The Labor Market, Inequality, and Health:
Four Empirical Essays

Inaugural-Dissertation zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaft

Fachbereich Wirtschaftswissenschaft
Freie Universität Berlin

vorgelegt von

Johannes Mattis Beckmannshagen, M.Sc.

Berlin, 29. Juni 2022

Erstgutachter
Prof. Dr. Carsten Schröder, DIW Berlin & Freie Universität Berlin

Zweitgutachter
Prof. Dr. Peter Haan, DIW Berlin & Freie Universität Berlin

Datum der Disputation: 18. August 2022

Collaborations with Co-authors and Publications

Chapter 1: Inspections and Compliance: Enforcement of the Minimum Wage Law

- Chapter 1 is based on an article published in *FinanzArchiv: Public Finance Analysis*, which is co-authored in equal parts with Alexandra Fedorets.
- Beckmannshagen, M., & Fedorets, A. (2021). Inspections and Compliance: Enforcement of the Minimum-Wage Law. *FinanzArchiv: Public Finance Analysis*, *77*(1), 1-58.

Chapter 2: Earnings Inequality and Working Hours Mismatch

- Chapter 2 is based on an article published in *Labour Economics*, which is co-authored in equal parts with Carsten Schröder.
- Beckmannshagen, M., & Schröder, C. (2022). Earnings Inequality and Working Hours Mismatch. *Labour Economics*, *76*, 102184.

Chapter 3: Out for Good: Labor Market Effects of Transitory and Persistent Health Shocks

- Chapter 3 is based on an article published as a working paper, which is co-authored in equal parts with Johannes König.
- Beckmannshagen, M., & Koenig, J. (2022). Out for Good: Labor Market Effects of Transitory and Persistent Health Shocks. SSRN Working Paper 4036482.

Chapter 4: Job Tasks and Workers' Health

- Chapter 4 is based on an unpublished article, which is written in sole authorship.
- Beckmannshagen, M. (2022). Job Tasks and Workers' Health.

Acknowledgements

Writing this dissertation would not have been possible without the immense support and assistance of my beloved family, friends, and colleagues.

First, I want to thank my supervisors Carsten Schröder and Peter Haan. Carsten convinced me that pursuing a PhD is an option in the first place. Throughout the 3½ years he was always approachable, and I really appreciate his professional advice and supervision. I learned a lot from him starting in the *Measuring Inequality* class but especially through our collaboration thereafter. Our fruitful collaboration has yielded a chapter of this thesis and his constructive feedback also helped to improve the dissertation overall. Moreover, Carsten always helped out in organizational matters and provided several reports that I needed for my scholarship—sometimes on very short notice. Peter agreed to become my second supervisor halfway through my PhD phase and was always accessible whenever I needed an opinion on professional or organizational matters. I consider myself very lucky to have such great support by my two supervisors.

I also thank Timm Bönke and Luca Stella for serving as committee members in the disputation committee for this thesis.

I thank the Hans-Böckler-Foundation for granting me the scholarship, which allowed me to pursue the PhD.

Moreover, I want to thank Alexandra Fedorets, who hired me as a student assistant back in 2015 and taught me all the fundamentals of empirical research. Alexandra supervised my master thesis and even after I stopped working for her, she was a huge help throughout my whole PhD phase. We also collaborated on a chapter of this thesis, and I really appreciate her advice—not only on professional matters.

Further, I want to thank Johannes König for his great guidance in the past two years. Even before we agreed to collaborate on a chapter of this thesis, he spent a lot of time in helping me to structure my research ideas. Similarly, he coached me for my chapter in sole authorship and provided great impulses to get the best out of it. Our collaboration was not only very productive, I particularly enjoyed the days of working together with him in person after a long time of Webex sessions.

I also want to thank Maximilian Longmuir. Max started with his PhD a year ahead of me. I really benefited from this because I could always rely on his advice dissertation-wise but especially regarding organizational matters with the Böckler Foundation and our scholarship. I particularly remember how Max helped me to overcome a phase of serious self-doubt by being a great chat over some beers one night at Pappelplatz.

A large thanks also to my other Böckler-teammates, especially my office mate Rick Glaubitz, Marie Wachinger and Christine Körner. We organized a great conference together and I always enjoyed being around such inspiring colleagues.

Moreover, I want to thank my colleagues from DIW Berlin, especially Daniel Graeber and Johannes Seebauer, for taking the time and providing really valuable feedback on large parts of this dissertation. Thanks also to Matteo Targa, Christoph Halbmeier, Lorenz Meister and Viola Hilbert for providing such a nice office atmosphere as well as my old SOEP machine room teammates Irakli Sauer, Milan Stille, Tabea Naujoks, Konstantin Göbler and Malte Müller for making my time at DIW a lot more fun and enjoyable.

Outside of the academic world, I want to thank the Weddinger Wiesel Herren 2 Basketball Team for always providing the opportunity to completely let go of anything PhD-related and just go out there and play basketball for a couple of hours.

To my friends, Nils L., Susi, Lukas, Sören, David, Inga, Jan K., Till, Jasper, Mocca, Jona, Jan Z., Mischa, Felix, Diana, Nils O., Koko, Ole, Pauli, Passi, and all the Nordhorners: You are a blast to be around and always cheer me up. I truly appreciate having you in my life.

The same holds for my family, in particular, my mom Karin, my dad Peter, as well as Nik and Eve, who always supported me along the way. It was and is just great to know that you always have my back and try everything to make things easier for me, for which I am sincerely thankful.

Lastly, I want to express my greatest gratitude to my soon-to-be wife Lea. Not only did you handle all my good and bad moods during the past 3½ years. The COVID pandemic also basically made us colleagues, sharing the same desk for 2 of those years. I think it’s incredible how well we managed that and that’s largely because of you and your great consideration for me and my situation. Your work ethic is contagious, and your unconditional moral support is simply amazing. Thank you for hanging in there with me so far—and hopefully for much longer.

Contents

1 Inspections and Compliance: Enforcement of the Minimum Wage Law[1]

1.1 Introduction

Among politicians, the introduction of the statutory minimum wage in Germany is widely considered to be a great success. Despite the pessimistic predictions of massive employment destruction in the simulation models (Knabe et al., 2014), the first *ex post* evaluations show that the observed job loss due to the reform is minor (e.g., Bellmann et al., 2016; Caliendo et al., 2018). However, the positive effects of the reform are also moderate: the lower part of the wage distribution grew substantially less than predicted by simulation models (Caliendo et al., 2017; Müller and Steiner, 2013). This discrepancy between the *ex ante* predictions and *ex post* estimations can be explained by the adjustment mechanisms to the minimum wage introduction that were not considered by the simulations, such as imperfect compliance with the law.[2]

Our paper is the first to empirically address the causal influence of inspections by enforcement authorities on non-compliance in a developed economy. To do this, we use regional variation in inspections and non-compliance levels, as well as the exogenous additional burden on the inspection authorities induced by the refugee influx in Europe. The most similar paper to ours is Ronconi (2010), who uses regional-level data for Argentina to document correlations between enforcement effort and compliance to different labor market regulations. To overcome the underlying endogeneity, the study uses variation in staffing of enforcement agencies by election cycles and shows that more enforcement efforts help to reduce non-compliance. Given the magnitude of the debates on minimum wages, empirical evidence on non-compliance is rare. Some authors name non-compliance as a reason for the non-existence of disemployment effects (Yaniv, 2006; Metcalf, 2008; Basu et al., 2010) and give some rough estimations of its magnitude (Caliendo et al., 2017). Overall, the relationship between inspections and non-compliance remains insufficiently studied, despite the potential threats from non-compliance, such as a reduction in the effective market wage (Yaniv, 2006) or creation of a competitive advantage for non-compliant firms (Benassi, 2011).

[1] This is a post-peer-review, pre-copyedit version of an article published in Finanzarchiv/Public Finance Analysis. The final authenticated version is available online at: https://doi.org/10.1628/fa-2021-0001.

[2] Other examples include re-negotiation of working time arrangements (Caliendo et al., 2017) or adjustments to product prices (Lemos, 2008; Aaronson and French, 2007; MaCurdy, 2015).

Becker (1968) postulates that criminal behavior is dependent on the difference between its benefits and costs. Ashenfelter and Smith (1979) formulate this problem specifically for the case of non-compliance with minimum wages, identifying two influence channels that can be adjusted by public authorities: the probability of being caught for criminal behavior by inspection effort and the fines imposed for detected violations. In Germany, enforcement is achieved through employer inspections by the "Finanzkontrolle Schwarzarbeit" (Financial Control of Illicit Employment, hereafter 'FCIE'), a branch of the German customs authorities. Historically, Main Customs Offices (MCO) are unevenly distributed across regions and have different resources at their disposal, such that regions are exposed to substantial differences in inspection effort.

Based on administrative data from the Federal Ministry of Finance and regional data from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (2015), we first show that population density and GDP per capita are the main explanatory factors for the inspection effort. Using Socio-Economic Panel (SOEP) data, we then quantify the incidence of non-compliance in the jurisdictional areas of each regional MCO. For this purpose, we calculate the number of employees entitled to the minimum wage but earning less than the legal minimum (€8.50) in 2014-2016. An OLS estimation reveals a positive correlation between non-compliance and inspection efforts, which confirms the risk-based nature of inspections. In order to overcome this endogeneity issue, we employ both a fixed-effects estimation and an instrumental variable based on the regional variation in the influx of refugees in 2015, which is exogenous to non-compliance given the existing initial distribution rule of refugees across regions. Refugee influx belongs to the competences of the MCOs. In 2015, FCIE personnel were delegated to support the Federal Office for Migration and Refugees (Deutscher Bundestag, 2016a), implying that in regions with a larger influx, fewer resources were available for labor market inspections. Both estimation strategies show that the causal influence of inspection efforts on non-compliance is virtually non-existent. Finally, we perform a series of robustness checks, primarily addressing the potential measurement error in the SOEP data. When measuring non-compliance by the share of employees earning below €8.00, i.e., deliberately lowering the measurement threshold, we find a small negative effect of inspection density on non-compliance. However, overall we conclude that the effect on inspections on non-compliance is very limited.

Our results call for the critical assessment of the enforcement design beyond increasing inspections efforts or imposing higher fines. Based on the international evidence, potential extensions include imposing reputational costs for non-compliance, improving supply chain monitoring, as well as strengthening labor organizations and workers' legal protection.

This paper is structured as follows. Section 1.2 summarizes international evidence on non-compliance. Section 1.3 reviews the institutional background of the minimum wage in Germany and its empirical evaluation. Section 1.4 describes our data sources and Section 1.5 provides descriptive statistics. Section 1.6 presents our estimation results while robustness checks are shown in Section 1.7. Section 1.8 reviews limitations and potential extensions of this paper. Section 1.9 discusses the policy implications and concludes.

1.2 International Evidence

The pioneering study for the field is Ashenfelter and Smith (1979), which formulates the model of non-compliance to minimum wages based on a cost-and-benefit model[3] and, among others, predicts that higher *non-compliance* should be found among low-wage groups. Using the Current Population Survey (CPS), they test these predictions empirically, finding higher *compliance* rates among typical low-wage groups. Ashenfelter and Smith (1979) argue that this unexpected result might be due to an effective enforcement strategy. Although typical low-wage firms have a greater incentive to violate the minimum wage law, it could be offset by enforcement authorities concentrating on these firms.

Although the theoretical framework of Ashenfelter and Smith (1979) lays the groundwork for the research of minimum wage non-compliance, their empirical results and, in particular, their choice of compliance measure are criticized.[4] Sellekaerts and Welch (1984) use the share of sub-minimum wages to measure non-compliance.[5] Based on the CPS, they document non-compliance rates between 3.5 and 9.29%, which are higher among the typical low-wage groups, thus in line with the predictions of the theoretical framework.

