Mean dep. var.	0.082	0.084	0.095
Obs	13	67	67
R^2	0.316	0.160	0.175

Notes: The table shows OLS regressions. In columns (1) and (3), the Communist vote share in the 2017 Czech elections serves as the dependent variable. Column (1) relies on the latest available (late 1946) regional data on staying anti-fascist Germans (certified anti-fascists or Germans subject to potential future transports and therefore likely anti-fascists; Luža, 1964) in % of 2017 population. The units of observations are the 13 regions as of 1950 covering the entire Czech Republic. Columns (2) and (3) use data on the number of staying Germans as of late 1946 from Urban (1964) for 67 *Sudetenland* counties. Since this data source does not separately show German anti-fascists as opposed to German indispensable industrial workers, we attempt to estimate the number of anti-fascists as the residual of the regression presented in Column (2), where we regress the share of staying Germans (in % of 2017 population) on the share of industry on county employment in 1930. In column (3), we use the residuals from the model in column (2) (i.e., variation in staying Germans unexplained by industry structure) as a proxy for anti-fascist Germans. Significance levels (robust standard errors): *** 0.01, ** 0.05, * 0.1.

ond, and more importantly, the presence of non-displaced anti-fascist Germans may be endogenous with respect to permanent differences in local Communist voting preferences—for example, strong Czech Communist elites might have been better able to protect their ethnic German party fellows. Notwithstanding these reservations, the magnitude of the nation-wide cross-sectional relationship in column (1) is significant as it implies that a 1% increase in the population share of anti-fascist German stayers after the war comes with a 0.5% increase in today's Communist vote share. The results in column (3) are qualitatively similar, but not quantitatively comparable due to the approximation procedure and the presence of measurement error.[60]

3.7.2 Multiplier effect

In the final step of our analysis we consider whether left-leaning German stayers had a significant multiplier (spillover) effect, as reflected in today's election outcomes, on their newly re-settled local communities. We thus ask whether the effects we estimate can be reasonably explained by the offspring of stayers (intergenerational transmission of values) alone, or whether they require spillovers of values into the non-stayer population.

Approaching this issue requires several simplifying assumptions. We assume no mobility differences and no differences in inter-ethnic marriages and in fertility of post-war anti-fascist stayers relative to their newly settled neighbours and their offspring. Table 3.8 provides two back-of-the-envelope calculations of such simplified multipliers based on our regression results; it relates counts of anti-fascists to counts of Communist votes. Column (1) relies on the cross-sectional nation-wide relationship presented in Table 3.7, where we find that a one percentage point increase share of anti-fascist German stayers relative to the 2017 population across 13 regions of the Czech Republic corresponds to a 0.5 percentage point increase in the 2017 Communist vote share (line (a) in Table 3.8). However, only around one of two residents of the Czech Republic turned out to vote in 2017. We assume uniform turnout rates (Table 3.14 shows no discontinuity in voter turnout across the demarcation line) and translate the population share-vote share coefficient

[60] The standardized beta coefficients for the population share of anti-fascist Germans are 0.6 in column (1) and 0.2 in column (3).

Table 3.8: Multiplier estimates

		All Czech lands	*Sudetenland* subsample
		(1)	(2)
(a)	Estimate from Table 3.7	0.540	
(b)	Valid votes in national election 2017	5,050,251	
(c)	Population 2017	10,578,820	
(d)	German population 1930		43,406
(e)	Discontinuity in anti-fascist Germans from Figure 3.3		0.028
(f)	"Excess" anti-fascist Germans 1946		1,215
(g)	Valid votes 2017		7,290
(h)	Discontinuity in Communist vote shares from Table 3.2		0.079
(i)	"Excess" Communist votes 2017		576
(j)	Multiplier *Communist votes 2017 per anti-fascist German 1946*	0.258	0.474

Notes: The table reports back-of-the-envelope calculations of the multiplier effect of anti-fascist Germans staying in Czechoslovakia after 1946 on Communist votes in the most recent 2017 Czech national election. Column (1) refers to the cross-sectional estimate from Table 3.7, column (1). The multiplier in line (j) equals (a) multiplied by (b) divided by (c). Column (2) combines information from Figure 3.3 and Table 3.2 and corresponds to an RD causal effect. The multiplier now equals (i) divided by (f), where (i) and (f) are in turn the products of rows (d) and (e), and (g) and (h), respectively.

from Table 3.7 into a stayer count-vote count multiplier by dividing the coefficient with the vote turnout rate. This gives a multiplier of about 0.3 (line (j)), which says that ten anti-fascist German stayers in 1946 come with approximately three Communist votes in the 2017 election. Given the total count of anti-fascist German stayers reported by Luža (1964), this would imply that some 6 to 7% of the 2017 Czech Communist votes had these specific German roots.[61]

Our second back-of-the-envelope calculation is based on our causal RD estimates; it confirms the magnitude of the tentative cross-sectional multiplier. In column (2), we refer to the sub-sample of municipalities along the northern half of the *Sudetenland* demarcation line, for which we observe the number of anti-fascist German stayers in local archives. A total of 43,406 Germans lived in these US-liberated municipalities as of 1930 (line (d)). Figure 3.3 focuses on these municipalities and shows a surplus of anti-fascist German stayers across the demarcation

[61] In total, 393,100 votes were cast for the Communist party in 2017; the number of post-war anti-fascist Germans is reported at 104,880 (anti-fascists, provisional citizenship/'special treatment' and Germans subject to potential future transfer). The numbers reported by Kučera (1992) are also close to 100,000.

line of 2.8% in terms of the 1930 population (line (e)); this implies 1,215 additional anti-fascist German stayers who were able to stay thanks to the presence of the US Army (f). Within this sample of municipalities, a total of 7,290 valid votes were cast in the 2017 Czech national election (g). We know from our RD estimates in Table 3.2 that Communist vote shares increase by about 8 percentage points of valid votes at the demarcation line (h). Thus, the US liberation is associated with 576 additional Communist votes (i). When we relate the absolute number of 'excess' anti-fascist Germans to 'excess' Communist votes, we obtain a multiplier of 0.47 (j), which implies that ten staying anti-fascist Germans in 1946 account for four to five Communist votes in 2017.

These are sizeable effects, but they do not necessitate that German stayers were able to spread their values among their new neighbours, as these effects are consistent with the post-war political value structure of the population being preserved through the generations until today. This could be achieved by full transmission of values within the families of stayers or by a combination of imperfect within-family transmission and oblique society-wide transmission. Given that at least three generations bridge the seven decades between treatment and effect, including five decades of the Communist regime and two decades of transition to democracy, we find the preservation of these far-left values strongly supportive of the notion that extremism has historical origins that begun with a 'small seed' of political development (Giuliano and Tabellini, 2020).

3.8 Conclusion

We provide the first causal evidence on the long-term impact of stayers exempted from ethnic cleansing. Three million *Sudeten* Germans were expelled from the Czech borderlands after World War Two. However, some 200,000 Germans were allowed to stay, many because they were liberated by the US Army and not by the Red Army. We study the legacy of anti-fascist Germans in post-war Czechoslovakia using quasi-experimental variation and find a substantial and lasting political-value footprint of this left-leaning minority in today's Czech Republic. Communist vote shares, active Communist party cells, far-left values, and social policies are more pronounced in *Sudetenland* today where more anti-fascist

Germans stayed after the war. Our evidence on how far-left political values take hold in re-settled communities extends the literature documenting long-lasting Communist preferences (see Fuchs-Schündeln and Schündeln, 2020, for a survey).

The finding that stayers who evade expulsion can have long-lasting effects on political values and voting behavior in re-settled populations complements the literature showing that immigrant's political values act similarly upon established societies thanks to cultural transmission (e.g., Dippel and Heblich, 2021; Ochsner and Roesel, 2020; Giuliano and Tabellini, 2020). Our evidence implies that ethnic cleansing does not prevent expression of identity by a small minority of stayers. Even Germans in a Slavic country following World War Two's atrocities have been able to affect political landscapes in newly formed societies. Ethnic Germans appear to have already been well represented among local Communist elites in the 1950s, i.e., shortly after the war, and this may be linked to the local roots of staying German anti-fascists (as compared to the re-settling Slavic majority). The effects we measure go well beyond the Communist regime, where state ideology was aligned with the far-left values of anti-fascists. They imply strong persistence of far-left values among stayers based on within-family inter-generational transmission. Overall, our findings provide new support for the 'small seed' mechanism of political development, which in our case corresponds to staying minorities integrating with newly arriving majorities.

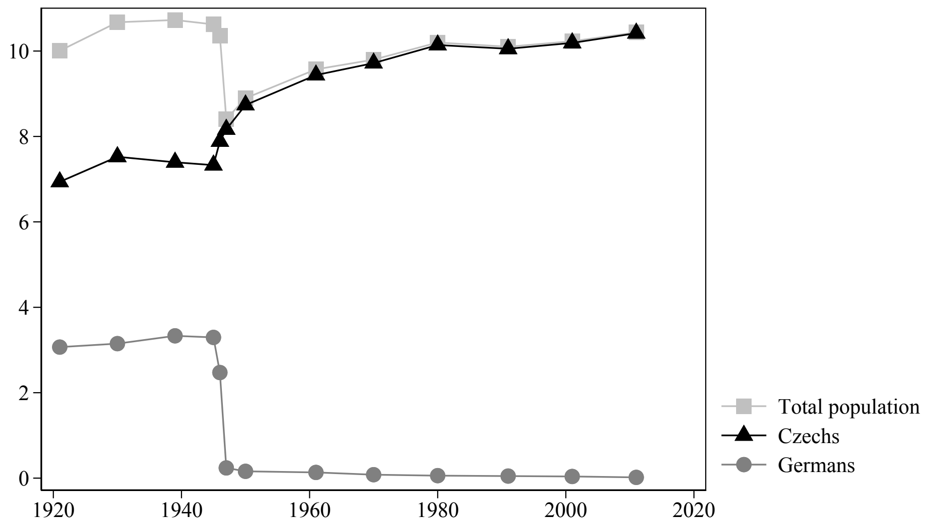
More broadly, our results shed new light on the inter-generational transmission of multi-dimensional identity. Evidence that ethnic-identity choices respond to incentives is well-established (e.g., Algan et al., 2020; Fouka, 2019; Atkin et al., 2020). In our case, German stayers had two identities: an ethnic and a political one. We find more active Communists today with German family roots where more anti-fascists avoided displacement, but we find a similar extent of ethnic assimilation. Among anti-fascist Germans, political identity may have supplanted their suppressed ethnic identity, and persisted when ethnic roots were no longer salient. Fading German identity of anti-fascists is in line with theoretical models which predict well-connected representatives of a minority will assimilate faster (Verdier and Zenou, 2017). Future research can investigate how integration policies affect the joint identity choice across ethnic, religious, and political dimensions, both within re-settled societies after ethnic cleansing and in established host societies facing immigration.[62]

[62]For example, Abdelgadir and Fouka (2020) explore the effect of suppression of immigrant religious expression on both their nationality and religious identity.

3.9 Supplementary figures and tables

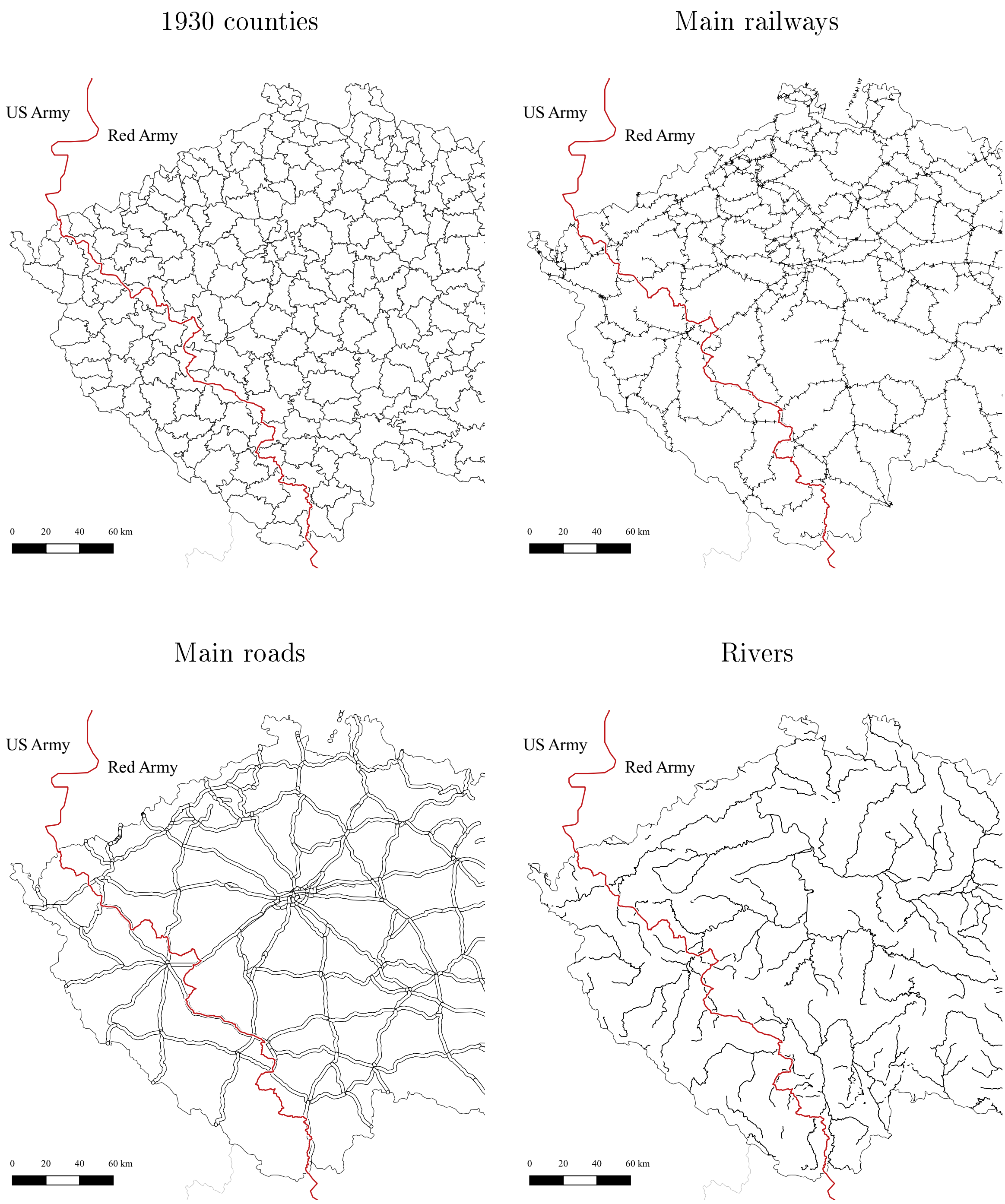
This Online Appendix provides supplementary material and is for online publication only.

Figure 3.4: Population in the Czech lands (in millions)



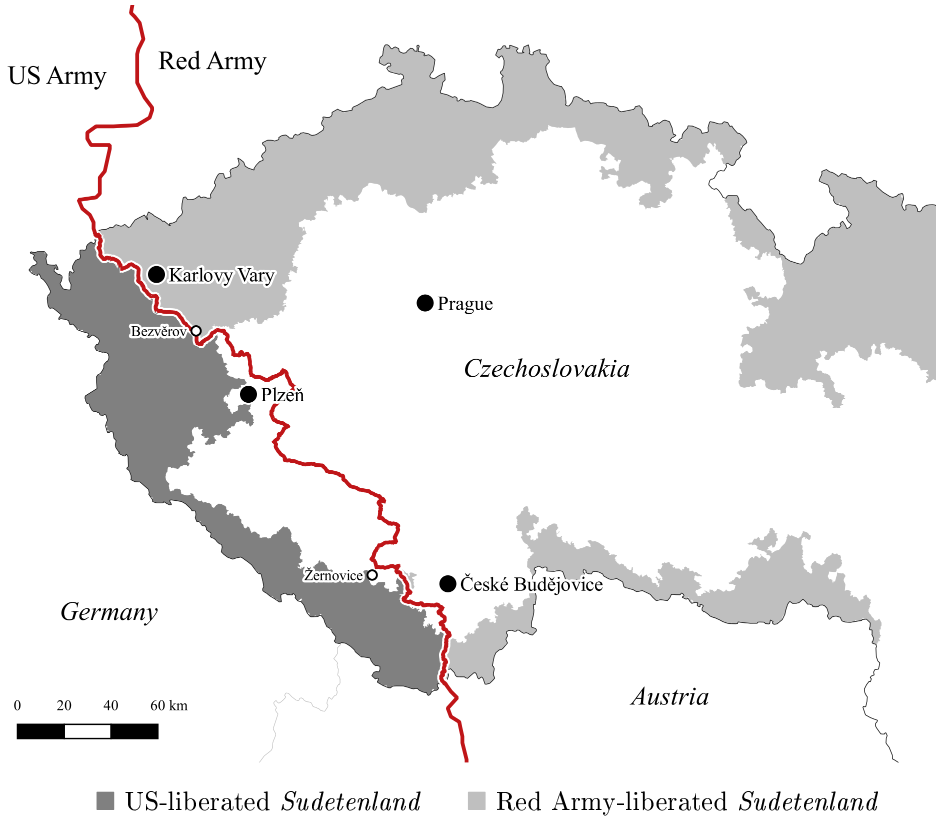
Notes: The figure shows total population of the Czech Republic (Czech lands consisting of Bohemia, Moravia and Silesia) between 1921 and 2011 (light gray), and population by self-declared ethnicity (black and dark gray). The German population (dark gray bullets) was almost entirely expelled in 1945 and 1946 and partly replaced by residents mainly from Czech hinterlands and Slovakia. 'Czechs' refers to all other non-German residents (black triangles).

Figure 3.5: Demarcation line and pre-existing infrastructure



Notes: The maps compare the demarcation line between US and Red Army forces in 1945 Czechoslovakia (red line) to county boundaries as of 1930, main roads, main railways, and rivers.

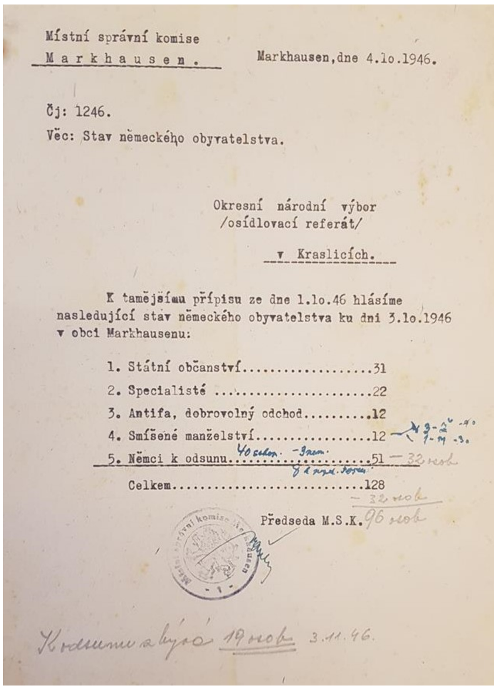
Figure 3.6: Demarcation line between US and Red Army forces in 1945 Czechoslovakia



Notes: The map zooms into Figure 3.1 in the main text. The red line represents the demarcation line between US and Red Army forces in 1945 Czechoslovakia, which runs from Karlovy Vary over Plzeň to České Budějovice (black dots). Prague is the capital city. The US-liberated regions of *Sudetenland* are in dark gray, the Red Army-liberated regions are in light gray. *Sudetenland* was settled by ethnic Germans and annexed by Nazi Germany in October 1938. The white-shaded area (within the Czechoslovak black boundaries) are the Czech main lands. We exclude from all analyses the regions south of Žernovice (white dot), where the demarcation coincided with (pre-displacement) ethnic divisions.

Figure 3.7: Registration/deportation lists (samples)

Markhausen (Hraničná)
(municipality)



Místní správní komise
Markhausen.

Markhausen, dne 4.1o.1946.

Čj: 1246.
Věc: Stav německého obyvatelstva.

Okresní národní výbor
/osídlovací referát/

v Kraslicích.