Weil (2005) examines minimum wage compliance in the US apparel industry, which is a typical low-wage industry. The author, surveying randomly selected contractors in the Los Angeles area, finds that 54% of firms were non-compliant with the minimum wage legislation. Although this study relies on a small sample, it contributes a policy-relevant assessment of minimum wage enforcement by employer inspections: it concludes that the economic incentives for non-compliance always outweigh the incentive for compliance with the minimum wage. This result underlines the necessity of alternative approaches to employer inspections. Clemens and Strain (2020) also emphasize that enforcement regimes play an important role in shaping compliance rates.

Non-compliance was also an issue in the UK following the introduction of its National Minimum Wage (NMW) in 1999. For example, the study by Metcalf (2008),

[3]In Appendix 2.7.1, we provide an overview of this theoretical framework and its development over time.

[4]Ashenfelter and Smith (1979) compare the number of individuals earning exactly the minimum wage to the number of sub-minimum wages prior to minimum wage introduction to measure compliance. This measure is likely to overestimate non-compliance.

[5]In our paper, we employ this non-compliance measure; see Section 1.5 for details.

despite existing measurement imprecision of non-compliance rates, points to the main enforcement problems: the low probability of being caught and of low fines if caught. The author concludes that non-compliance might be a reason why the NMW had almost no impact on employment.

The Low Pay Commission[6] publishes yearly reports on the NMW and specifically dedicates one chapter to non-compliance, documenting that non-compliance rates increase after each minimum wage increase before gradually falling, thus implying that employers take time to adapt to new rates (Low Pay Commission, 2017). Calculation of non-compliance in this case is likely to be downward biased, as the wage information stems from employer surveys, where under-reporting of non-compliance is more likely.

In 2010, the Department for Business Innovation & Skills of the UK government announced its NMW Compliance Strategy, thus initiating a multifaceted approach to achieve higher compliance levels. In addition to enlarging the enforcement body, an additional focus was put on education and information, including public naming of non-compliant firms (Department for Business, Innovation and Skills, 2010, p.10). Through the potential reputation damage, the violating firm faces higher costs of non-compliance. Thus, the government increased public awareness both of the minimum wage and of the legal prosecution of violators. Since then, the incidence of sub-minimum wages has declined by about 18 percent according to Low Pay Commission (2015).

1.3 The Minimum Wage in Germany

1.3.1 Institutional Background

In July 2014, the ruling German grand coalition decided to introduce a statutory minimum wage of €8.50 per hour effective January 1, 2015.[7] Simultaneously, the Minimum Wage Commission (MWC) was formed and assigned to advise about future adjustments to the minimum wage level.[8]

The minimum wage is binding for the vast majority of employees. Only a few groups are exempt: employees younger than 18, trainees, some types of interns, and the former long-term unemployed during their first six months of re-employment (MiLoG, 2014, § 22). Additionally, in sectors with sector-specific minimum wages on the basis of collective agreements, the general minimum wage would not become

[6]The Low Pay Commission is a public advisory body counseling the UK government on the NMW.

[7]Previously, wage setting was characterized primarily by collective bargaining. In some sectors, collective bargaining agreements were declared generally binding and imposed sector-specific minimum wages.

[8]Thereafter, the MWC decided upon a general rule that it would recommend minimum wage adjustments following the development of collective wage agreements over the previous two years.

effective until 2017.[9] The minimum wage in Germany is highly binding, as it affected 10 to 14 percent of employment, corresponding to around 4 million employees (Amlinger et al., 2014).

1.3.2 Legal Enforcement

When introducing the minimum wage, the German government made the Financial Control of Illicit Employment (FCIE), a branch of the customs authority, responsible for enforcing the new law (MiLoG, 2014, § 14). Thus, in addition to detecting illegal employment and violations of sector-specific minimum wages, the FCIE's remit was expanded to enforcing the general minimum wage. The plans to extend the FCIE's workforce were slowly fulfilled.

The FCIE chooses firms in which they conduct inspections either randomly or following tips.[10] If violations are detected and employers convicted, they face fines of up to €500,000 (MiLoG, 2014, § 21 (3)). These fines do not include back payments to those employees who were paid sub-minimum wages. Such back payments must be claimed by affected employees in a civil court.

1.3.3 Evaluation of the Minimum Wage in Germany

The Minimum Wage Commission biennially publishes reports providing a comprehensive overview of the effects following the minimum wage introduction.[11] In the third, and most recent, report, the MWC concludes that the introduction of the legal minimum unambiguously increased hourly wages at the lower end of the distribution (Minimum Wage Commission, 2020). This result holds whether data from the Earnings Survey, a voluntary firm-level survey, or the Socio-Economic Panel (SOEP) are analyzed. Caliendo et al. (2017) also document wage growth at the lower end of the hourly wage distribution based on the SOEP and a causal identification strategy utilizing differences in regional exposure to the reform. However, this does not translate to a significant effect on monthly earnings, due to a reduction in contractual hours.

The MWC reports also present descriptive and causal evidence on employment effects of the minimum wage. Descriptively, the number of regular jobs in Germany continuously grew. However, the results of causal analyses are mixed (Minimum Wage Commission, 2020, p.101). While most authors agree on a negative effect on marginal employment[12], results on total employment vary from small negative effects (Caliendo et al., 2018; Bossler and Gerner, 2020) to no significant effects

[9] This was the case for the meat industry, hair dressers, agriculture, temporary work in eastern Germany, the garment sector, and laundry services.

[10] E.g., through the minimum wage hotline at the Federal Ministry of Labor and Social Affairs.

[11] See Minimum Wage Commission (2016, 2018, 2020).

[12] See, e.g., Caliendo et al. (2018); Garloff (2019); Schmitz (2019).

(Garloff, 2019; Dustmann et al., 2022), if not positive effects (Ahlfeldt et al., 2018; Stechert, 2018).

According to the published studies, both employment effects and distributional effects—if existent—are quantitatively relatively small; especially when compared to *ex ante* predictions of simulation models (Müller and Steiner, 2013).

1.3.4 Indications of Non-Compliance in Germany

One potential explanation for the smaller-than-expected changes induced by the minimum wage reform is non-compliance. Caliendo et al. (2017) use the Socio-Economic Panel to show that, in 2015, following the introduction of minimum wages, about 2.1 million eligible employees were paid less than €8.50. In 2016, this number decreased to 1.8 million (Burauel et al., 2017, 2020). The incidence of non-compliance remains even after omitting the legally exempted groups[13] and considering the measurement error in survey data. Non-compliance is especially pronounced among those employed with a mini-job (Pusch and Seifert, 2017; Bachmann et al., 2017).

In addition to individual survey data, non-compliance was also detected in employer surveys. The MWC report states that, based on the data from the Earnings Survey, about 1.0 million employees were detected as being paid less than €8.50 in 2015 (Minimum Wage Commission, 2018). In 2016, this number drops to 0.7 million. The numbers based on the Earnings Survey may be biased for several reasons. Besides measurement error (which is likely to be lower than in employee surveys like the SOEP), the Earnings Survey in 2015 and 2016 was conducted on a voluntary basis with very low response rates, which raises concerns about the representativeness of the data. Moreover, these data do not include those small firms that are more likely to pay low wages and where non-compliance is more pronounced. Last, but not least, gathering data for this survey often relies on bookkeeping systems (such as Datev or SAP) that automatically detect and report cases of sub-minimum wages, giving an opportunity to correct the error in the system.

An additional indication of non-compliance stems from specialized surveys. For instance, Fedorets and Schröder (2019) document that about a half of dependent employees with pre-reform wages below €8.50 report having experienced circumvention measures.

1.4 Data

Our empirical analysis relies on the data on inspections and fines broken down by regional jurisdiction of MCO offices (henceforth, *FCIE data*) to quantify inspection

[13] McGuinness et al. (2020) point out the importance of a precise exclusion of the legally exempted group from the analysis based on survey data.

efforts. Our second data source is the German Socio-Economic Panel (SOEP), which allows us to compute non-compliance measures for the same regional units as in the FCIE data.

Analyzing the data at the aggregate regional level implies that we expect inspections not only to have a straightforward effect on the inspected firms, but also a spillover effect on other firms in the region. The presence of higher inspection efforts and the knowledge about them being a threat to non-compliant behavior acts as a deterrent for businesses in the specific regional jurisdiction.

1.4.1 FCIE Data

FCIE enforcement activities in 2015-2016 are summarized in the responses of the federal government to parliamentary inquiries (Deutscher Bundestag, 2016b). For 2014, we acquired data directly from the Federal Ministry of Finance (2017a,b) based on the Freedom of Information Act. The information on enforcement is available at the aggregate level of the 41 regional MCOs[14] from which the FCIE operates: number of inspections (2014-2016), initiated investigations (2015-2016), and fines imposed (2015-2016).

1.4.2 Socio-Economic Panel (SOEP)

For the identification of non-compliance among the eligible population, we employ the German Socio-Economic Panel (SOEP), v33.1 for the 2013 to 2016 waves (see Goebel et al., 2019). The SOEP is a representative longitudinal study that contains usual socio-demographic information, together with information on monthly earnings and hours worked per contract and in actuality.[15] Moreover, it contains detailed regional information that allows us to identify in which MCO jurisdiction area each respondent lives. After restricting the sample to contain only eligible employees with valid information on monthly earnings and weekly working hours, we have about 10,000 individual observations per year.

Based on the SOEP data, we calculate hourly wages by dividing the gross monthly wage by 4.33[16] times the respective number of hours per week. Obtaining hourly wages from survey data through this method poses the risk of measurement error, as both earnings and hours worked could suffer from misreporting. Bound et al.

[14] For the administrative districts of the MCOs, see `https://www.service.bund.de/Content/DE/Behoerden/Suche/Formular.html?` and search for "Hauptzollamt", last accessed on May 19, 2022.

[15] Information on working hours is a common blind spot in the administrative data. In Germany, administrative data sources, such as the SIAB, only contain a dummy variable for working full or part time. Caliendo et al. (2018) shows that differentiated information on working hours is especially useful for studying minimum wage effects.

[16] 52 weeks in a year are divided by 12 months.

(2001) analyze the incidence of measurement error in survey data and find no clear pattern for the misreporting of earnings, but a tendency to overreport hours worked, potentially due to social desirability issues. To minimize this potential threat, for the calculation of hourly wages and the non-compliance measure, we use the reported working hours as specified in the working contract instead of the reported actual working hours. Additionally, although the measurement error issue is extremely important in quantification of the absolute number of jobs with detected non-compliance, it should play only a minor role when relating non-compliance to inspection efforts at the regional level, under the assumption that measurement error is distributed among the regional units (Galvin, 2016). We additionally address the issue of measurement error in a series of robustness checks in Section 1.7.

1.4.3 Additional Data Sources

In order to control for the state of regional economy, we used lagged information on economic activity, including population density, GDP per capita, marginal employment, unemployment, unemployment benefit II recipients, employment in manufacturing, employment in services, and self-employment[17] provided for 2012 at the NUTS-3 level (Federal Institute for Research on Building, Urban Affairs and Spatial Development, 2015). These regional aggregates are included in the regression analysis to control for regional characteristics that are potentially correlated with inspection density and non-compliance probability.

For the applied instrumental variable approach, the number of accommodated refugees living in each MCO region is calculated, based on data from the Federal Statistical Office (Federal Statistical Office, 2017a). The data refer to the most broadly formulated definition of refugees staying in Germany, including those with accepted, declined, or yet to be determined asylum statuses.[18]

1.5 Descriptive Statistics

1.5.1 Inspection Effort

Based on the FCIE data, Table 1.1 shows the overall number of inspections, investigations, and fines in 2013-2016. In 2013-2014, before the introduction of the minimum wage, the average number of inspections was about 63-64,000. In 2015-2016, the number of inspections dropped by about 30 percent. The drop in the number of

[17] All employment and unemployment characteristics calculated as rates.