K tamějšímu přípisu ze dne 1.1o.46 hlásíme nasledující stav německého obyvatelstva ku dni 3.1o.1946 v obci Markhausenu:

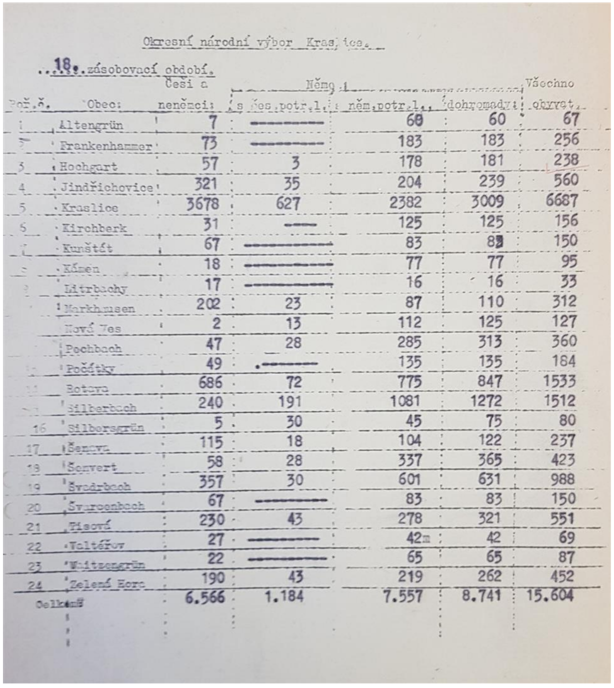
1. Státní občanství....................31
2. Specialisté 22
3. Antifa, dobrovolný odchod..........12
4. Smíšené manželství.................12
5. Němci k odsunu.....................51 − 32 osob

Celkem..............................128
− 32 osob
96 osob

Předseda M.S.K.

19 osob 3.11.46.

Graslitz (Kraslice)
(county)



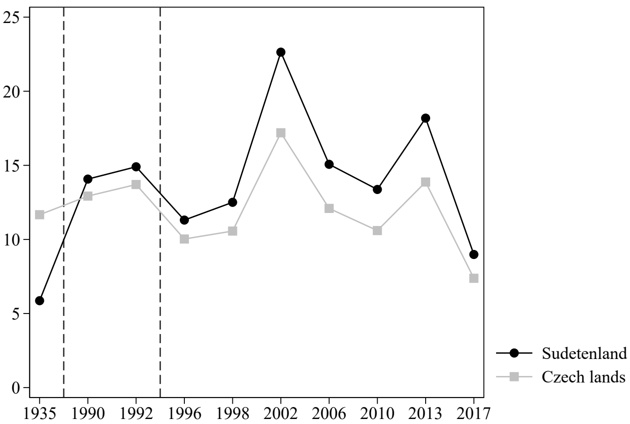
Okresní národní výbor Kraslice.

18. zásobovací období.

Poř. č.	Obec:	Češi a neněmci:	Němci: s čes. potr. l.	Němci: něm. potr. l.	Němci: dohromady:	Všechno obyvat.
1	Altengrün	7	———	60	60	67
	Frankenhammer	73	———	183	183	256
3	Hochgart	57	3	178	181	238
4	Jindřichovice	321	35	204	239	560
5	Kraslice	3678	627	2382	3009	6687
6	Kirchberk	31	———	125	125	156
	Kunštát	67	———	83	83	150
	Kámen	18	———	77	77	95
	Litrbachy	17	———	16	16	33
	Markhausen	202	23	87	110	312
	Nová Ves	2	13	112	125	127
	Pechbach	47	28	285	313	360
	Počátky	49	———	135	135	184
	Rotava	686	72	775	847	1533
	Silberbach	240	191	1081	1272	1512
16	Silbersgrün	5	30	45	75	80
17	Šenava	115	18	104	122	237
18	Šenvert	58	28	337	365	423
19	Švadrbach	357	30	601	631	988
20	Švarcenbach	67	———	83	83	150
21	Tisová	230	43	278	321	551
22	Valtéřov	27	———	42	42	69
23	Waitzengrün	22	———	65	65	87
24	Zelená Hora	190	43	219	262	452
	Celkem:	6.566	1.184	7.557	8.741	15.604

Notes: The documents show two samples of registration/deportation lists collected from local archives in *Sudetenland* and used in Figure 3.3. The lists refer to the municipality of Markhausen (Hraničná) and the county of Graslitz (Kraslice) and were compiled by the local national committee of Kraslice county (*Okresní národní výbor Kraslice, ONV Kraslice*). Documents are printed with the permission of *Státní okresní archiv Sokolov.*

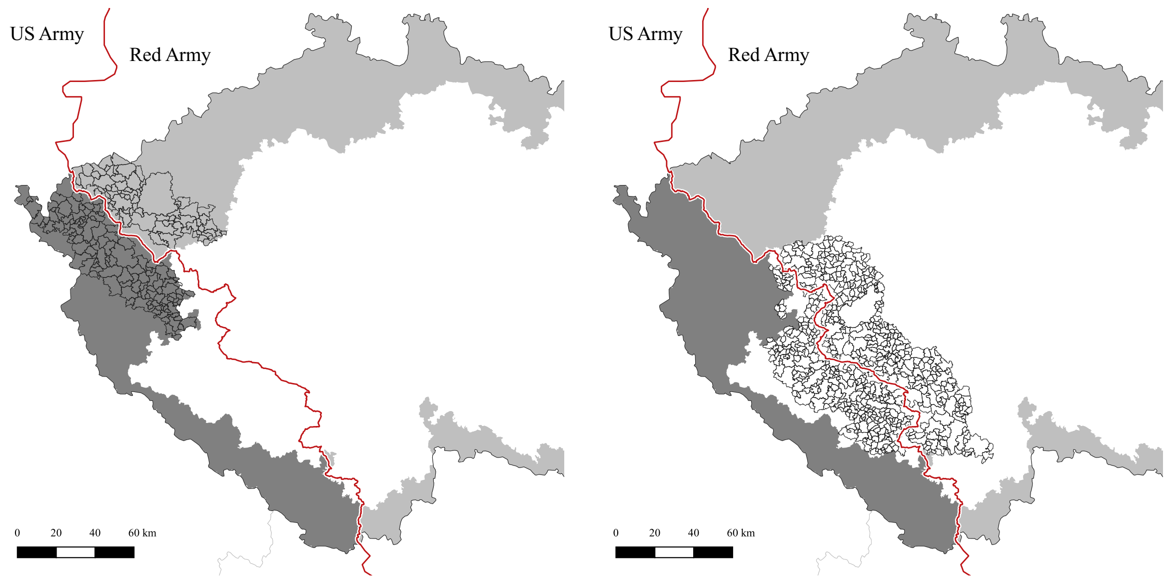
Figure 3.8: Communist vote share (in % of valid votes)



Notes: The figure shows vote shares of the Czech communist party (KSČ/KSČM) in national parliamentary elections in 1935 and from 1990 to 2017. Black lines with squares show vote shares in the formerly German-settlement areas of *Sudetenland*, gray lines with bullets refer to vote shares in the Czech main lands. Vertical dashed lines separate Czechoslovakia before expulsions (1935), democratic Czechoslovakia after the Velvet Revolution (1990, 1992), and modern Czech Republic (1994 to 2017), which is the main focus of our analysis. We omit the 1946 national election when Germans were not eligible to vote.

Figure 3.9: Sample municipalities in *Sudetenland* and in the Czech main lands

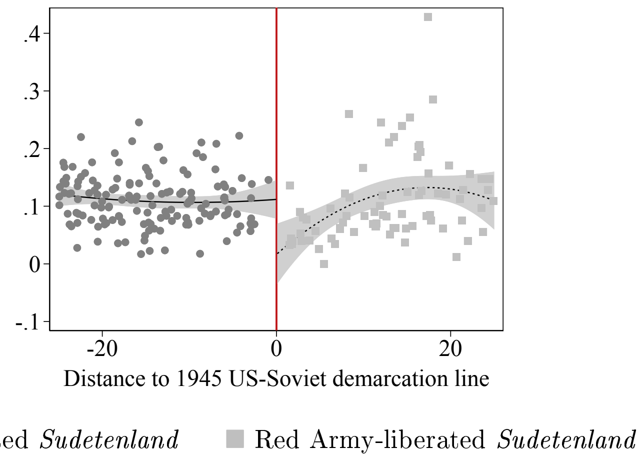
Sudetenland Czech main lands



■ US-liberated *Sudetenland* ■ Red Army-liberated *Sudetenland*

Notes: The maps show the two samples of municipalities we use in this study. The red line is the demarcation line between US and Red Army forces in 1945 Czechoslovakia. Black lines within Czechoslovakia are municipality boundaries for municipalities included in a sample. The left-hand map refers to the sample of the German-populated *Sudetenland* municipalities, the right-hand map shows the Czech main lands. We exclude municipalities more than ±25 km from the demarcation line, municipalities divided by the boundaries of *Sudetenland* or the demarcation line, and municipalities located south of Žernovice, where the demarcation line coincided with ethnic divides.

Figure 3.10: Communist party vote shares 2017 (RD plots)



Notes: The graph plots Communist party vote shares in the 2017 national election in municipalities against the distance to the demarcation line. We use only municipalities withing a maximum distance of 25 km to the 1945 demarcation line. Dots in dark gray represent US-liberated municipalities, dots in light gray are municipalities liberated by the Red Army.

Table 3.9: Names of county-level Communist party leaders (1959)

	Names of local Communist leaders		
	Total	German	% German
	(1)	(2)	(3)
US zone 1945	242	35	14.5%
Soviet zone 1945	240	29	12.1%

Notes: The table presents the share of German surnames among the 546 local Communist party leaders in the year 1959 in eight Czech counties around the demarcation line (the 1950 counties of Aš, Cheb, Kraslice, Mariánské Lázně, and Sokolov sum up to the US Army region, the Red Army region is the sum of the counties of Kadaň, Karlovy Vary, Ostrov, Podbořany, and Toužim). Names are hand collected from local archives.

Table 3.10: Population and houses (relative to 1930)

	Sudetenland		Czech main lands	
	Population	Houses	Population	Houses
	(1)	(2)	(3)	(4)
1900	0.103	0.106	-0.042	-0.090
	(0.113)	(0.078)	(0.067)	(0.062)
1910	0.013	0.048	-0.044	-0.070
	(0.079)	(0.060)	(0.059)	(0.053)
1921	0.059	0.051	-0.026	-0.060
	(0.051)	(0.055)	(0.041)	(0.044)
1930	–	–	–	–
	–	–	–	–
1950	0.210**	0.013	-0.029	-0.020
	(0.092)	(0.122)	(0.042)	(0.044)
1961	0.122	0.091	0.036	-0.036
	(0.133)	(0.158)	(0.065)	(0.076)
1970	0.046	0.095	0.004	-0.011
	(0.194)	(0.147)	(0.079)	(0.100)
1980	0.045	0.015	-0.008	-0.059
	(0.203)	(0.147)	(0.097)	(0.127)
1991	0.011	-0.027	-0.022	-0.121
	(0.187)	(0.148)	(0.109)	(0.158)
2001	0.006	-0.020	0.003	-0.144
	(0.182)	(0.139)	(0.125)	(0.173)
2011	-0.051	-0.180	-0.042	-0.118
	(0.164)	(0.195)	(0.142)	(0.199)

Notes: The table shows RD estimates at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. We use a local-linear RD procedure including a data-driven optimal bandwidth choice (Calonico et al., 2017). The units of observation are municipalities, the dependent variables are population and houses relative to 1930. Columns (1) and (2) show estimates for regions historically settled by ethnic Germans (*Sudetenland*), columns (3) and (4) refer to the Czech main lands. We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.11: Pre-war Communist vote shares

	Sudetenland	Czech main lands	Difference
	(1)	(2)	(3)
US zone 1945	0.033	0.042	-0.009
Soviet zone 1945	0.050	0.044	0.007
Difference	-0.017	-0.001	-0.015

Notes: The table shows Communist (KSČ) vote shares in the 1935 Czechoslovak national elections at the 1945 demarcation line between US and Red Army forces. The units of observation are counties. We impose a 25 km bandwidth around the demarcation line. Column (1) shows estimates for regions originally settled by ethnic Germans (*Sudetenland*), column (2) refers to the Czech main lands. Rows refer to US- and Red Army-liberated regions. Column (3) and the third row show mean differences. Significance levels: *** 0.01, ** 0.05, * 0.1 (none to report).

Table 3.12: *Sudetenland* sub-samples

	Communist vote share 2017	
	Sudetenland	
	North	South
	(1)	(2)
US zone 1945	0.059**	0.122
	(0.025)	(0.133)
Geography controls	No	No
Population controls	No	No
Mean dep. var.	0.109	0.108
RD bandwidth	18.649	19.708
Eff. obs.	91	76

Notes: The table shows estimates for two regional sub-samples of *Sudetenland* corresponding to the baseline local-linear RD specification in column (3) of Table 3.2. We split *Sudetenland* municipalities into a north and a south sub-sample relative to the village of Bezvěrov. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.13: Pseudo treatments

	Communist vote share 2017				
	Sudetenland			Czech main lands	
	Pseudo dem. line +25km	Pseudo dem. line −25km	Pseudo dem. line Ohře river	Pseudo dem. line +25km	Pseudo dem. line −25km
	(1)	(2)	(3)	(4)	(5)
US zone 1945	0.013	0.014	-0.004	-0.018	0.005
	(0.029)	(0.016)	(0.011)	(0.015)	(0.021)
Geography controls	No	No	No	No	No
Population controls	No	No	No	No	No
Mean dep. var.	0.114	0.111	0.104	0.107	0.101
RD bandwidth	24.159	30.143	31.208	14.280	10.997
Eff. obs.	132	218	421	362	132

Notes: The table shows various pseudo-treatment analyses, building on our baseline RD specification (see, Table 3.2, columns (3) and (6)). Columns (1) to (3) shows estimates for regions historically settled by ethnic Germans (*Sudetenland*), columns (4) to (5) refer to the Czech main lands. We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. In columns (1), (2), (4) and (5), we shift the demarcation line 25 km to the East and to the West. In column (3), we us a pseudo demarcation line running from East to West along the Ohře river, which cuts through *Sudetenland*. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.14: Vote shares in national elections, 1996–2017

	Vote shares									
	Sudetenland					Czech main lands				
	Communist party	Far-left parties	Centrist parties	Far-right parties	Voter turnout	Communist party	Far-left parties	Centrist parties	Far-right parties	Voter turnout
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1996	0.051*	0.070**	-0.128***	0.065***	0.039	0.019	0.023	0.007	-0.026	0.004
	(0.029)	(0.036)	(0.046)	(0.025)	(0.060)	(0.015)	(0.020)	(0.027)	(0.023)	(0.021)
1998	0.042	0.070**	-0.106***	0.039**	0.004	0.018	0.006	-0.010	0.010	-0.006
	(0.025)	(0.028)	(0.033)	(0.020)	(0.038)	(0.017)	(0.020)	(0.020)	(0.012)	(0.017)
2002	0.120***	0.131***	-0.133**	0.003	-0.005	0.007	0.004	0.016	-0.013	-0.001
	(0.042)	(0.048)	(0.055)	(0.015)	(0.050)	(0.028)	(0.029)	(0.033)	(0.019)	(0.021)
2006	0.076**	0.076**	-0.079***	0.003	-0.037	0.013	0.013	-0.017	0.005	0.007
	(0.030)	(0.030)	(0.031)	(0.009)	(0.063)	(0.020)	(0.020)	(0.021)	(0.011)	(0.028)
2010	0.087**	0.083**	-0.108**	0.047*	-0.059	0.009	0.012	-0.028	0.001	-0.002
	(0.034)	(0.034)	(0.051)	(0.024)	(0.064)	(0.022)	(0.022)	(0.026)	(0.011)	(0.019)
2013	0.110***	0.109***	-0.123***	0.010	-0.086	0.006	0.010	-0.019	0.010	-0.005
	(0.040)	(0.039)	(0.046)	(0.023)	(0.062)	(0.021)	(0.021)	(0.023)	(0.013)	(0.022)
2017	0.079***	0.080***	-0.096**	0.019	-0.096**	0.004	0.004	0.013	-0.010	-0.001
	(0.026)	(0.026)	(0.043)	(0.028)	(0.047)	(0.017)	(0.016)	(0.026)	(0.021)	(0.024)

Notes: The table shows the effect for US-liberated regions (RD estimates) at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. We use a local-linear RD procedure with a data-driven optimal bandwidth choice (Calonico et al., 2017). The unit of observation are municipalities, the dependent variables are vote shares for the Communist party (KSČM), ideological camps and voter turnout in all democratic elections in Czech Republic since 1996. Columns (1) to (5) show estimates for regions historically settled by ethnic Germans (*Sudetenland*), columns (6) to (10) refer to the Czech main lands. We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust RD standard errors/standard errors clustered at county level): *** 0.01, ** 0.05, * 0.1.

Table 3.15: Social Democrats (ČSSD)

	Social democratic vote share 2017	
	Sudetenland	Czech main lands
	(1)	(2)
US zone 1945	-0.015	0.008
	(0.023)	(0.014)
Geography controls	No	No
Population controls	No	No
Mean dep. var.	0.068	0.079
RD bandwidth	20.009	12.527
Eff. obs.	145	302

Notes: The table replicates our baseline RD specifications (Table 3.2, columns (3) and (6)) for the vote shares of the Social Democratic party (ČSSD) in the 2017 Czech national elections. The units of observation are municipalities. Column (1) shows estimates for regions originally settled by ethnic Germans (*Sudetenland*), column (2) refers to the Czech main lands. We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.16: Communist party cells

	Communist party cell (yes = 1)	
	Sudetenland	Czech main lands
	(1)	(2)
US zone 1945	0.121***	0.031
	(0.044)	(0.040)
Geography controls	No	No
Population controls	No	No
Mean dep. var.	0.343	0.142
RD bandwidth	16.632	8.149
Eff. obs.	805	1,428

Notes: The table shows the effects for US-liberated regions (RD estimates) at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. We use a local-linear RD procedure with a data-driven optimal bandwidth choice (Calonico et al., 2017). The unit of observation are municipalities, the dependent variables is an indicator for the presence of a local Communist party cell standing in local (municipal) elections. We pool all local elections in modern Czech Republic (between 1994 and 2018). Column (1) shows estimates for regions historically settled by ethnic Germans (*Sudetenland*), column (2) refers to the Czech main lands. We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust RD standard errors/standard errors clustered at county level): *** 0.01, ** 0.05, * 0.1.

Table 3.17: Trust (LITS micro data)

	Sudetenland	Czech main lands
	(1)	(2)
Trust		
General	-0.374	-0.293
	(0.679)	(0.256)
Government	-1.133*	-0.326
	(0.652)	(0.226)
Local government	0.210	-0.315
	(0.665)	(0.223)
Parties	-0.549	-0.146
	(0.667)	(0.220)
Neighbors	0.710	0.434*
	(0.608)	(0.247)
New contacts	0.087	0.297
	(0.726)	(0.210)
Foreigners	-1.218*	0.533***
	(0.736)	(0.199)
Geography controls	Yes	Yes
Sociodemographic controls	Yes	Yes
Year fixed effects	Yes	Yes
Bandwidth	25.000	25.000
Max. obs.	126	197

Notes: The table shows coefficients for US-liberated regions from ordered probit specifications at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. The units of observation are individual respondents in the Life in Transition Survey, the dependent variables are answers to survey questions. We pool survey II (2010) and III (2016) and include year fixed effects. Geography controls are longitude and latitude of the respondent. Sociodemographic controls are age and gender. Column (1) shows estimates for regions originally settled by ethnic Germans (*Sudetenland*), column (2) refers to the Czech main lands. We impose a 25 km bandwidth around the demarcation line. We exclude residents from municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.18: Ideology of settlers from Czech main lands (1947 census)

		Origin of settlers									
		Czech main lands	*Thereof* (County and corresponding Communist vote share 1935)							Communist voters 1935	Communist vote share 1935
			Benešov 9%	Beroun 21%	Blatná 1%	... ...	Zábřeh 8%	Zlín 5%	Znojmo 3%	(Predicted)	(Predicted)
		(1)	(2)	(3)	(4)	(5)–(116)	(117)	(118)	(119)	(120)	(121)
Karlovy Vary	Red Army zone	33443	390	660	371	...	11	134	33	3822	11%
Aš	US zone	7520	21	84	115	...	3	21	12	665	9%
Cheb	US zone	19592	128	387	414	...	9	36	0	1756	9%
Jáchymov	Red Army zone	11053	94	193	233	...	0	140	27	1341	12%
Kadaň	Red Army zone	14112	166	743	217	...	0	31	38	2007	14%
Kraslice	Divided	5122	28	67	66	...	0	10	0	535	10%
Mariánské Lázně	US zone	18938	645	131	387	...	0	29	2	1824	10%
Podbořany	Red Army zone	11612	28	288	256	...	0	28	3	1333	11%
Sokolov	US zone	15197	101	314	179	...	2	45	6	1630	11%
Toužim	Divided	11131	54	40	281	...	0	9	14	1146	10%
Karlovy Vary region		146709	1655	2907	2519	...	25	483	135	15951	11%
Sudetenland		1119952	6050	8319	6459	...	1782	11617	4319	117836	11%

Notes: The table reports the total number of settlers from the Czech main lands to *Sudetenland* in column (1). Columns (2) to columns (119) break down the total number of settlers by their county of origin; the header also reports the Communist vote share in the county in the 1935 Czechoslovak national election (we omit columns (5) to (116) due to space limits). Column (120) reports the predicted number of Communist voters, derived as the sum of settlers times Communist vote share 1935. Column (121) is the number of predicted Communist voters divided by the total number of settlers. Karlovy Vary region as of 1950 includes 10 counties and somewhat correspond with the map in Figure 3.2.