[18] This measure also includes refugees who lived in Germany before 2014. We argue that capacities of the MCOs were tied to help out the overwhelmed migration offices. Thus, to capture whether or not the regional migration offices were likely overwhelmed, the total number of refugees is the appropriate measure (in contrast to, e.g., the relative number of refugees per regional population or number of incumbent firms).

inspections in 2015 has several explanations. First, with the introduction of the general minimum wage, inspections required more resources. Secondly, Germany experienced a massive influx of refugees in 2015 and 2016, which also demanded MCO resources. Thirdly, due to a re-structuring within the MCOs at the end of 2014, when the section responsible for conducting inspections was assigned to another subject area (Federal Ministry of Finance, 2014), the capacity for inspections also declined (Gewerkschaft der Polizei – Zoll, 2014).

Table 1.1: Inspections and fines 2013-2016

	2013	2014	2015	2016
Inspections	64,001	63,014	43,637	40,374
Investigations, total	135,016	137,292	128,432	126,315
Investigations, MiLoG			705	1,651
Fines, total (€)	44,700,000	46,700,000	43,400,000	48,700,000
Fines, MiLoG (€)			200,000	1,500,000

Note: Displayed are the number of inspections conducted, investigations initiated, and fines imposed by the FCIE 2013-2016. Sources: BT-Drs. 18/4403, 18/7525, 18/11475.

Despite the drop in the inspection numbers, the total number of investigations based on inspections only went down slightly. The total sum of fines declined in 2015, but then rose in 2016. Expectedly, the fines related to the Minimum Wage Law (MiLoG) are at a very low level in 2015, when the MCOs chose to warn most non-compliant employers, whereas in 2016 they were more likely to assess fines for non-compliance. The average fine per inspection is about €1,000 in 2015 and €1,200 in 2016. The average fines per investigation were €340 in 2015 and €385 in 2016. When counting only investigations related to the Minimum Wage Law, the average fines per investigation were €284 in 2015 and €908 in 2016. In the following, we use the information on inspections in the regression analysis (see Section 1.6) and information on investigations and fines for robustness checks (see Section 1.7). Unfortunately, the available information is limited: for instance, it is unknown what proportion of inspections were initiated due to tip-offs versus those randomly initiated.

Information on the total number of inspections varies by regional MCO, which are very different in their size and jurisdictional area. As we do not have information on the staffing of each regional MCO, we assume that the total number of inspections is a proxy for the MCO inspection resources ($inspections_r$). To make this comparable over regions, we further calculate inspection density ($dens_r$) as a ratio of the total number of inspections in the MCO area to the number of businesses in this area ($businesses_r$)[19]:

$$dens_r = \frac{inspections_r}{businesses_r}. \tag{1.1}$$

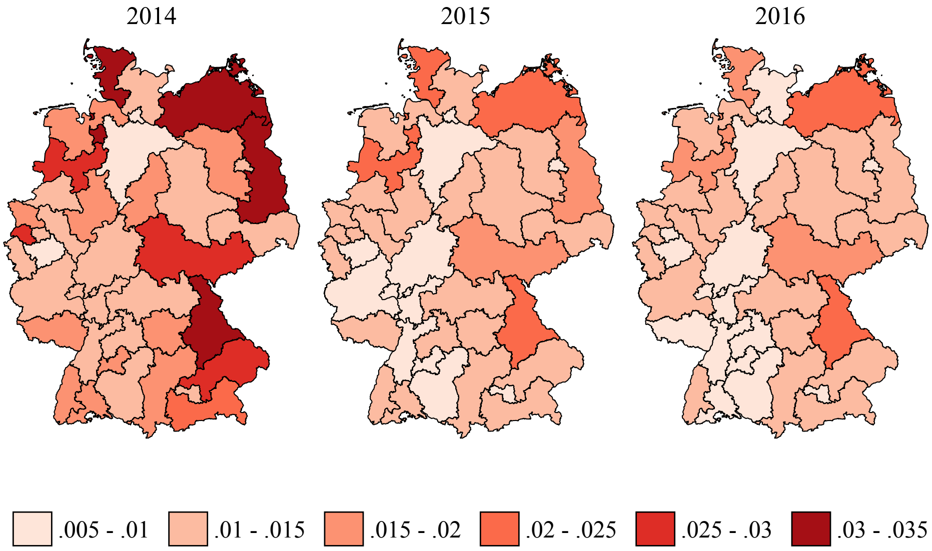
Figure 1.1 shows inspection densities across the 41 MCO areas between 2014 and 2016, with darker shading indicating a higher inspection density. Inspection densities fall between 2014 and 2016 and, on average, they are higher in eastern Germany. In 2014, the highest inspection density was in Bremen, where 3.4 percent of all businesses were inspected. In contrast, the inspection density in Hamburg was much lower: 0.7 percent of all businesses. In 2015, the inspection density in Bremen shrank to 2.4 percent, but still remained the highest density among all MCOs. In Hamburg, the inspection density in 2015 declined only slightly, to 0.5 percent. In 2016, the changes were minor: in Bremen, the inspection density dropped to 2 percent, whereas in Hamburg it slightly grew to 0.6 percent. Overall, these numbers document high variation in inspection density between regions and some variation in inspections within a region over time.

1.5.2 Incidence of Non-Compliance

Based on eligible employees with valid information on monthly wages and contractual working hours from the SOEP, Table 1.2 depicts the distribution of hourly wages in 2013-2016. Before the minimum wage was introduced, about 12 percent of employees earned less than €8.50. After the minimum wage introduction, this share declined to 9 percent in 2015 and to 8 percent in 2016. Notably, only about 2 percentage points of the distribution are located just below the legal minimum, in the interval between 8 and €8.50. Based on the weighting factors from the SOEP, the share of 7.9 percent in 2016 corresponds to 1.7 million eligible employees who earn less than the legal minimum.[20]

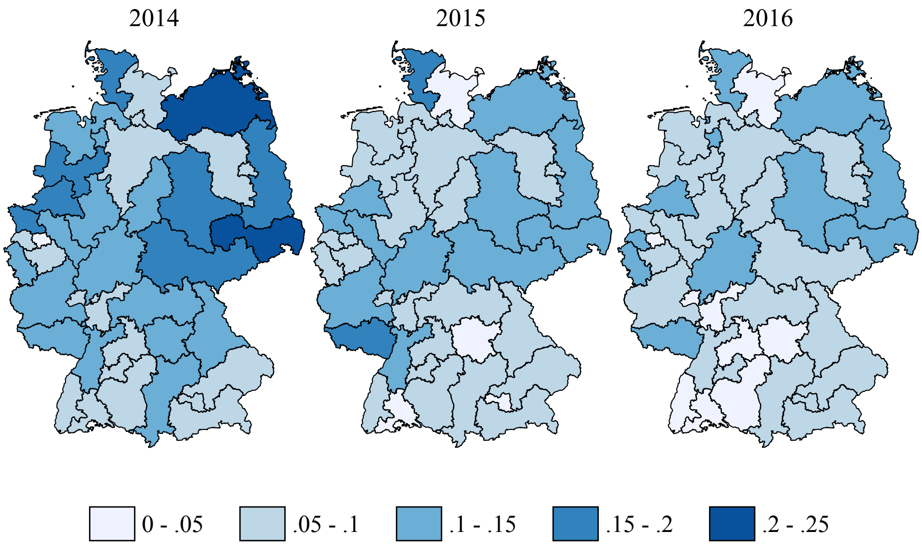
[19]See `https://www.regionalstatistik.de/genesis/online`. Officially, FCIE inspections are conducted among the businesses with at least one regular employee. However, the Federal Statistical Office does not single out such businesses in its statistics. Therefore, we use the total number of businesses in the area under the assumption that it correlates highly with the number of businesses with at least one regular employee.

[20]Table 1.9 in the Appendix depicts the distribution of *actual* hourly wages. Given that most employees work longer hours in actuality than is stated by their contract, the shares below the €8.50-threshold is higher. Therefore, the non-compliance measure based on contractual hours is more conservative.



Note: Displayed values are the sum of inspections in each area of responsibility relative to the number of businesses active in each respective area. Sources: Deutscher Bundestag (2016b); Federal Ministry of Finance (2017a,b).

Figure 1.1: Inspections conducted (relative) by the FCIE in 2014-2016



Note: Displayed values are the share of the eligible population with a contractual hourly wage less than €8.50. Regional units: areas of responsibility of the 41 Main Customs Offices. Source: SOEP v33, own calculations, weighted using the SOEP weighting factors.

Figure 1.2: Share of hourly wages below €8.50 in 2014-2016

Table 1.2: Distribution of contractual hourly wages 2013-2016

	2013	2014	2015	2016
<8 €	0.103	0.095	0.069	0.060
≥8 € - <8.50 €	0.021	0.021	0.020	0.019
≥8.50 € - ≤9 €	0.026	0.027	0.032	0.030
>9 €	0.850	0.857	0.879	0.891
Observations	11,060	10,220	9,496	9,012

Note: Displayed are shares of employees in specific contractual hourly wage bins 2013-2016. Source: SOEP v33 (Minimum wage eligible individuals), own calculations, weighted using the SOEP weighting factors.

Table 1.3 compares the average characteristics of employees earning sub-minimum wages with all eligible employees and shows that women, migrants, East-Germans, as well as employees in part-time and marginal employment are over-represented among eligible employees earning less than the minimum wages. This is in line with international evidence on non-compliance (Galvin, 2016).

Table 1.3: Characteristics of individuals with sub-minimum wages vs. all eligibles

	Hourly wage < €8.50		All eligible individuals	
	2015	2016	2015	2016
Age	42.01	42.39	43.49	43.65
Female	0.70	0.69	0.49	0.49
Non-German citizenship	0.15	0.19	0.10	0.10
East	0.27	0.26	0.19	0.18
Full-time	0.47	0.46	0.78	0.78
Part-time	0.17	0.19	0.14	0.14
Marginal	0.35	0.35	0.06	0.06
Contractual hourly wage	6.64	6.91	18.65	18.90
Observations	846	715	9,496	9,012

Note: Displayed are descriptive statistics of employees with sub-minimum wages versus all other eligible employees. Source: SOEP v33 (Minimum wage eligible individuals), own calculations, weighted using the SOEP weighting factors.

The incidence of observed wages below €8.50 is unevenly distributed across regions. Figure 1.2 shows the share of hourly wages below €8.50 in each MCO area from 2014 to 2016. In 2014, the share of workers incidence varies from about 5 to 25 percent of the eligible employees. In 2015 and 2016, hourly wages grow, leading to a decrease in the share of eligible employees with sub-minimum wages. For instance, in 2016, the overall share of sub-minimum wages was the highest in the jurisdiction of MCO Muenster (13.6 percent), and the lowest in the jurisdiction of MCO Singen (3.1 percent).

1.6 Regression Analysis

In order to elicit the relationship between inspections and non-compliance, we first study the correlation pattern between the inspection density in a region and its economic characteristics.[21] In particular, we estimate the following OLS regression:

$$dens_{r,t} = \alpha + \beta_1\, PopDens_{r,2012} + \beta_2\, GDP_{r,2012} + \gamma X_{r,t} + \epsilon_{r,t}. \tag{1.2}$$

Note: $X_{r,t}$ here contains control variables for the share among the working population of: marginal employees, unemployed, unemployment benefit II recipients, employees in the industrial sector, employees in the services sector, self-employed, and recipients of wages below €8.50 in the previous year.

Table 1.4 shows that a higher population density and a higher GDP per capita are associated with a lower inspection density in all specifications and together explain more than one-third of its variation (specifications 1 and 2). Other regional factors, such as employment and its structure, as well as unemployment characteristics, also relate substantially and significantly to the inspection density and explain a large proportion of its variation (specification 3). When we control for the share of eligible employees with sub-minimum wages in the previous year, the share of the explained variation increases slightly. A one percentage point higher share of employees with sub-minimum wages in the previous year is related to 0.2 percent increase in the inspection density.

This confirms that the degree of inspection density is not randomly distributed on the spatial level but instead depends on regional factors that are correlated with the minimum wage bite, i.e., the share of individuals affected by the minimum wage introduction.[22] This supports the aforementioned hypothesis that a simple regression of non-compliance rates on inspection density would suffer from endogeneity and simultaneity bias.