Table 3.19: Czechoslovak national election in May 1946

	Communist vote share 1946					
	Sudetenland			Czech main lands		
	Para-metr. RD	Para-metr. RD	Local-lin. RD	Para-metr. RD	Para-metr. RD	Local-lin. RD
	(1)	(2)	(3)	(4)	(5)	(6)
US zone 1945	0.056	0.090	-0.024	0.027	0.017	-0.111
	(0.110)	(0.102)	(0.230)	(0.052)	(0.047)	(0.085)
Geography controls	No	Yes	No	Yes	Yes	No
Population controls	No	Yes	No	Yes	Yes	No
Mean dep. var.	0.611	0.611	0.655	0.451	0.451	0.466
RD bandwidth	25.000	25.000	5.564	25.000	25.000	4.658
Eff. obs.	183	183	23	555	555	112
R^2	0.949	0.954	–	0.911	0.917	–

Notes: The table shows the effect for US-liberated regions (RD estimates) at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia based on a parametric (quadratic-interacted) polynomial approach without/with control variables (columns (1), (2), (4), and (5), bandwidth: 25 km) and a local-linear RD specification including a data-driven optimal bandwidth choice (Calonico et al., 2017). The employed data correspond to municipalities within a 25 km bandwidth on both sides of the demarcation line. The units of observation are municipalities, the dependent variable is the vote share of the Communist party (KSČ) in the Czechoslovak national elections in May 1946 (Germans were not eligible to vote and resettlement not yet finished). Columns (1) to (3) show estimates for regions originally settled by ethnic Germans (*Sudetenland*), columns (4) to (6) refer to the Czech main lands. We exclude municipalities south of Žernovice, where ethnicity divides corresponded with the demarcation line. Geography controls are the distance to the external (German) border, distance to the nearest main road, distance to the nearest railway line, mean altitude and slope (difference between maximum and minimum altitude). Population controls are logged population in 1930. Significance levels (Conley (2010) standard errors/robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.20: Sectoral employment shares, 1950

	Sectoral share	
	Sudetenland	Czech main lands
	(1)	(2)
Agriculture	0.486	0.050
	(0.619)	(0.127)
Industry	-0.504	-0.014
	(0.702)	(0.102)
Crafting	0.024	0.023
	(0.023)	(0.017)
Other sectors	0.046	-0.040
	(0.127)	(0.047)

Notes: The table shows the effects for US-liberated regions (RD estimates) at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. We use a local-linear RD procedure with a data-driven optimal bandwidth choice (Calonico et al., 2017). The employed data correspond to municipalities within a 25 km bandwidth on both sides of the demarcation line. The units of observation are municipalities, the dependent variables are sectoral employment shares as of the 1950 census. Column (1) shows estimates for regions historically settled by ethnic Germans (*Sudetenland*), column (2) refers to Czech main lands. We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.21: Sectoral employment shares, 2001

	Sectoral share	
	Sudetenland	Czech main lands
	(1)	(2)
Agriculture	0.145**	-0.009
	(0.065)	(0.066)
Industry	-0.130	0.003
	(0.086)	(0.041)
Retail	0.022	-0.023
	(0.030)	(0.021)
Transport	-0.034	0.001
	(0.030)	(0.012)
Public sector, health, education	0.011	0.043
	(0.030)	(0.026)

Notes: The table shows the effects for US-liberated regions (RD estimates) at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. We use a local-linear RD procedure with a data-driven optimal bandwidth choice (Calonico et al., 2017). The units of observation are municipalities, the dependent variables are sectoral employment shares as of the 2001 census. Column (1) shows estimates for regions historically settled by ethnic Germans (*Sudetenland*), column (2) refers to Czech main lands. We exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.22: War injuries and displacement (LITS micro data)

	Sudetenland	Czech main lands
	(1)	(2)
Were you, your parents or any of your grandparents ...		
... physically injured or killed during WWII?	-0.027	-0.081
	(0.153)	(0.062)
... forced to move as a result of WWII?	-0.025	-0.090
	(0.231)	(0.064)
Geography controls	Yes	Yes
Sociodemographic controls	Yes	Yes
Year fixed effects	Yes	Yes
Bandwidth	25.000	25.000
Max. obs.	115	194

Notes: The table shows the marginal effects for US-liberated regions from probit specifications at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. The units of observation are individual respondents in the Life in Transition Survey, the dependent variables are answers to survey questions. We pool survey waves II (2010) and III (2016) and include year fixed effects. Geography controls are longitude and latitude of the respondent. Socio-demographic controls are age and gender. We impose a 25 km bandwidth around the demarcation line. Column (1) shows estimates for regions originally settled by ethnic Germans (*Sudetenland*), column (2) refers to the Czech main lands. We exclude residents from municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust standard errors): *** 0.01, ** 0.05, * 0.1.

Table 3.23: Monuments and memorials

	Number of monuments					Share of US army monuments	
	Total	WWII	US Army	Red Army	German	Total	WWII
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
US zone 1945	-6.070	-1.062	0.039	-0.231	-2.630	0.011	0.062
	(4.912)	(0.871)	(0.059)	(0.260)	(1.750)	(0.008)	(0.038)
Geography controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dep. var.	5.562	0.918	0.082	0.192	2.315	0.008	0.044
Bandwidth	25.000	25.000	25.000	25.000	25.000	25.000	25.000
Obs.	73	73	73	73	73	73	31
Adj. R^2	0.414	0.398	0.480	0.341	0.428	0.159	0.471

Notes: The table shows OLS estimates comparing US- and Red Army-liberated regions in 1945 Czechoslovakia. The units of observation are municipalities, the dependent variable is the number of local monuments and memorials corresponding to a given type of events, including World War Two (WWII), liberating forces, and German history. Geography controls are the distance to the external border, distance to the next main road, distance to the next railway line, mean altitude and slope (difference between maximum and minimum altitude). Population controls are logged population in 1930 and logged present-day population. We use a sub-sample of *Sudetenland* municipalities along the norther half of the *Sudetenland* demarcation line withing a maximum distance of 25 km around the demarcation line. Significance levels (standard errors clustered at municipalities): *** 0.01, ** 0.05, * 0.1.

Table 3.24: Mobility in *Sudetenland*

	% Local born residents	
	All cohorts	Born before 1945
	(1)	(2)
US zone 1945	0.072	0.019
	(0.050)	(0.056)
Geography controls	No	No
Population controls	No	No
Mean dep. var.	0.413	0.103
RD bandwidth	12.208	17.174
Eff. obs.	72	119

Notes: The table shows the effects for US-liberated regions (RD estimates) at the demarcation line between US- and Red Army-liberated regions in 1945 Czechoslovakia. We use a local-linear RD procedure with a data-driven optimal bandwidth choice (Calonico et al., 2017). The units of observation are municipalities, the dependent variable is the share of residents born in the municipality. Data come from the 2001 census. We use regions originally settled by ethnic Germans (*Sudetenland*) and exclude municipalities south of Žernovice, where ethnic divides corresponded with the demarcation line. Significance levels (robust RD standard errors): *** 0.01, ** 0.05, * 0.1.

3.10 Data description and sources

This Online Appendix describes our data sources and is for online publication only.

3.10.1 Election data

National elections 1990, 1992, 1996, 1998, 2002, 2006, 2010, 2013, 2017: We retrieved data at the municipality level from the election website of the Czech Statistical Office (`https://www.volby.cz`). We focus on KSČ/KSČM, the Communist party, and ČSSD, the Social Democratic party. We code as far-left the following set of parties: KSČM, ČSNS, LEV 21, Občané 2011, RDS, STOP, SŽJ, SDS, SDL, Levý blok, HSS, Volební seskupení zájmových svazů v ČR, and Československé demokratické fórum. We code as far-right the following parties: BPS, CESTA, ČHNJ, Česká národní fronta, ČP, DSSS, Politika 21, KONS, Koruna Česká, Moravané, Národní strana, Národ Sobě, ND, NEZ/DEM, Volte Pravý Blok www.cibulka.net, ŘN-VU, REAL, Rozumní, Blok proti islamizaci – Obrana domova, SPD, SPR-RSČ, Svobodní, Unie H.A.V.E.L. 17, Úsvit, Národní demokratická strana, Volba pro budoucnost, Nové hnutí, Strana venkova - spojené občanské síly, Republikáni, MoDS, ČMUS, HSMS, HSD-SMS/HSDMS, Strana republikánské a národně demokratické jednoty. The remainder are considered centrist parties.

National elections 1935, 1946: Data at the municipality level (1946) are hand-collected from the following source: *Zprávy Státního Úřadu Statistického Republiky Československé, 27 (1946), Řada B, Číslo 24-25, 26-28, 29-30, 31-33, Prague.* We transform the data to the present territorial status of municipalities. Data at the county level (1935) are hand collected from *Český statistický úřad (2008): Výsledky hlasování podle okresů v letech 1920 – 1946, Prague.*

Local (municipality) elections 1994, 1998, 2002, 2006, 2010, 2014, 2018: We retrieved the data at the municipality level (including candidate names) from the election website of the Czech Statistical Office (`https://www.volby.cz`).