[21] Given that the inspection density is provided at the level of 41 MCOs and the economic characteristics are available at the NUTS-3 level, we aggregate the latter to the MCO areas.

[22] In 2014, the correlation coefficient between the share of wages below €8.50 and GDP per capita is -0.47; between the share of wages below €8.50 and population density it is -0.43.

Table 1.4: Pooled OLS regression: determinants of inspection density in 2015 and 2016

Log inspection density	(1)	(2)	(3)	(4)
Log population density	-0.398***	-0.323**	-0.274**	-0.214
	(0.085)	(0.122)	(0.131)	(0.134)
East	0.069	0.019	0.289	0.226
	(0.107)	(0.113)	(0.389)	(0.340)
Log GDP per capita		-0.195	-0.486	-0.381
		(0.177)	(0.571)	(0.564)
Log marginal employment			1.319**	1.179**
			(0.636)	(0.550)
Log unemployment			0.829	0.693
			(0.574)	(0.551)
Log unemp. benefit II			-0.384	-0.308
			(0.475)	(0.469)
Log industrial employment			1.122***	1.083***
			(0.330)	(0.301)
Log employment in services			1.737***	1.550**
			(0.605)	(0.588)
Log self-employed			0.982**	1.049**
			(0.442)	(0.430)
Wages below €8.50, prev. year				0.212*
				(0.106)
Year dummy (=1 in 2016)	-0.084***	-0.084***	-0.086***	-0.023
	(0.020)	(0.020)	(0.021)	(0.037)
Observations	82	82	82	82
R^2	0.307	0.316	0.539	0.572

Note: All variables in logs. Robust standard errors. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Source: Federal Ministry of Finance (amount of inspections), INKAR (regional indicators), SOEP v33 (Wages below €8.50), own calculations.

In the next step, we model the individual probability of earning less than the minimum wage to depend on the inspection density in the prior year as well as individual-specific and regional-specific control variables. Specifically, we estimate the following probit regression, with $u85^t_{i,r}$ taking the value of 1 if an individual earns less than €8.50:

$$u85^t_{i,r} = \alpha^t + \beta^t \, insp^{t-1}_r + \gamma^t X^t_{i,r} + \epsilon^t_{i,r}, \quad t = 2015, 2016. \qquad (1.3)$$

Note: All control variables included in vector $X^t_{i,r}$ are displayed in the regression tables.

Table 1.5 shows the estimated marginal effects at means for 2015 and 2016. Specification (1) includes only the main explanatory variable of lagged inspection density. Specifications (2) and (3) additionally include individual-level and regional-level control variables. Finally, specification (4) contains both individual and regional-level control variables.

Without any control variables, the marginal effect of inspection density on non-compliance probability indicates a significant positive correlation between the two variables: a one percent increase in inspection density is statistically associated with a 2.6 (2.1) percentage points higher non-compliance probability in 2015 (2016). This relationship cannot be interpreted as being causal due to the endogeneity of inspection efforts. Specifications (2) to (4) support this suggestion. When controlling for personal characteristics, as in (2), the positive relationship remains only weakly significant. In specifications (3) and (4), inclusion of regional indicators leaves only insignificant marginal effects. Overall, the results reveal that gender, foreign citizenship, low education, marginal employment, as well as employment in the hospitality sector and in small firms are significantly related to the higher probability of sub-minimum wages. Regional characteristics also play their role. Thus, residence in the South is related to lower non-compliance risk, while the opposite applies to individuals in the former eastern Germany (see specification 4).

To address the endogeneity of inspections, we perform an additional estimation in first differences. This approach eliminates potential level effects caused by permanently higher levels of inspection density in certain regions. The change in non-compliance probability is regressed on the change in inspection density and the control variables:

$$\Delta u85_{i,r}^{16-15} = \alpha^{16-15} + \beta^{16-15} \Delta insp_r^{15-14} + \gamma^{16-15} \Delta X_{i,r}^{16-15} + \Delta\epsilon_{i,r}^{16-15}, \tag{1.4}$$

where $\Delta u85_{i,r}^{16-15}$ is equal to 1 when an individual's wages grew from being below the minimum wage threshold to above the minimum wage threshold, and zero otherwise.

Table 1.6 displays the marginal effects at means of this probit estimation in differences. Note that not only are all time-invariant regressors dropped from the regression, but also the individual time-fixed effect, implying that the remaining variance is based only on status changes from non-compliance to compliance, as well as the respective changes in the explanatory variables. The marginal effects of the remaining variables can be interpreted as effects of changes in the respective independent variables on changes of the non-compliance probability. The regression yields almost no correlation of the probability of leaving the non-compliance status and inspection density. The marginal effects are positive, but very small and insignificant. Even when no control variables are included, no significant relationship can be found, as was the case in the respective regression in levels.

Table 1.5: Relationship between noncompliance and inspection density: marginal effects at means of probit estimation

$u85_{i,r,t}$	(1)		(2)		(3)		(4)	
	2015	2016	2015	2016	2015	2016	2015	2016
Log $insp_{r,t-1}$	0.026***	0.021**	0.021*	0.016	-0.013	-0.004	-0.008	-0.007
	(0.010)	(0.010)	(0.011)	(0.011)	(0.008)	(0.006)	(0.005)	(0.004)
Female			0.023***	0.025***			0.021***	0.023***
			(0.005)	(0.005)			(0.004)	(0.004)
Age			-0.000	-0.000			-0.000	-0.000
			(0.000)	(0.000)			(0.000)	(0.000)
German			-0.015**	-0.021***			-0.021***	-0.026***
			(0.006)	(0.007)			(0.006)	(0.006)
Tert. education			-0.039***	-0.035***			-0.037***	-0.033***
			(0.007)	(0.007)			(0.005)	(0.005)
Marginally empl.			0.101***	0.073***			0.095***	0.071***
			(0.010)	(0.008)			(0.007)	(0.006)
Temporary contract			0.033***	0.022***			0.029***	0.020***
			(0.008)	(0.007)			(0.008)	(0.007)
Firm w/ ≥ 10 empl.			-0.047***	-0.047***			-0.041***	-0.041***
			(0.007)	(0.007)			(0.004)	(0.005)
Empl. in hotels & rest.			0.069***	0.053***			0.063***	0.047***
			(0.012)	(0.008)			(0.009)	(0.007)
Log population density					-0.029***	-0.024***	-0.015**	-0.017***
					(0.008)	(0.006)	(0.007)	(0.005)
East					-0.010	0.002	0.003	0.010
					(0.020)	(0.015)	(0.013)	(0.010)
Log GDP p. capita					0.026	0.000	-0.004	-0.011
					(0.030)	(0.031)	(0.025)	(0.024)
Log marginal empl.					0.005	0.010	0.001	0.004
					(0.025)	(0.017)	(0.016)	(0.010)
Log unemployment					0.043	-0.007	0.044	-0.004
					(0.045)	(0.043)	(0.035)	(0.029)

Log unemp. benefit II					-0.003	0.031	-0.017	0.021
					(0.036)	(0.037)	(0.028)	(0.023)
Log industry					0.037**	0.015	0.026**	0.013
					(0.015)	(0.013)	(0.011)	(0.008)
Log services					0.016	0.037	0.031	0.033
					(0.029)	(0.032)	(0.024)	(0.022)
Log self-empl.					0.018	0.003	0.003	0.006
					(0.018)	(0.013)	(0.011)	(0.007)
Log share < €8.50 in t-1					0.045***	0.017***	0.028***	0.010***
					(0.008)	(0.006)	(0.006)	(0.004)
Observations	7,035	7,062	7,035	7,062	7,035	7,062	7,035	7,062
Pseudo R^2	0.004	0.002	0.222	0.195	0.027	0.018	0.253	0.224

Note: $r \in \{1, 41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.6: Changes of noncompliance probability and inspection density: marginal effects at means of probit estimation in differences

$\Delta\ u85_{i,r,16,15}$	(1)	(2)
Δ Log $insp_{r,15,14}$	-0.017	-0.017
	(0.017)	(0.017)
Δ Age		0.007
		(0.005)
Δ German		-0.015
		(0.031)
Δ Marginally employed		-0.053***
		(0.015)
Δ Temporary contract		0.002
		(0.009)
Δ Firm w/ $\geq$ 10 employees		0.014
		(0.009)
Δ Empl. in hotels & restaurants		-0.060**
		(0.030)
Observations	5,690	5,690
Pseudo R^2	0.001	0.031

Note: $r \in \{1, 41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. Taking differences eliminates all time-invariant control variables, such as gender or the regional indicators. The dependent variable is defined to equal 1 if an individual moved from a wage lower than €8.50 in 2015 to at least €8.50 in 2016. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

As a second approach to tackle the endogeneity issue caused by the non-random distribution of inspections, we suggest an instrumental-variable approach. This approach relies on the various responsibilities of the MCOs, each demanding resources. In particular, we use the fact that almost 1 million refugees arrived in Germany in 2015, with the number of asylum seekers in Germany more than doubling between 2014 and 2016 (Federal Statistical Office, 2017a). In order to cope with the situation appropriately, personnel were withdrawn from the FCIE and reassigned to support the Federal Office for Migration and Refugees (Deutscher Bundestag, 2016a). This can be considered to be an exogenous shock to the inspection capacity of MCOs.[23] Thus, the number of refugees accommodated in a region is used as instrument for inspection density: in regions accommodating more refugees, the

[23]The underlying idea of this instrument is similar to Ronconi (2010), who exploits the existence of electoral cycles in labor inspection staffing using panel data on Argentina for 1995-2002.

FCIE was more shorthanded, thus resulting in a reduced inspections capacity.[24] The instrumental variable approach operationalizes this principle in the following two-stage estimation:

$$\begin{aligned} 1^{st}\, stage: \quad & insp_{r,2015} = \alpha_1 + \beta_1\, rfg_{r,2016} + \gamma_1\, X_{i,r,2015} + \epsilon^1_{i,r,2015}, \\ 2^{nd}\, stage: \quad & u85_{i,r,2016} = \alpha_2 + \beta_2\, \widehat{insp}_{r,2015} + \gamma_2\, X_{i,r,2016} + \epsilon^2_{i,r,2016}. \end{aligned} \tag{1.5}$$

The instrumental variable, $rfg_{r,2016}$, describes the number of accommodated refugees in each MCO region at the end of 2016. As the dependent variable is binary, we apply a probit model using maximum likelihood estimation.

The relevance of the instrument can be tested by inspecting the first stage regression. Table 1.10 in Appendix 1.10.2 reveals that the instrument's coefficient is highly significant and the F-statistic sufficiently high, indicating a strong instrument (Angrist and Pischke, 2014).

Whether the exclusion assumption holds cannot be tested statistically. At first glance, it seems plausible that the number of accommodated refugees in a region is independent of an individual's probability to earn less than €8.50. At the same time, the distribution of refugees depends on population and tax revenue (Federal Office for Migration and Refugees, 2016, p. 16), factors that potentially also correlate with minimum wage compliance. Under the assumption that—conditional on the observed regional controls included in the regression—the distribution of refugees is otherwise independent from non-compliance incidence, it can be treated as exogenous. Before refugees can enter the German labor market, local authorities must decide upon their resident status. Due to the large influx of refugees in 2015, the respective decision processes took an extended period of time, making it very unlikely that refugees entered the labor market in 2015, right after arriving in Germany. This strengthens our argument for the exclusion assumption that the number of refugees in a region does not directly affect non-compliance probability in that region.[25]

Table 1.7 shows the marginal effects at the means of this instrumental variable probit estimation. The regression yields a small but insignificant negative effect of the instrumented inspection density in 2015 on the individual probability to be affected by sub-minimum wages. Only in specification (3), including regional level controls, is the marginal effect positive, at 0.004; albeit also insignificant. The effects of the personal characteristics and regional indicators are in line with the results in Table 1.5.

[24]We deliberately do not relate the number of refugees to the regional population size, as we want our instrument to depict the additional burden on MCO personnel, not the region.

[25]Also note, that there are no refugees in the sample of minimum wage eligible employees used for the regression analysis. This rules out a direct impact of refugees on the non-compliance measure.

Table 1.7: Accommodated refugees as instrument for inspection density: marginal effects at means of IV probit estimation

$u85_{i,r,2016}$	(1)	(2)	(3)	(4)
Log $\widehat{insp}_{r,2015}$	-0.007	-0.008	-0.026	-0.025
	(0.021)	(0.018)	(0.021)	(0.016)
Female		0.025***		0.023***
		(0.005)		(0.005)
Age		-0.000		-0.000
		(0.000)		(0.000)
German		-0.019***		-0.026***
		(0.006)		(0.006)
Tert. education		-0.037***		-0.033***
		(0.007)		(0.005)
Marginally employed		0.074***		0.072***
		(0.009)		(0.007)
Temporary contract		0.023***		0.020***
		(0.007)		(0.007)
Firm w/ $\geq$ 10 employees		-0.048***		-0.041***
		(0.007)		(0.005)
Empl. in hotels & restaurants		0.054***		0.047***
		(0.008)		(0.008)
Log population density			-0.028***	-0.021***
			(0.007)	(0.006)
East			0.001	0.009
			(0.016)	(0.012)
Log GDP p. capita			-0.009	-0.017
			(0.037)	(0.029)
Log marginal employment			0.010	0.005
			(0.017)	(0.010)
Log unemployment			-0.004	-0.001
			(0.050)	(0.034)
Log unemp. benefit II			0.030	0.020
			(0.041)	(0.026)
Log industry			0.022	0.019*
			(0.014)	(0.010)
Log services			0.045	0.039
			(0.036)	(0.026)

Log self-employed			0.007	0.009
			(0.013)	(0.007)
Log share < €8.50 in prev. year			0.020***	0.014***
			(0.007)	(0.005)
Observations	7,062	7,062	7,062	7,062

Note: r ∈ {1, 41} ≡ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. (Pseudo-)R^2 cannot be given in IV probit estimation. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Overall, the regression analysis implies that inspections follow a risk-based strategy, where the most inspections effort is observed in the regions with higher non-compliance. At the same time, removing this endogeneity reveals near-zero coefficients of the relationship between inspections and non-compliance, implying that—at that point in time—inspections were broadly ineffective at preventing non-compliance.

1.7 Robustness Checks

We perform several robustness checks of our main results. Table 1.8 gives an overview of the coefficients of the variables of interest: in the upper panel, we display the summary of our main results (i.e. estimation results of equations (1.3), (1.4) and (1.5) from Tables 1.5, 1.6 and 1.7), which we then compare to the results of the following robustness checks.[26]

Robustness Check 1: MCO area aggregate. Our main analysis relies on the individual-level data from the SOEP. In this robustness check, we aggregate the individual-level information from the SOEP and conduct the estimation on the basis on MCO areas. Thus, the dependent variable is the share of sub-minimum wages in each MCO area.

Robustness Check 2: Alternative enforcement measures. In addition to focusing on the inspection density to measure the enforcement efforts by the authorities, we considered the number of initiated investigations and the sum of total fines that were imposed due to minimum wage violations as our major explanatory variables.[27]

[26]The detailed results of all robustness checks are found in the Appendix 1.10.3.

[27]Note, that an inspection is an on-site procedure of documenting the workings of a firm based on studying their paperwork (such as working contracts, working time records etc.) and multiple interviews with employees and owners of the firm. During an inspection, multiple violations can be detected, thus multiple investigations can be initiated, and multiple fines can be imposed

Robustness Check 3: Contemporary effects. In addition to examining lagged effects of inspections in the prior year on the incidence of non-compliance the year after, we also conducted the whole analysis investigating contemporary effects. However, our preferred specification remains the one with the lagged effects, as we believe that the effect of inspections efforts does not unfold instantaneously. Moreover, the SOEP data are primarily gathered from February to July, so we expect our non-compliance measure to react to inspections from the previous calendar year.

Robustness Check 4: Lower non-compliance threshold. To control for potential measurement error, we also conducted our analyses setting the threshold for minimum wage non-compliance to €8.00 instead of €8.50. This allows for inaccuracy of our hourly wage measure by over 5 percent and increases the probability that all cases we classify as non-compliance are indeed paid below the legal minimum.

Robustness Check 5: Exclusion of marginal employment. Several studies find negative effects of the minimum wage introduction on marginal employment. As this employment type was particularly affected by the minimum wage introduction, non-compliance might be especially widespread among marginal employees. We excluded marginal employees from the sample to check whether our analysis of non-compliance is particularly driven by marginal employees.

Robustness Check 6: Restriction to full-time employment only. As an additional approach to minimize the potential threat of measurement error in reported working hours, we only considered full-time employees in our sample. This group is the least flexible in their working hours, which minimizes the measurement error compared to part-time workers.

All robustness checks confirm the positive relationship between enforcement efforts and non-compliance incidence when no regional control variables are included. Running the analysis on an aggregate level or considering alternative enforcement measures as well as contemporary effects also corroborates the findings of our primary analysis that there is no significant relationship between enforcement and sub-minimum wages as soon as regional controls are included in the analysis.

---

following one inspection. In 2015, the correlation between inspections and investigations (at the regional level) is at 0.65 in absolute values and at 0.49 relative to the number of firms in each region. The correlation between the number of inspections and the fines in 2015 is at 0.03 in absolute values and at 0.36 relative to the number of firms in each region. The information on investigations and fines is only available for 2015, so only the regression of sub-minimum wages in 2016 can be conducted. To address the size and composition of MCO areas, both variables were divided by the number of firms operating in each area, analogously to our approach when considering inspections.

Table 1.8: Overview of primary analysis and robustness checks

	(1)		(2)		(3)		(4)	
	2015	2016	2015	2016	2015	2016	2015	2016
Primary analysis								
Main	0.026***	0.021**	0.021*	0.016	-0.013	-0.004	-0.008	-0.007
FD		-0.017		-0.017				
IV		-0.007		-0.008		-0.013		-0.017
Check 1: MCO area aggregate								
Main	0.031**	0.039***	-0.000	0.019	-0.011	0.005	-0.007	-0.013
FD		0.039***		0.012		0.060**		0.034
Check 2: Alternative enf. measures								
Main (dep. var.: Investigations)		0.021***		0.023***		-0.005		-0.002
Main (dep. var.: Fines)		0.000		0.001		-0.001		-0.001
Check 3: Contemporary effects								
Main	0.037***	0.028***	0.028**	0.023**	-0.015	-0.003	-0.009	-0.007
FD		0.025		0.024				
IV		-0.008		-0.008		-0.029		-0.028
Check 4: Wages < 8.00								
Main	0.011	0.006	0.008	0.004	-0.018***	-0.010**	-0.012***	-0.010**
FD		-0.011		-0.011				
IV		-0.009		-0.010		-0.028		-0.028**
Check 5: Excl. marginal empl.								
Main	0.023**	0.022*	0.018*	0.017*	-0.008	-0.002	-0.004	-0.000
FD		-0.009		-0.010				
IV		0.005		-0.000		-0.010		-0.009
Check 6: Only Full-time empl.								
Main	0.018	0.020*	0.014	0.016*	-0.014**	-0.001	-0.006*	0.002
FD		-0.000		-0.002				
IV		0.003		-0.002		0.011		0.005

Note: 'Main' regressions refer to the basis probit estimation as specified in Equation (1.3). 'FD' refers to regressions in first differences as specified in Equation (1.4). 'IV' refers to the 2SLS instrumental variable approach as specified in Equation (1.5). IV regressions are only included in this table if instrument relevance was confirmed in the first stage regression. The top panel, *Primary analysis*, reports the main coefficients of Tables 1.5, 1.6 and 1.7. Specification (1) contains only the respective enforcement measure as explanatory variable. Specification (2) additionally includes individual level control variables. Specification (3) includes regional level control variables. Specification (4) includes both, individual and regional level control variables. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

However, the robustness checks specifically addressing potential measurement error in working hours and hourly wages give interesting insights. When including regional controls, considering the alternative non-compliance threshold of €8.00 or restricting the sample to only full-time employees yields significant negative coefficients of inspection density. Consequently, a one percent increase of inspection density is associated with a lower probability of having wages below €8.00 of 0.01 to 0.02 percent. When conducting the instrumental variable approach to address endogeneity this association increases to 0.03. Restricting the sample to only full-time employees yields a weakly significant negative relationship for the year 2015. Accordingly, a one percent increase of inspection density is associated with a lower non-compliance probability of around 0.01 percent. In 2016, there is not a significant relationship.

Thus, addressing endogeneity and potential measurement error can yield a significant negative effect of inspections on non-compliance. However, the coefficient remains very low in its magnitude, implying that the effect of inspections on non-compliance remains non-substantial in our empirical framework. Therefore, the robustness checks support our main finding that inspections were ineffective in hindering minimum wage non-compliance.

1.8 Qualifications and Extensions

By using MCO-level information on conducted employer inspections this study is based on a unique data source. Data on the inspection efforts are scarce as the customs authorities only record the number of inspections conducted, the number of investigations initiated, and the sum of fines that are exposed by each MCO. While even these data are not easily accessible, they only offer limited research opportunities for a detailed assessment of the overall enforcement efforts. For example, we cannot distinguish between inspections that were randomly initiated and inspections that were initiated after tip-offs. Also, we have no information about the size of an inspected firm. In the data a random inspection of a small firm only employing a handful of workers is not distinguishable from an inspection of a very large firm after a tip-off. This form of unobserved heterogeneity disguises that the different characteristics of inspections have major consequences for the inspection capacity of a MCO and also for the deterrent effect inspections are supposed to entail.

Also, the main analysis of this paper was carried out, when only two waves of data after the introduction of the minimum wage were available. This was timely, because the study deals with the imperfect implementation of the reform, a topic of high relevance when the first evaluations of the reform were conducted. At the same time, longer time-series of information on inspections and sub-minimum wages would increase the research opportunities regarding the effectiveness of inspections in preventing non-compliance. In particular, a longer time-series would allow for fixed effects regressions with more variation over time. In the main analysis of this paper, we perform a regression in first differences to eliminate region-specific level effects and only investigate differences in inspection density and non-compliance probability. However, we can only perform the first differences estimation for one difference after the minimum wage introduction (2016-2015), at the same using the differences in inspection density between 2015 and 2014 as the main explanatory variable. The difference in inspection density between 2015 and 2014 corresponds temporally to the extension of the inspections' scope—i.e., also investigating potential violations against the newly introduced minimum wage—and the internal re-structuring of the FCIE. For future research, a fixed effects analysis of a longer time-series without such untraceable interferences may be particularly informative of the relationship between minimum wage non-compliance and the authorities' inspection efforts.

Another data limitation concerns measurement error. Our measure of hourly wages is based on survey data. As discussed above, to qualify wages below €8.50 as cases of non-compliance, it is crucial for the hourly wages to be reported correctly. Based on the SOEP, we calculate hourly wages by dividing monthly earnings by the supplied hours. The measure relies on the correct reporting of working hours by the respondents and is thus subject to the same difficulties the enforcement authorities face, when conducting inspections: workers might overreport their supplied working hours, for example, because of social desirability, while employers might underreport an employee's supplied for working hours, falsely pretending to be compliant. As a consequence, the absolute number of sub-minimum wages is much higher based on the SOEP than based on the Earnings Survey (see Section 1.3) and the true number likely lies somewhere in between. As this measurement error does not systematically differ between regions it does not pose a direct threat to the empirical strategy of this paper. However, in the larger context, the absence of reliable information on working hours does not only impede the enforcement efforts by the authorities. It also prevents a reliable and exact measure of non-compliance incidence, which would be of interest for policymakers and researchers.