3.10.2 Population data

Total population 1900, 1910, 1921, 1930, 1950, 1961, 1970, 1980, 1991, 2001, 2011: Data at the municipality level are from *Český statistický úřad (2015): Historický lexikon obcí České republiky - 1869 - 2011, Počet obyvatel a domů podle krajů, okresů, obcí, částí obcí a historických osad/lokalit v letech 1869 - 2011, Česká republika, Prague.*

Total population 2017: Data at the municipality level are from the Small Lexicon of Municipalities of the Czech Republic 2017, published by the Czech Statistical Office.

Population by ethnicity 1920, 1930, 1950, 1961, 1970, 1980, 1991, 2001, 2011: Data for the Czech lands are from the Historická data v GIS projecty (Zpřístupnění historických prostorových a statistických dat v prostředí GIS, `http://www.historickygis.cz`) by the Urbánní a regionální laboratoř, available at (`http://web.natur.cuni.cz/ksgrrsek/urrlab_vystupy/download`).

Population by ethnicity, denomination and foreigners 1930: Data are hand-collected from publications of the 1930 census: *Ministerstvo Vnitra a Státní Úřad Statistický (1934): Statistický lexikon obcí v Republice Československé : I., Země Česká, Prague.* We transform the data to the present territorial status of municipalities.

Population by ethnicity 1939: Data for the Czech lands on the German population as of May 1939 are from Bohmann (1959, p. 247); we proxy figures for the Czech population by the 1942 population of the 'Protectorate of Bohemia and Moravia' (Bohmann, 1959, p. 194).

Population by ethnicity 1945: Data for the Czech lands on the German population as of April/May 1945 are from Bohmann (1959, p. 252); we proxy figures for the Czech population by the 1944 population of the 'Protectorate of Bohemia and Moravia', taken from *Státní úřad statistický (1948): Pohyb obyvatelstva v roce 1944, Československá Statistika, Svazek 176, Prague.*

Population by ethnicity 1946: Data for the Czech lands are compiled as follows: Bohmann (1959, p. 202) estimates the total number of German expellees in 1946 at 2,232,541. We add this number to the staying 239,911 Germans to derive the number of Germans still living in the Czech lands by late 1945/early 1946. We proxy figures for the Czech population in early 1946 by the 1945 Czech population of the Czech lands, taken from *Státní úřad statistický (1949): Pohyb obyvatelstva v roce 1945, Československá Statistika, Svazek 178, Prague.*

Population by ethnicity 1947: Data at the political county level and for the Czech lands in total are from Urban (1964) (data as of 27 January 1947); we proxy figures for the Czech population in early 1947 by the 1946 Czech population of the Czech lands, taken from *Státní úřad statistický (1949): Pohyb obyvatelstva v roce 1946, Československá Statistika, Svazek 181, Prague.*

Population by ethnicity 2001: Data at the municipality level are from *Český statistický úřad (2014): Basic data about municipalities in 2001, 4. Population by nationality, Prague.*

German Population in 1930, 1939, 1943, 1944, 1946 (February, April, July, October, December) and 1947 (January): County-level data for 1930 and 1939 as described above ('Population by ethnicity'). County-level data for 1943 and 1944 are collected from *Statistisches Bundesamt (1953): Zivilbevölkerung des Deutschen Reiches 1940-1945, Arb.-Nr. VIII/19/I, Wiesbaden.* Political county-level data for the German population in 1946 (February, April, July, October, December) and 1947 (January) are from Řeháček (2011, p. 259).

Population by sectoral shares 1930, 1950, 2001: Municipality-level data are hand-collected from publications of the 1950 census: *Státní úřad statistický (1958): Sčítání lidu v republice československé ke dni 1. března 1950, díl IV, Hospodářský lexikon obcí, Prague.* We transform the data to the present territorial status of municipalities. Data at the municipality level for 2001 are from *Český statistický úřad (2014): Basic data about municipalities in 2001, 4. Population by economic activities (economic branches), Prague.* Data on industrial shares at the county level in 1930 are the Historická data v GIS projecty by the Urbánní a regionální laboratořhe, available at (`http://web.natur.cuni.cz/ksgrrsek/urrlab_vystupy/download`).

Anti-fascist Germans: We have collected the number of Germans on the municipality level by late 1946 from local archives in Karlovy Vary (`http://www.soaplzen.cz/soka-kv`) and in Sokolov (`http://www.soaplzen.cz/soka-so`). The data cover municipalities in the former counties of Karlovy Vary, Kraslice and Loket. Data on anti-fascist Germans at the level of 13 Czech regions are from Luža (1964).

Migration matrizes: We digitized census data from 22. May 1947 at the county level which includes information on the residence of the respondents by 1 May 1945 from: *Státní úřad statistický (1951): Soupis Obyvatelstva v Československu v letech 1946 a 1947, Československá Statistika, Svazek 184, Prague.*

Population by age and local born status: Data at the municipality level were provided upon request by the Czech Statistical Office.

Local communist party leaders: Data on 546 local Communist party leaders in ten Czech counties in 1959 were collected from local archives. Details are available upon request.

3.10.3 Micro data

Life in Transition Survey: We use the Life in Transition Survey (LITS) micro dataset and geocode the residence of the respondents which are available for waves II (2010) and III (2016). Data are available at the website of the European Bank for Reconstruction and Development (`https://www.ebrd.com/what-we-do/economic-research-and-data/data/lits.html`).

3.10.4 Geodata

Country boundaries: Data on country boundaries as of 1930 are from MPIDR (Max Planck Institute for Demographic Research) and CGG (Chair for Geodesy and Geoinformatics, University of Rostock) (2013): MPIDR Population History GIS Collection – Europe (partly based on © EuroGeographics for the administrative boundaries), Rostock. Boundaries of *Sudetenland* as of the Munich Agreement of 1938 are from: Jiří Nenutil, Martin Váňa, Lukáš Funk: Územní ztráty československa po Mnichovské dohodě na území dnešní čR(Německý zábor). Realizováno z projektu SGS-2013-052 "Právní skutečnosti nacistické okupace a jejích důsledků"řešitel JUDr. Vilém Knoll, Ph.D.

Local boundaries: Data for historical county boundaries are from shape files from the Historická data v GIS projecty by the Urbánní a regionální laboratořhe, available at (`http://web.natur.cuni.cz/ksgrrsek/urrlab_vystupy/download`). Shape files for present-day municipality boundaries as of 2008 are retrieved from the Český úřad zeměměřický a katastrální (`https://www.cuzk.cz`).

Demarcation line: We geocode the demarcation line between US and Red Army forces in 1945 Czechoslovakia based on the information from (Pecka, 1995, p. 61). Our translation reads as follows: *'The demarcation line was created in May 1945 (see map on the page 60) and it was approximately crossing along the railroad Honí Dvořiště, Velešín, České Budějovice; it overlapped with the main road between Kosov and Kamenný Újezd and headed West towards Vltava valey, Kremž, Brloh and Nová Ves. Further, it followed the road to Netolice, Vitějovice, Strunkovice nad Blatnicí, Bavorov, Vodňany, and Radčice. Passing the quota 466 directed to Chvaletice, Křtětice, Božovice, Ražice, Heřmaň, around Putim, on the left flank of Otava around Písek to Oldřichov. Chlaponice, Mladotice, Nová hospoda and then along the road Písek-Plzeň to Sedlec, Blatná, Lnáře, Kasejovice, towards Životice, Nepomuk, Spálené Poříčí, Nezvěstice, Šťáhlavy, and Nord-West via villages Raková, Rokycany, Borek, Svojkovice, Volduchy, Březina, Bezděkov, Stupno, Všenice, Střapole, Kříše. Then turned around Plzeň to Chrást, Třemošná, Horní Bříza, Kaznějov, Nečtiny following the road to Karlovy Vary through villages Třebouň, Toužim, Útvina, Krásné Údolí, East of the city of Teplá along the railroad Bečov nad Teplou-Krásný Jez, following the ridges of Slavkovský forest to Jalový Dvůr near Loket, Vřesová, Jindřichovice, Kraslice, Stříbrná, Bublava and through German teritory to Plavno-Saská Kamenice up to Labe.'* Geodata on the demarcation line in Austria are from Ochsner and Roesel (2020) and for Germany self-compiled based on information from US Military Archives.

Roads, railways, and rivers, roughness, distances: We used the location of roads, railways, and rivers and surface roughness as provided by DIVA-GIS (`http://www.diva-gis.org/gdata`). Distances to the Czech external border and to the demarcation line are computed using GIS.

3.10.5 Other data

Monuments and memorials: We collected geocoded data on war memorials from the website of the Society for Military Memorial Places (`https://www.vets.cz`).

German names: We purchased name matches for all 1.3 million candidates standing in Czech local elections between 1994 and 2018 with the website *forebears.io.* (`https://www.forebears.io`). We code a name as German when the original name or a name converted to the German pronunciation is most frequent to Germany or Austria. For example, 'č' becomes 'tsch'. The authors provide all details on request.

War bombings: We geocode bombing incidences during World War Two in individual municipalities reported by Pecka (1995).

Local public infrastructure: Data on health facilities (2016), kindergartens (2017), water mains (2018), and schools (2016) are from the Small Lexicon of Municipalities of the Czech Republic, annually published by the Czech Statistical Office.

Bibliography

Abdelgadir, A. and Fouka, V. (2020). Political secularism and Muslim integration in the West: Assessing the effects of the French headscarf ban. *American Political Science Review*, 114(3):707–723.

Acemoglu, D., Hassan, T. A., and Robinson, J. A. (2011). Social structure and development: A legacy of the Holocaust in Russia. *The Quarterly Journal of Economics*, 126(2):895–946.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5):1369–1401.

Akbulut-Yuksel, M. and Yuksel, M. (2015). The long-term direct and external effects of Jewish expulsions in Nazi Germany. *American Economic Journal: Economic Policy*, 7(3):58–85.

Alesina, A. and Fuchs-Schündeln, N. (2007). Goodbye lenin (or not?): The effect of communism on people's preferences. *American Economic Review*, 97(4):1507–1528.

Algan, Y., Malgouyres, C., Mayer, T., and Thoenig, M. (2020). The economic incentives of cultural transmission: Spatial evidence from naming patterns across France. *The Economic Journal*, forthcoming.