1.9 Discussion and Conclusion

Three years after Germany decided to introduce the minimum wage, its enforcement remains relevant. This paper documents a positive and significant correlation between inspection efforts and the incidence of non-compliance at the regional level. This result can be explained by the risk-based nature of inspections and implies their endogeneity. We address this by using a fixed-effects estimation and an instrumental variable based on the influx of refugees to Germany, which diverted resources away from inspections. Our results show that inspections efforts do not lead to reduced non-compliance probability at the individual level.

This empirical result can have several explanations. On the one hand, it may be explained by insufficient data. We agree that a longer panel would allow for more elaborate and accurate estimations. On the other hand, the near-zero correlation coefficients between inspection efforts and non-compliance can also imply that the enforcement strategy is ineffective. Indeed, the main criticism of the current enforcement strategy is that is does not create enough disincentives for firms to comply with the minimum wage law. In particular, both the detection probability and the imposed fines are very low.

This conclusion provokes an important question: are labor market inspection economically desirable despite their seemingly low efficiency? Our profound conviction is that the answer is yes. In order to assess efficiency of labor market inspections, one can roughly calculate the amount of foregone earnings due to non-compliance to the minimum wage. Among workers who earn below €8.50 in 2015, the average wage gap to the minimum wage threshold was €1.86 per hour. Multiplying each worker's wage gap with the respective contractual working hours yields an average earnings gap of €212 per month, or €2,544 per year. Extrapolating with SOEP weighting factors, this sums up to €401 million per month, or €4.8 billion per year.[28] Although this is likely an upper bound estimation, the amount of potentially foregone earnings is immense. Investing into labor market inspections and better enforcement thus seems justified. Beyond the straightforward monetary rationale, labor market inspections contain multiple elements, such as inspection of work permits, contracts, working conditions, and are, therefore, a key part of legal enforcement. Abolishing legal enforcement by on-site inspections could provoke unprecedented amounts of fraud in the labor market. However, we also acknowledge that there are multiple ways to improve the inspections practices, to make them both more wide-spread and more efficient.

First, the costs of an inspection can be driven down by improving the quality of working hours records by using digital services, so that these records are kept at the work site and can be immediately used for an inspection. At the moment, gathering information on earnings and working hours is completed during face-to-face interviews, which are very time demanding. Fines for not keeping records of working hours are relatively low. Many firms make use of their right to hand in working time records with a delay of up to 7 days, which opens an opportunity for manipulation. Additionally, profiling firms in terms of its registered personnel and tax information before an inspection could also save expensive resources. Overall, automation would

[28]This calculation is rough and assumes inelastic working hours and employment, as well as it disregards potential measurement errors. The calculation is strongly dependent on the estimation of the amount of non-compliance. Replacing the SOEP-based estimation of non-compliance by the respective estimation based on an employer-based Earnings Survey for 2015, yields the total annual sum of foregone earnings of about €2 billion.

allow for making some inspection procedures low-cost and, therefore, widespread. Thus, on-site inspections could be conducted in a more focused way, involving less resources per inspection, which would also help to increase the number of on-site inspections and the non-compliance detection probability.

Second, the costs of non-compliance for employers must increase. The most straightforward way is to increase the fines. Additionally, one can think of imposing reputation costs for non-compliant behavior. For instance, "naming and shaming", as practiced in the UK, could be adopted, publishing lists of violating firms. This strategy aims at the competitive disadvantage for a firm after appearing on such list and imposes reputation losses that may, in the long term, be higher than losses from fines (Karpoff and Lott Jr., 1993). Alternatively, instead of penalizing misbehavior one could also focus on incentivizing lawful behavior (Ehrlich, 1996). Establishing a *positive list* to make customers aware of lawful wages could create a competitive advantage for firms complying with the law.

Third, minimum wage compliance could be improved by requiring firms to monitor their subcontractors, as proposed by Weil (2005), or by empowering labor organizations to adequately enforce workers' rights (Galvin, 2016).

Fourth, information campaigns could help both to improve the awareness of employees of their rights and steps to take in case of violations. Together with employers' associations, an information campaign can aim at the widespread understanding that paying the minimum wage has a profoundly beneficial impact not just for the firm's staff but also the whole economy.

Minimum wage compliance appears highly socially desirable, as it aims at higher earnings for low-wage workers. In contrast, non-compliance not only relates to economic disadvantages for workers but it also undermines the credibility of policymakers and potentially facilitates elaborate economic fraud. However, it may also be argued that policymakers, employers, and even workers can have incentives for minimum wage non-compliance (Basu et al., 2010), arising from the need for greater economic flexibility and job preservation. However, we believe that when an overwhelming majority of respondents of public opinion polls declares their support for minimum wages (Fedorets and Schröder, 2019), they refer to compliant minimum wages. A weakly compliant minimum wage has multiple drawbacks for the society: lower wages and purchase power, unpaid taxes, abuse of power relations between employees and employers, competition distortion, and political incredulity, which may result in greater economic and political instability and increased fraud incidence in society.

1.10 Appendix

1.10.1 Theoretical Framework

Although labor market models differ in their predictions about the implications of minimum wages on employment, they agree that a binding minimum wage constitutes an increase in costs for the input factor of labor. This increase concerns for, in particular, low-skilled workers, as they are more likely to be at the low end of the hourly wage distribution. Firms might react to an increase in labor costs in several ways. Commonly used models focus on explicit and implicit employment cuts (Neumark et al., 2004), passing the wage increase on consumers through higher prices (Lemos, 2008; Aaronson and French, 2007; MaCurdy, 2015), substituting low-skill labor with high-skill labor (Fairris and Bujanda, 2008; Neumark and Wascher, 2003), or substituting low-skill labor with capital (Aaronson and Phelan, 2017). These models assume full compliance with the minimum wage law.

Ashenfelter and Smith (1979) are the first to formulate a theory for non-compliance to the minimum wage (M). According to their seminal paper, the profit-maximizing employer will weigh the expected profits of violating the minimum wage, $\mathrm{E}(\pi)$ by paying lower wages (w), against the profits of complying with the minimum wage $\pi(M,r,p)$.[29] For the latter, there are two determining factors: the probability that the minimum wage violation will be detected by public enforcement (λ) and the potential penalty a non-complying firm faces after being caught (D). Thus, a profit-maximizing firm would choose not to comply with the minimum wage if:

$$E(\pi) - \pi(M,r,p) = (1-\lambda)\,[\pi(w,r,p) - \pi(M,r,p)] - \lambda D > 0. \tag{1.6}$$

Source: Ashenfelter and Smith (1979, p. 335).

In the framework of Ashenfelter and Smith (1979), penalty D is exogenous, meaning it is a fixed amount independent of the extent of the violation. In this setting, the deterrent supposed to make firms comply with the minimum wage goes through two channels that public authorities can adjust. They can increase the probability that violating firms are caught by increasing the law enforcement's inspection efforts and they can increase the fines imposed if violations are detected. Ashenfelter and Smith (1979) conclude that firms are less likely to comply if the difference between market and minimum wage is large and with an increasing elasticity of labor demand. Accordingly, those firms facing difficulties paying the minimum wage and, if complying, would be likely to lay off workers, have a higher incentive to violate the minimum wage law.

Grenier (1982) argues that the imposed fine depends on the severity of the minimum wage violation. Thus, he modifies the penalty structure of the model of

[29] r denotes the price for other production factors and p denotes the output price.

Ashenfelter and Smith (1979). The penalty D should be considered as endogenous and dependent on the difference between the statutory minimum wage and the free market wage, which leads to contrary conclusions. According to him, the incentive to comply is lower for firms paying wages just below the minimum wage because these firms are also facing lower potential penalties. Further, employers less responsive to changes in labor costs have a lower elasticity of labor demand and are more likely to pay less than the legal minimum.

Chang and Ehrlich (1985) side with Ashenfelter and Smith, arguing that the incentive for non-compliance is higher, the larger is the difference between minimum and market wages, regardless of the penalty structure. Thus, people with particularly low wages would be at high risk of being affected by non-compliance. Chang and Ehrlich (1985) also argue that the size of the imposed fine entailed by non-complying behavior presents the determining channel through which public enforcement authorities could efficiently disincentivize non-compliance with the minimum wage law. Indeed, the combination of low detection probability with high fines is said to be the least resource-consuming and, therefore, optimal (Becker, 1968; Galvin, 2016).

Yaniv (1988) analyzes minimum wage non-compliance on monopsonistic labor markets and shows that, for employers with market-power, non-compliance with the minimum wage can be the profit-maximizing choice. He also introduces partial compliance—a case where employers pay a share of their employees in line with the minimum wage law while also paying some employees less than legally obligated. He concludes that, in monopsonistic labor markets, less enforcement efforts are necessary to discourage non-compliance sufficiently. Compared to competitive labor markets, the incentive for non-compliance would be *a priori* smaller in monopsonistic markets.

As shown before, depending on the labor market structure, a minimum wage above the competitive market level leads to an increase in labor costs for firms. While Ashenfelter and Smith (1979) and Grenier (1982) exclusively concentrate on the evasion of this potential increase in costs, Chang and Ehrlich (1985) also link it to the channel of avoiding a rise in labor costs by laying off employees. They argue that, due to the potential fine a firm has to pay if it is caught non-complying, namely λD in Equation (1.6), the firm's marginal costs of labor increase compared to the marginal costs on a competitive market without the minimum wage.

Thus, even if firms do not comply with the minimum wage, they face higher expected costs for labor resulting in a reduction of employment. The employment level of a non-complying firm would be, therefore, higher than in the full-compliance case, but lower than it would be in a competitive labor market without a minimum wage.

This view is challenged by Yaniv (2006), who argues that minimum wage non-compliance also reduces the effective free market wage. Thus, the positive effect of the detection risk on the marginal labor costs, which Chang and Ehrlich (1985) focus on, would be offset by this reduction in marginal costs of labor due to the market wage decrease. Consequently, a minimum wage with imperfect compliance would not have any effect on employment levels compared to a competitive labor market without a minimum wage.

1.10.2 Additional Tables

Table 1.9: Distribution of actual hourly wages 2013-2016

	2013	2014	2015	2016
<7 €	0.134	0.118	0.097	0.081
≥8 € - <8.50 €	0.021	0.022	0.022	0.021
≥8.50 € - ≤9 €	0.027	0.028	0.031	0.035
>9 €	0.817	0.832	0.850	0.864
Observations	12,059	10,934	10,215	9,677

Note: Displayed are shares of employees in specific actual hourly wage bins 2013-2016. Source: SOEP v33 (Minimum wage eligible individuals), own calculations, weighted using the SOEP weighting factors.