Amuri, F. D. (2011). Monitoring , monetary incentives and workers ' rents in determining absenteeism.

Antman, F. (2012). Gender, educational attainment, and the impact of parental migration on children left behind. *Journal of Population Economics*, 25(4):1187–1214.

Antman, F. M. (2011). International migration and gender discrimination among children left behind. *American Economic Review*, 101(3):645–649.

Antman, F. M. (2013). The impact of migration on family left behind. In Constant, A. F. and Zimmermann, K. F., editors, *International Handbook on the Economics of Migration*, Chapters, chapter 16, pages 293–308. Edward Elgar Publishing.

Arbatli, C. E. and Gokmen, G. (2018). Ethnic capital and sub-national development: Armenian and Greek legacy in post-expulsion Turkey. Working paper, National Research University Higher School of Economics, Moscow.

Arbatli, C. E. and Gomtsyan, D. (2019). Voting retrospectively: Critical junctures and party identification. *European Economic Review*, 119:356–390.

Atkin, D., Colson-Sihra, E., and Shayo, M. (2020). How do we choose our identity? a revealed preference approach using food consumption. *Journal of Political Economy*, forthcoming.

Avdeenko, A. (2018). Long-term evidence of retrospective voting: A natural experiment from the German Democratic Republic. *European Economic Review*, 103:83–107.

Babecký, J., Berson, C., Fadejeva, L., Lamo, A., Marotzke, P., Martins, F., and Strzelecki, P. (2019). Non-base wage components as a source of wage adaptability to shocks: evidence from European firms, 2010–2013. *IZA Journal of Labor Policy*, 8(1):1–18.

Barmby, T. (2002). Worker absenteeism: A discrete hazard model with bivariate heterogeneity. *Labour Economics*, 9(4):469–476.

Bartošek, K. and Pichlík, K. (1951). *Hanebná role amerických okupantů v západních Čechách v roce 1945*. Svoboda, Prague.

Bauer, T. K., Braun, S., and Kvasnicka, M. (2013). The economic integration of forced migrants: Evidence for post-war Germany. *The Economic Journal*, 123(571):998–1024.

Becker, S. O., Boeckh, K., Hainz, C., and Woessmann, L. (2016). The empire is dead, long live the empire! Long-run persistence of trust and corruption in the bureaucracy. *The Economic Journal*, 126(590):40–74.

Becker, S. O. and Ferrara, A. (2019). Consequences of forced migration: A survey of recent findings. *Labour Economics*, 59:1–16.

Becker, S. O., Grosfeld, I., Grosjean, P., Voigtländer, N., and Zhuravskaya, E. (2020). Forced migration and human capital: Evidence from post-WWII population transfers. *American Economic Review*, 110(5):1430–1463.

Beine, M., Docquier, F., and Rapoport, H. (2008). Brain drain and human capital formation in developing countries: Winners and losers. *The Economic Journal*, 118(528):631–652.

Bell-Fialkoff, A. (1993). A brief history of ethnic cleansing. *Foreign Affairs*, 72(3):110–121.

Beneš, J. (1998). *Německá příjmení u Čechů*. Univerzita J.E. Purkyně, Ústí nad Labem.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2011). Formation and persistence of oppositional identities. *European Economic Review*, 55(8):1046–1071.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2016). Bend it like beckham: Ethnic identity and integration. *European Economic Review*, 90:146–164. Social identity and discrimination.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319.

Blattman, C. (2009). From violence to voting: War and political participation in Uganda. *American Political Science Review*, 103(2):231–247.

Bohmann, A. (1959). *Das Sudetendeutschtum in Zahlen*. Sudetendeutscher Rat, Munich.

Brandes, D. (2001). *Der Weg zur Vertreibung 1938–1945*. Oldenbourg, Munich.

Braun, S. T. and Dwenger, N. (2020). Settlement location shapes the integration of forced migrants: Evidence from post-war Germany. *Explorations in Economic History*, 77:101330.

Brewer, M., Zilio, F., and Crossley, T. (2019). What Do We Really Know about the Employment Effects of the UK ' s National Minimum Wage ?

Brosig-Koch, J., Helbach, C., Ockenfels, A., and Weimann, J. (2011). Still different after all these years: Solidarity behavior in East and West Germany. *Journal of Public Economics*, 95(11):1373–1376.

Bundesministerium für Vertriebene, Flüchtlinge und Kriegsgeschädigte (1957). *Die Vertreibung der deutschen Bevölkerung aus der Tschechoslowakei*, volume 1. dtv, Munich.

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

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Cantoni, D., Hagemeister, F., and Westcott, M. (2020). Persistence and activation of right-wing political ideology. Working paper, University of Munich, Munich.

Card, D. (1992a). Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage.

Card, D. (1992b). Using regional variation in wages to measure the effects of the federal minimum wage. *Ilr Review*, 46(1):22–37.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The Effect of Minimum Wages on Low-Wage Jobs: Evidence from the United States Using a Bunching Estimator.

Chaney, E. and Hornbeck, R. (2016). Economic dynamics in the Malthusian era: Evidence from the 1609 Spanish expulsion of the Moriscos. *The Economic Journal*, 126(594):1404–1440.

Charnysh, V. (2019). Diversity, institutions, and economic outcomes: Post-WWII displacement in Poland. *American Political Science Review*, 113(2):423–441.

Clayton, T. C., Thompson, M., and Meade, T. W. (2008). Recent respiratory infection and risk of cardiovascular disease: Case-control study through a general practice database. *European Heart Journal*, 29(1):96–103.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92:1–45.

Conley, T. G. (2010). Spatial econometrics. In Durlauf, S. N. and Blume, L. E., editors, *Microeconometrics*, pages 303–313. Palgrave Macmillan UK, London.

Csillag, M. (2017). THE INCENTIVE EFFECTS OF SICKNESS - ANALYSIS OF A " NATURAL EXPERIMENT ". Number 2016, pages 1–21.

De Paola, M. (2010). Absenteeism and peer interaction effects: Evidence from an Italian Public Institute. *Journal of Socio-Economics*, 39(3):420–428.

De Paola, M., Scoppa, V., and Pupo, V. (2014). Absenteeism in the Italian Public Sector: The Effects of Changes in Sick Leave Policy. *Journal of Labor Economics*, 32(2):337–360.

Dell, M. (2010). The persistent effects of Peru's mining Mita. *Econometrica*, 78(6):1863–1903.

Dickerson, B. J. (2006). The liberation of Western Czechoslovakia 1945. http://web.

archive.org/web/20180709144950/http://www.militaryhistoryonline.com/wwii/articles/liberation1945.aspx. Accessed: 13 November 2019.

Dippel, C. and Heblich, S. (2021). Leadership in social networks: Evidence from the Forty-Eighters in the Civil War. *American Economic Review*, 111(1):1–35.

Doucouliagos, H. and Stanley, T. D. (2009). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, 47(2):406–428.

Duspivová, K., Fischer, J., Kantorová, P., Matějka, M., Mrázek, P., Procházka, M., Spáčil, P., Smolka, V., Šustová, Š., Vltavská, K., and Zelený, M. (2013). Hodnocení vlivu minimální mzdy na sociálně ekonomický vývoj ČR Certifikovaná metodika. Technical report, Prague.

Eder, C. and Halla, M. (2016). The Long-lasting Shadow of the Allied Occupation of Austria on its Spatial Equilibrium. IZA Discussion Papers 10095, IZA.

Eder, C. and Halla, M. (2018). On the origin of the german east-west population gap. Economics Working Papers 2018-17, Department of Economics, Johannes Kepler University Linz.

Edin, P.-A., Fredriksson, P., and Åslund, O. (2003). Ethnic enclaves and the economic success of immigrants—evidence from a natural experiment. *The Quarterly Journal of Economics*, 118(1):329–357.

Egan, P. J. (2020). Identity as dependent variable: How americans shift their identities to align with their politics. *American Journal of Political Science*, 64(3):699–716.

Eriksson, T. and Pytlikova, M. (2004). Firm-level consequences of large minimum-wage increases in the Czech and Slovak Republics. *Labour*, 18(1):75–103.

European Commission's DG for Employment Social Affairs & Inclusion (2017). Employment, Social Affairs and Inclusion.

Fialová, K. and Mysíková, M. (2009). The minimum wage: Labor market consequences in the Czech Republic. *Finance a Uver - Czech Journal of Economics and Finance*, 59(3):255–274.

Fischer, P. and Kodet, R. (2013). Americans in West Bohemia in 1945. *West Bohemian Historical Review*, 3(1):138–155.

Foitzik, J. (1983). Kadertransfer: Der organisierte Einsatz sudetendeutscher Kommunisten in der SBZ 1945/46. *Vierteljahrshefte für Zeitgeschichte*, 31(2):308–334.

Fontana, N., Nannicini, T., and Tabellini, G. (2017). Historical roots of political extremism: The effects of Nazi occupation of Italy. CESifo Working Paper Series 6838, CESifo, Munich.

Fouka, V. (2019). How Do Immigrants Respond to Discrimination? The Case of Germans in the US During World War I. *American Political Science Review*, 113(2):405–422.

Fouka, V. (2020). Backlash: The unintended effects of language prohibition in us schools after world war i. *Review of Economic Studies*, 87(1):204–239.

Franzel, E. (1967). *Die Vertreibung: Sudetenland 1945/1946*. Podzun-Verlag, Bad Nauheim.

Friedman, J. N., Olsen, T., Pistaferri, L., and Chetty, R. (2011). Adjustment Costs , Firm Responses , and Micro vs . Macro Labor Supply Elasticities : Evidence from Danish Tax Records. *Quarterly Journal of Economics*, page 57.

Fuchs-Schündeln, N. and Schündeln, M. (2020). The long-term effects of Communism in Eastern Europe. *Journal of Economic Perspectives*, 34(2):172–191.

Gelman, A. and Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456.

Gerlach, D. W. (2007). *For Nation and Gain: Economy, Ethicity and Politics in the Czech Borderlands, 1945–1948*. PhD thesis, University of Pittsburgh.

Giuliano, P. and Tabellini, M. (2020). The seeds of ideology: Historical immigration and political preferences in the United States. CEPR Discussion Paper DP14784, CEPR, London.

Glassheim, E. (2016). *Cleansing the Czechoslovak borderlands: Migration, environment, and health in the former Sudetenland.* University of Pittsburgh Press, Pittsburgh.

Grosfeld, I. and Zhuravskaya, E. (2015). Cultural vs. economic legacies of empires: Evidence from the partition of Poland. *Journal of Comparative Economics*, 43(1):55–75.

Guzi, M., Mikula, Š., and Huber, P. (2019). Old sins cast long shadows: The long-term impact of the resettlement of the sudetenland on residential migration. IZA Discussion papers 12536, IZA, Bonn.

Harasztosi, P. and Lindner, A. (2019). Who Pays for the Minimum Wage? *American Economic Review*, 109(8):2693–2727.

Herrmann, M. A. and Rockoff, J. E. (2010). DOES MENSTRUATION EXPLAIN GENDER GAPS IN WORK ABSENTEEISM?

Hoffmann, R. J., Heißig, K., and Kittel, M. (2010). *Odsun – Die Vertreibung der Sudetendeutschen*, volume 2. Sudetendeutsches Archiv, Munich.

Huber, K., Lindenthal, V., and Waldinger, F. (2020). Discrimination, managers, and firm performance: Evidence from "Aryanizations" in Nazi Germany. *Journal of Political Economy*, forthcoming.

Hyytinen, A., Meriläinen, J., Saarimaa, T., Toivanen, O., and Tukiainen, J. (2018). When does regression discontinuity design work? evidence from random election outcomes. *Quantitative Economics*, 9(2):1019–1051.

Ichino, A. and Moretti, E. (2009). Biological Gender Differences , Absenteeism , and the Earnings Gap. *American Economic Journal: Applied Economics*, 1(1):183–218.

Ivlevs, A., Nikolova, M., and Graham, C. (2019). Emigration, remittances, and the subjective well-being of those staying behind. *Journal of Population Economics*, 32(1):113–151.

Jurajda, Š. and Kovač, D. (2021). Names and Behavior in a War. *Journal of Population Economics*, 34(1):1–33.

Kaufmann, C. (1996). Possible and impossible solutions to ethnic civil wars. *International Security*, 20(4):136–175.

Klepsch, P. (2013). 'Gebt Eure Uhren her, ihr werdet nie mehr welche brauchen'. In Kalckhoff, A., editor, *Versöhnung durch Wahrheit*, pages 273–276. Heimatkreis Saaz, Stuttgart.

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

Kučera, J. (1992). Die rechtliche und soziale Stellung der Deutschen in der Tschechoslowakei Ende der 40er Jahre und Anfang der 50er Jahre. *Bohemia*, 33(2):322–337.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Luža, R. (1964). *The Transfer of the Sudeten Germans: A Study of Czech-German Relations, 1933-1962*. New York University Press, New York.

Machin, S., Manning, A., and Rahman, L. (2003). Where the Minimum Wage Bites Hard: Introduction of Minimum Wages to a Low Wage Sector. *Journal of the European Economic Association*, 1(1):154–180.

Mareš, M. (2008). Communist and Post-communist Parties in the Czech Republic and Slovakia. In Backes, U. and Moreau, P., editors, *Communist and Post-communist Parties in Europe*, pages 285–310. Vandenhoeck & Ruprecht, Göttingen.

Martinez, L., Jessen, J., and Xu, G. (2020). A glimpse of freedom: Allied occupation and political resistance in East Germany. ESOC Working Paper No. 18, Empirical Studies of Conflict Project, Princeton University.

Mendelsohn, R. (2010). *Liberation*. Global Publishing Solutions (distributed by the U.S. Embassy's Public Affairs Section), Prague.

Meyer, R. and Wise, D. (1983). The Effects of the Minimum Wage on the Employment and Earnings of Youth. *Journal of Labor Economics*, 1(1):289–308.

Milkie, M. A., Nomaguchi, K. M., and Denny, K. E. (2015). Does the amount of time mothers spend with children or adolescents matter? *Journal of Marriage and Family*, 77(2):355–372.

Ministry of Labor and Social affairs (2018). Unemployment rate.

Ministry of Labour and Social Affairs (2017). Informace o vyplacených dávkách v resortu MPSV ČR v prosinci 2017. Technical report, Ministry of Labour and Social Affairs.

Ministry of Labour and Social Affairs (2018). the Evolution of the Minimum Wage since 1991.

Mišterová, I. (2013). The reception of the presence of the U.S. Army in Pilsen in 1945 in local periodicals. *West Bohemian Historical Review*, 3(1):156–179.

Montalvo, J. G. and Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, 95(3):796–816.

Němeček, J. (2002). *Cesta k dekretům a odsunu Němců: datová příručka*. Littera Bohemica, Prague.

Nunn, N. (2008). The long-term effects of Africa's slave trades. *The Quarterly Journal of Economics*, 123(1):139–176.

Nunn, N. and Wantchekon, L. (2011). The slave trade and the origins of mistrust in Africa. *American Economic Review*, 101(7):3221–52.

Ochsner, C. (2017). Dismantled once, diverged forever? a quasi-natural experiment of red army misdeeds in post-wwii europe. ifo Working Paper 240, ifo Institute, Munich.

Ochsner, C. and Roesel, F. (2017). Activated history – the case of the Turkish Sieges of Vienna. CESifo Working Paper 6586, CESifo, Munich.

Ochsner, C. and Roesel, F. (2020). Migrating Extremists. *The Economic Journal*, 130(628):1135–1172.

OECD (2018). Minimum relative to average wages of full-time workers.

OECD (2019). STAN Industrial Analysis.

Paringer, L. (1983). Women and Absenteeism: Health or Economics? *The American Economic Review*, 73(2):123–127.

Pecka, J. (1995). *Na demarkační čáře: Americká armáda v Čechách v roce 1945.* Ústav pro soudobé dějiny AV ČR, Prague.

Pertold, F. (2019). What if they take it all? Heterogeneous impact of sickness absence reform. *Economics of Transition and Institutional Change*, 27(3):647–671.

Pettersson-Lidbom, P. and Thoursie, P. S. (2013). Temporary Disability Insurance and Labor Supply: Evidence from a Natural Experiment*. *Scandinavian Journal of Economics*, 115(2):485–507.

Pogue, F. C. (1954). *The Supreme Command.* Office of the Chief of Military History, Washington D.C.

Řeháček, K. (2011). *Němci na Karlovarsku v letech 1945-1948.* Karel Řeháček, Plzeň.

Reindl-Mommsen, M. (1967). Die Sudetendeutschen in der Tschechoslowakei nach 1945. *Bohemia*, 8(1):315–324.

Rozenas, A., Schutte, S., and Zhukov, Y. (2017). The political legacy of violence: the long-term impact of stalin's repression in ukraine. *The Journal of Politics*, 79(4):1147–1161.

Ruiz, I. and Vargas-Silva, C. (2013). The economics of forced migration. *The Journal of Development Studies*, 49(6):772–784.

Schechtman, J. B. (1953). Postwar population transfers in Europe: A survey. *The Review of Politics*, 15(2):151–178.

Schneider, U. (1995). Die deutschen Antifaschisten in der ČSR im Jahre 1945, die Situation nach Kriegsende und Aussiedlung in die sowjetische Besatzungszone Deutschlands (SBZ). In Sudetendeutsches Archiv, editor, *Odsun: Die Vertreibung der Sudetendeutschen*, pages 257–276. Sudetendeutsches Archiv, Munich.

Scoppa, V. (2010). The influence of inertia on contract design: Contingency planning in information technology service contracts. *Managerial and Decision Economics*, 31(April):503–515.

Segura, G. M. and Fraga, L. R. (2008). Race and the recall: Racial and ethnic polarization in the California recall election. *American Journal of Political Science*, 52(2):421–435.

Semrad, A. (2015). Immigration and educational spillovers: Evidence from Sudeten German expellees in post-war Bavaria. Discussion Papers in Economics, University of Munich, Department of Economics, Munich.

Slapnicka, H. (2000). Oberösterreichische Demarkationslinien und die Fortsetzung nach Norden. *Oberösterreichische Heimatblätter*, 54(1/2):73–76.

Suppan, A. (2006). Austrians, Czechs, and Sudeten Germans as a community of conflict in the twentieth century. Center for Austrian Studies Working Paper 06-1, Institute for Modern History, University of Vienna, Vienna.

Testa, P. (2020). The economic legacy of expulsion: Lessons from postwar Czechoslovakia. *The Economic Journal*, forthcoming.

The Czech Statistical Office (2017). Statistical yearbook of the Czech Republic 2017. Technical report, Prague.

The Czech Statistical Office (2019). Wages, Labor costs - time series.

The World Bank Group (2019). GDP growth (annual %).

Turnwald, W. (1951). *Dokumente zur Austreibung der Sudetendeutschen.* Selbstverlag der Arbeitsgemeinschaft zur Wahrung Sudetendeutscher Interessen, Munich.

Urban, R. (1964). *Die sudetendeutschen Gebiete nach 1945.* Alfred Metzner Verlag, Frankfurt am Main, Berlin.

Valencia Caicedo, F. (2018). The mission: Human capital transmission, economic persistence, and culture in South America. *The Quarterly Journal of Economics*, 134(1):507–556.

Verdier, T. and Zenou, Y. (2017). The role of social networks in cultural assimilation. *Journal of Urban Economics*, 97:15–39.

Voigtländer, N. and Voth, H.-J. (2012). Persecution perpetuated: The medieval origins of anti-semitic violence in Nazi Germany. *The Quarterly Journal of Economics*, 127(3):1339–1392.

Waldinger, F. (2010). Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany. *Journal of Political Economy*, 118(4):787–831.

Waldinger, F. (2012). Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *The Review of Economic Studies*, 79(2):838–861.

Wiedemann, A. (2016). *"Pojď s námi budovat pohraničí!" Osídlování a proměna obyvatelstva bývalých Sudet 1945-1952.* Prostor, Praha.

Wilde, A. (2015). *Ausprägungen räumlicher Identität in ehemaligen sudetendeutschen Gebieten am Beispiel der Bezirke Tachov (Tachau) und Sokolov (Falkenau).* Universitätsverlag Potsdam, Potsdam.

Ziebarth, N. R. and Karlsson, M. (2010). A natural experiment on sick pay cuts, sickness absence, and labor costs. *Journal of Public Economics*, 94(11-12):1108–1122.