Table 1.10: Refugees as an instrument for inspections: first-stage regression statistics

	(1)	(2)	(3)	(4)
Coefficient	-0.327***	-0.324***	-0.344***	-0.344***
p-value	0.002	0.001	0.001	0.001
F-statistic	11.497	11.747	13.723	13.794
Partial R^2	0.221	0.221	0.241	0.241
Observations	7,062	7,062	7,062	7,062

Note: These are statistics of the first-stage regression following Equation (1.5). They confirm the relevance of the instrument used in the regression yielding the results of Table 1.7. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

1.10.3 Robustness Checks

Table 1.11: Relationship between noncompliance and inspection density: OLS regression on MCO level

$u85_{i,r,t}$	(1) 2015	(1) 2016	(2) 2015	(2) 2016	(3) 2015	(3) 2016	(4) 2015	(4) 2016
Log $insp_{r,t-1}$	0.031**	0.039***	-0.000	0.019	-0.011	0.005	-0.007	-0.013
	(0.013)	(0.013)	(0.014)	(0.018)	(0.013)	(0.013)	(0.014)	(0.012)
Female			-0.120	-0.037			-0.100	-0.338**
			(0.133)	(0.198)			(0.173)	(0.136)
Age			0.010***	0.006*			0.004	0.004
			(0.003)	(0.003)			(0.004)	(0.003)
German			0.151	0.094			0.018	0.031
			(0.099)	(0.071)			(0.132)	(0.079)
Tert. education			-0.012	0.048			-0.015	-0.111
			(0.088)	(0.070)			(0.116)	(0.075)
Marginally empl.			-0.222	-0.239			-0.198	0.062
			(0.228)	(0.196)			(0.376)	(0.144)
Temporary contract			0.062	0.302*			-0.010	0.477**
			(0.127)	(0.153)			(0.146)	(0.184)
Firm w/ $\geq$ 10 empl.			-0.230	-0.449***			-0.322*	-0.558***
			(0.145)	(0.163)			(0.180)	(0.145)
Empl. in hotels & rest.			-0.028	0.263			-0.185	0.316*
			(0.207)	(0.227)			(0.250)	(0.166)
Log population density					-0.015	-0.025	-0.010	-0.058***
					(0.016)	(0.015)	(0.023)	(0.016)
East					-0.023	0.001	-0.045	0.036
					(0.038)	(0.035)	(0.062)	(0.031)
Log GDP p. capita					-0.011	0.074	-0.073	0.113*
					(0.071)	(0.078)	(0.079)	(0.062)
Log marginal empl.					0.012	-0.014	0.029	-0.045
					(0.048)	(0.053)	(0.067)	(0.045)
Log unemployment					0.213***	0.073	0.144	-0.033
					(0.070)	(0.075)	(0.087)	(0.070)

Log unemp. benefit II					-0.143**	-0.021	-0.094	0.056
					(0.053)	(0.059)	(0.071)	(0.061)
Log industry					0.033	-0.008	0.031	-0.019
					(0.027)	(0.027)	(0.036)	(0.032)
Log services					0.066	-0.089	0.133	-0.051
					(0.075)	(0.095)	(0.107)	(0.095)
Log self-empl.					0.031	-0.020	-0.015	-0.054
					(0.042)	(0.038)	(0.066)	(0.035)
Log share < €8.50 in t-1					0.045***	0.031***	0.035*	0.011
					(0.013)	(0.009)	(0.018)	(0.008)
Observations	41	41	41	41	41	41	41	41
Pseudo R^2	0.119	0.135	0.443	0.501	0.591	0.624	0.700	0.821

Note: $r \in \{1, 41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.12: Relationship between noncompliance and inspection density: OLS regeression on MCO level in first differences

Δ Log $u85_{r,t}$	(1)	(2)	(3)	(4)
Δ Log $insp_{r,t-1}$	0.039***	0.012	0.060**	0.034
	(0.013)	(0.036)	(0.022)	(0.034)
Δ Female		-0.192*		-0.235
		(0.103)		(0.147)
Δ Age		-0.001		0.005
		(0.005)		(0.007)
Δ German		-0.058		-0.026
		(0.297)		(0.227)
Δ Tert. education		-0.103		0.063
		(0.104)		(0.214)
Δ Marginally employed		-0.153		-0.167
		(0.253)		(0.310)
Δ Temporary contract		-0.005		0.059
		(0.091)		(0.093)
Δ Firm w/ $\geq$ 10 employees		-0.127		-0.170
		(0.139)		(0.187)
Δ Empl. in hotels & restaurants		0.178		0.097
		(0.292)		(0.374)
Δ Log population density			-0.782*	-0.696
			(0.441)	(0.515)
Δ Log GDP p. capita			0.582	0.945
			(1.329)	(1.969)
Δ Log marginal employment			-0.424	-0.840
			(0.694)	(0.867)
Δ Log unemployment			-2.731	-2.252
			(1.702)	(2.075)
Δ Log unemp. benefit II			2.067	2.080
			(1.350)	(1.696)
Δ Log industry			-1.372**	-1.298
			(0.602)	(0.785)
Δ Log services			-0.549	-1.129
			(1.127)	(1.699)
Δ Log self-employed			-0.178	-0.172
			(0.358)	(0.455)

Δ Log share < €8.50 share in prev. year			-0.027**	-0.026*
			(0.013)	(0.015)
Observations	41	41	41	41
R^2	0.135	0.209	0.312	0.452

Note: r ∈ {1,41} ≡ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors. Between the two years individuals might have moved, thus the MCO-level means of normally time-invariant variables can change and these are not completely eliminated as in the personal-level regression (expect indicator for East Germany). Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.13: Accommodated refugees as instrument for inspection density: 2SLS estimation, MCO level

$u85_{r,2016}$	(1)	(2)	(3)	(4)
$\widehat{\text{Log}\,insp}_{r,2015}$	0.114	-0.109	-0.058	-1.275
	(0.559)	(0.487)	(0.637)	(0.783)
Female		-0.915		-7.149*
		(3.441)		(3.865)
Age		0.184***		0.133*
		(0.063)		(0.073)
German		1.458		-0.894
		(1.244)		(1.804)
Tert. education		0.745		-2.635
		(1.303)		(2.048)
Marginally employed		-3.161		4.483
		(3.424)		(4.108)
Temporary contract		6.604***		12.875***
		(2.326)		(3.877)
Firm w/ ≥ 10 employees		-8.752***		-9.462***
		(2.635)		(2.953)
Empl. in hotels & restaurants		7.683*		8.986*
		(4.191)		(5.013)
Log population density			-0.465	-1.526***
			(0.333)	(0.505)
East			0.458	1.065
			(0.761)	(0.859)
Log GDP p. capita			1.767	2.130
			(1.941)	(1.350)
Log marginal employment			0.471	-0.329

			(0.950)	(1.114)
Log unemployment			1.595	-0.393
			(1.661)	(2.431)
Log unemp. benefit II			-0.749	0.966
			(1.261)	(1.904)
Log industry			0.010	0.422
			(0.534)	(0.775)
Log services			-1.139	0.320
			(1.929)	(1.968)
Log self-employed			-0.558	-0.221
			(0.771)	(0.908)
Log share < €8.50 in prev. year			0.647***	0.302
			(0.199)	(0.180)
Observations	41	41	41	41
R^2	0.039	0.549	0.608	0.718

Note: r ∈ {1,41} ≡ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors. First stage regressions reveal that the relevance condition is not satisfied (see Table 1.14). Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.14: First-stage regression statistics, MCO level

	(1)	(2)	(3)	(4)
Coefficient	-0.320***	-0.283***	-0.314***	-0.284**
p-value	0.0080	0.0049	0.0083	0.0353
F-statistic	7.825	9.202	7.977	5.061
Partial R^2	0.208	0.253	0.214	0.208
Observations	41	41	41	41

Note: First-stage statistics reveal that relevance condition of the instrument used in the regression yielding the results of Table 1.13 is not satisfied. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.15: Relationship between noncompliance and investigations per firm: marginal effects at means of probit estimation

$u85_{i,r,2016}$	(1)	(2)	(3)	(4)
Log $inv_{r,2015}$	0.021***	0.023***	-0.005	-0.002
	(0.005)	(0.005)	(0.007)	(0.006)
Female		0.024***		0.023***
		(0.005)		(0.004)
Age		-0.000		-0.000
		(0.000)		(0.000)
German		-0.024***		-0.026***
		(0.007)		(0.006)
Tert. education		-0.035***		-0.033***
		(0.006)		(0.005)
Marginally employed		0.072***		0.071***
		(0.007)		(0.006)
Temporary contract		0.022***		0.020***
		(0.007)		(0.007)
Firm w/ ≥ 10 employees		-0.045***		-0.041***
		(0.006)		(0.005)
Empl. in hotels & restaurants		0.051***		0.047***
		(0.008)		(0.007)
Log population density			-0.021***	-0.014***
			(0.007)	(0.006)
East			0.008	0.013
			(0.016)	(0.011)
Log GDP p. capita			0.010	-0.003
			(0.030)	(0.023)
Log marginal employment			0.007	0.002
			(0.016)	(0.010)
Log unemployment			-0.010	-0.004
			(0.038)	(0.025)
Log unemp. benefit II			0.029	0.019
			(0.033)	(0.021)
Log industry			0.010	0.008
			(0.012)	(0.008)
Log services			0.023	0.024
			(0.033)	(0.024)

Log self-employed			0.004	0.006
			(0.012)	(0.007)
Log share < €8.50 in prev. year			0.028***	0.017***
			(0.008)	(0.006)
Observations	7,062	7,062	7,062	7,062
Pseudo R^2	0.006	0.205	0.019	0.224

Note: $r \in \{1,41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. The main explanatory variable measures the number of investigations introduced by each MCO, relative to the number of firms in the MCO district in 2015. The number of investigations includes investigations initiated on the basis of three legislative acts (the MiLog, the AEntG and the AÜG), all of which impose legal minimum wages. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.16: Relationship between noncompliance and fines per firm: marginal effects at means of probit estimation

$u85_{i,r,2016}$	(1)	(2)	(3)	(4)
Log $fine_{r,2015}$	0.000	0.001	-0.001	-0.001
	(0.004)	(0.003)	(0.002)	(0.002)
Female		0.025***		0.023***
		(0.005)		(0.004)
Age		-0.000		-0.000
		(0.000)		(0.000)
German		-0.020***		-0.026***
		(0.006)		(0.006)
Tert. education		-0.036***		-0.032***
		(0.007)		(0.005)
Marginally employed		0.073***		0.071***
		(0.008)		(0.006)
Temporary contract		0.022***		0.020***
		(0.007)		(0.007)
Firm w/ $\geq$ 10 employees		-0.048***		-0.040***
		(0.007)		(0.005)
Empl. in hotels & restaurants		0.053***		0.047***
		(0.008)		(0.007)
Log population density			-0.021***	-0.015***
			(0.006)	(0.005)
East			0.006	0.014
			(0.014)	(0.010)

Log GDP p. capita			0.011	-0.001
			(0.030)	(0.023)
Log marginal employment			0.007	0.002
			(0.016)	(0.010)
Log unemployment			0.003	0.004
			(0.041)	(0.027)
Log unemp. benefit II			0.019	0.013
			(0.035)	(0.021)
Log industry			0.009	0.007
			(0.012)	(0.008)
Log services			0.027	0.025
			(0.031)	(0.022)
Log self-employed			0.005	0.006
			(0.012)	(0.007)
Log share < €8.50 in prev. year			0.026***	0.015***
			(0.007)	(0.005)
Observations	7,062	7,062	7,062	7,062
Pseudo R^2	0.000	0.193	0.019	0.224

Note: r ∈ {1,41} ≡ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. The main explanatory variable measures the amount of fines imposed by each MCO, relative to the number of firms in the MCO district in 2015. The number of fines includes fines imposed on the basis of three legislative acts (the MiLog, the AEntG and the AÜG), all of which set legal minimum wages. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.17: Accommodated refugees as instrument for investigations: marginal effects at means of IV probit estimation

$u85_{i,r,2016}$	(1)	(2)	(3)	(4)
$\widehat{\text{Log}\,inv}_{r,2015}$	-0.010	-0.011	-0.018	-0.023
	(0.029)	(0.027)	(0.025)	(0.020)
Female		0.026***		0.023***
		(0.006)		(0.004)
Age		0.000		-0.000
		(0.000)		(0.000)
German		-0.019***		-0.027***
		(0.006)		(0.006)
Tert. education		-0.038***		-0.033***
		(0.008)		(0.005)

Marginally empl.		0.075***		0.072***
		(0.010)		(0.007)
Temporary contract		0.024***		0.020***
		(0.007)		(0.007)
Firm w/ ≥ 10 empl.		-0.050***		-0.041***
		(0.009)		(0.005)
Empl. in hotels & rest.		0.055***		0.048***
		(0.010)		(0.008)
Log population density			-0.025**	-0.021**
			(0.010)	(0.009)
East			0.020	0.032
			(0.028)	(0.023)
Log GDP p. capita			0.011	0.000
			(0.031)	(0.025)
Log marginal empl.			0.010	0.007
			(0.017)	(0.012)
Log unemployment			-0.033	-0.043
			(0.055)	(0.045)
Log unemp. benefit II			0.048	0.050
			(0.043)	(0.034)
Log industry			0.012	0.011
			(0.013)	(0.010)
Log services			0.014	0.008
			(0.038)	(0.031)
Log self-empl.			0.000	-0.000
			(0.014)	(0.010)
Log share < €8.50 in t-1			0.031***	0.022***
			(0.010)	(0.008)
Observations	7,062	7,062	7,062	7,062

Note: r ∈ {1, 41} ≡ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. (Pseudo-)R^2 cannot be given in IV probit estimation. First stage regressions reveal that the relevance condition is not satisfied (see Table 1.18). Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.18: First-stage regression statistics

	(1)	(2)	(3)	(4)
Coefficient	-0.265	-0.263	-0.266**	-0.267**
p-value	0.114	0.112	0.014	0.014
F-statistic	2.614	2.636	6.592	6.613
Partial R^2	0.057	0.057	0.101	0.102
Observations	7,062	7,062	7,062	7,062

Note: First-stage statistics reveal that the relevance condition of the instrument used in the regression yielding the results of Table 1.17 is not satisfied. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.19: Accommodated refugees as instrument for fines: marginal effects at means of IV probit estimation

$u85_{i,r,2016}$	(1)	(2)	(3)	(4)
$\widehat{\text{Log}\mathit{fine}}_{r,2015}$	-0.002	-0.003	-0.006	-0.008
	(0.007)	(0.006)	(0.009)	(0.007)
Female		0.025***		0.023***
		(0.005)		(0.005)
Age		-0.000		-0.000
		(0.000)		(0.000)
German		-0.020***		-0.027***
		(0.006)		(0.007)
Tert. education		-0.036***		-0.033***
		(0.007)		(0.005)
Marginally employed		0.074***		0.073***
		(0.008)		(0.008)
Temporary contract		0.022***		0.020***
		(0.007)		(0.007)
Firm w/ ≥ 10 employees		-0.048***		-0.041***
		(0.007)		(0.005)
Empl. in hotels & restaurants		0.053***		0.048***
		(0.008)		(0.008)
Log population density			-0.025***	-0.021***
			(0.009)	(0.007)
East			0.012	0.022
			(0.019)	(0.016)

Log GDP p. capita			0.017	0.008
			(0.032)	(0.025)
Log marginal employment			0.008	0.004
			(0.016)	(0.010)
Log unemployment			0.023	0.032
			(0.063)	(0.044)
Log unemp. benefit II			0.002	-0.011
			(0.055)	(0.038)
Log industry			0.009	0.007
			(0.012)	(0.008)
Log services			0.028	0.026
			(0.030)	(0.021)
Log self-employed			0.001	0.001
			(0.014)	(0.009)
Log share < €8.50 in prev. year			0.022**	0.011
			(0.010)	(0.008)
Observations	7,062	7,062	7,062	7,062

Note: $r \in \{1, 41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. (Pseudo-)R^2 cannot be given in IV probit estimation. First stage regressions reveal that the relevance condition is not satisfied (see Table 1.20). Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.20: First-stage regression statistics

	(1)	(2)	(3)	(4)
Coefficient	-0.921***	-0.924***	-0.737*	-0.739*
p-value	0.002	0.002	0.088	0.087
F-statistic	10.744	10.830	3.066	3.087
Partial R^2	0.151	0.152	0.076	0.076
Observations	7,062	7,062	7,062	7,062

Note: First-stage statistics reveal that the relevance condition of the instrument used in the regression yielding the results of Table 1.19 is not statisfied. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.21: Contemporary relationship between noncompliance and inspection density: marginal effects at means of probit estimation

$u85_{i,r,t}$	(1)		(2)		(3)		(4)	
	2015	2016	2015	2016	2015	2016	2015	2016
Log $insp_{r,t}$	0.037***	0.028***	0.028**	0.023**	-0.015	-0.003	-0.009	-0.007
	(0.011)	(0.008)	(0.013)	(0.009)	(0.010)	(0.008)	(0.007)	(0.005)
Female			0.023***	0.024***			0.021***	0.023***
			(0.005)	(0.005)			(0.004)	(0.004)
Age			0.000	-0.000			-0.000	-0.000
			(0.000)	(0.000)			(0.000)	(0.000)
German			-0.015**	-0.022***			-0.021***	-0.026***
			(0.006)	(0.007)			(0.006)	(0.006)
Tert. education			-0.039***	-0.034***			-0.037***	-0.033***
			(0.007)	(0.007)			(0.005)	(0.005)
Marginally empl.			0.100***	0.072***			0.095***	0.071***
			(0.010)	(0.008)			(0.007)	(0.006)
Temporary contract			0.033***	0.022***			0.029***	0.020***
			(0.008)	(0.007)			(0.008)	(0.007)
Firm w/ ≥ 10 empl.			-0.047***	-0.047***			-0.041***	-0.041***
			(0.007)	(0.007)			(0.004)	(0.005)
Empl. in hotels & rest.			0.069***	0.053***			0.063***	0.047***
			(0.012)	(0.008)			(0.009)	(0.007)
Log population density					-0.030***	-0.024***	-0.015**	-0.017***
					(0.009)	(0.007)	(0.007)	(0.005)
East					-0.012	0.003	0.002	0.011
					(0.020)	(0.014)	(0.013)	(0.010)
Log GDP p. capita					0.023	-0.000	-0.006	-0.013
					(0.029)	(0.029)	(0.024)	(0.023)
Log marginal empl.					0.006	0.010	0.002	0.005
					(0.025)	(0.017)	(0.016)	(0.010)
Log unemployment					0.036	-0.007	0.041	-0.003
					(0.044)	(0.043)	(0.035)	(0.029)

Log unemp. benefit II					0.002	0.031	-0.013	0.020
					(0.036)	(0.037)	(0.028)	(0.023)
Log industry					0.037**	0.014	0.026**	0.013
					(0.015)	(0.012)	(0.012)	(0.008)
Log services					0.019	0.037	0.032	0.035
					(0.028)	(0.031)	(0.023)	(0.022)
Log self-empl.					0.016	0.003	0.002	0.005
					(0.018)	(0.013)	(0.011)	(0.007)
Log share < €8.50 in t-1					0.045***	0.017***	0.028***	0.011***
					(0.009)	(0.006)	(0.006)	(0.004)
Observations	7,035	7,062	7,035	7,062	7,035	7,062	7,035	7,062
Pseudo R^2	0.005	0.004	0.223	0.198	0.027	0.018	0.253	0.224

Note: $r \in \{1,41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.22: Changes of noncompliance probability and inspection density: marginal effects at means of probit estimation in differences

$\Delta\ u85_{i,r,16,15}$	(1)	(2)
Δ Log $insp_{r,16,15}$	0.025	0.024
	(0.024)	(0.022)
Δ Age		0.007
		(0.005)
Δ German		-0.015
		(0.030)
Δ Marginally employed		-0.052***
		(0.015)
Δ Temporary contract		0.003
		(0.009)
Δ Firm w/ $\geq$ 10 employees		0.014
		(0.009)
Δ Empl. in hotels & restaurants		-0.059**
		(0.029)
Observations	5,690	5,690
Pseudo R^2	0.002	0.031

Note: $r \in \{1,41\} \equiv$ MCO regions. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Standard errors are clustered at MCO level. Taking differences eliminates all time-invariant control variables, such as gender or the regional indicators. Dependent variable is defined to equal 1 if an individual moved from a wage lower than €8.50 in 2015 to at least €8.50 in 2016. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.23: Accommodated refugees as instrument: contemporary marginal effects at means of IV probit estimation

$u85_{i,r,2016}$	(1)	(2)	(3)	(4)
Log $\widehat{insp}_{r,2016}$	-0.008	-0.008	-0.029	-0.028
	(0.023)	(0.019)	(0.024)	(0.018)
Female		0.025***		0.023***
		(0.006)		(0.005)
Age		-0.000		-0.000
		(0.000)		(0.000)
German		-0.019***		-0.026***
		(0.006)		(0.006)
Tert. education		-0.037***		-0.033***
		(0.007)		(0.005)
Marginally employed		0.074***		0.072***
		(0.009)		(0.007)
Temporary contract		0.024***		0.020***
		(0.007)		(0.007)
Firm w/ ≥ 10 employees		-0.048***		-0.041***
		(0.007)		(0.005)
Empl. in hotels & restaurants		0.054***		0.048***
		(0.008)		(0.008)
Log population density			-0.029***	-0.022***
			(0.008)	(0.006)
East			0.002	0.011
			(0.016)	(0.012)
Log GDP p. capita			-0.020	-0.028
			(0.041)	(0.032)
Log marginal employment			0.014	0.008
			(0.017)	(0.011)
Log unemployment			-0.004	-0.001
			(0.047)	(0.032)
Log unemp. benefit II			0.028	0.018
			(0.039)	(0.024)
Log industry			0.024*	0.021*
			(0.015)	(0.011)
Log services			0.054	0.048*
			(0.038)	(0.027)

Log self-employed			0.003	0.005
			(0.013)	(0.007)
Log share < €8.50 in prev. year			0.022***	0.015***
			(0.007)	(0.005)
Observations	7,062	7,062	7,062	7,062

Note: r ∈ {1, 41} ≡ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. (Pseudo-)R^2 cannot be given in IV probit estimation. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.24: First-stage regression statistics, contemporary effects

	(1)	(2)	(3)	(4)
Coefficient	-0.316***	-0.313***	-0.294***	-0.294***
p-value	0.0059	0.0054	0.0009	0.0009
F-statistic	8.471	8.653	12.938	12.989
Partial R^2	0.196	0.197	0.217	0.217
Observations	7,062	7,062	7,062	7,062

Note: These are statistics of the first-stage regression following Equation (1.5) with contemporary effects. They confirm the relevance of the instrument used in the regression yielding the results of Table 1.23. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.25: Relationship between noncompliance and inspection density: marginal effects at means of probit estimation, considering wages under €8 as noncompliance

$u8_{i,r,t}$	(1)		(2)		(3)		(4)	
	2015	2016	2015	2016	2015	2016	2015	2016
Log $insp_{r,t-1}$	0.011	0.006	0.008	0.004	-0.018***	-0.010**	-0.012***	-0.010**
	(0.007)	(0.006)	(0.008)	(0.007)	(0.006)	(0.005)	(0.004)	(0.004)
Female			0.011***	0.017***			0.010***	0.015***
			(0.004)	(0.004)			(0.003)	(0.003)
Age			-0.000	-0.000			-0.000	-0.000
			(0.000)	(0.000)			(0.000)	(0.000)
German			-0.011**	-0.013**			-0.015***	-0.016***
			(0.005)	(0.005)			(0.005)	(0.005)
Tert. education			-0.025***	-0.019***			-0.023***	-0.018***
			(0.006)	(0.004)			(0.004)	(0.003)
Marginally employed			0.081***	0.055***			0.075***	0.053***
			(0.009)	(0.006)			(0.006)	(0.005)
Temporary contract			0.023***	0.005			0.020***	0.005
			(0.006)	(0.006)			(0.006)	(0.006)
Firm w/ $\geq$ 10 employees			-0.035***	-0.033***			-0.031***	-0.030***
			(0.005)	(0.004)			(0.004)	(0.003)
Empl. in hotels & restaurants			0.051***	0.032***			0.047***	0.029***
			(0.009)	(0.005)			(0.008)	(0.005)
Log population density					-0.027***	-0.015***	-0.013**	-0.011***
					(0.008)	(0.005)	(0.006)	(0.004)
East					0.002	0.005	0.010	0.008
					(0.017)	(0.012)	(0.010)	(0.008)
Log GDP p. capita					0.026	0.020	-0.001	0.005

					(0.029)	(0.022)	(0.022)	(0.015)
Log marginal employment					0.013	0.007	0.005	-0.002
					(0.024)	(0.017)	(0.014)	(0.010)
Log unemployment					0.011	0.010	0.018	0.007
					(0.039)	(0.026)	(0.029)	(0.019)
Log unemp. benefit II					0.009	0.007	-0.005	0.005
					(0.032)	(0.022)	(0.023)	(0.014)
Log industry					0.026*	0.012	0.018	0.009
					(0.015)	(0.009)	(0.011)	(0.006)
Log services					0.013	0.006	0.025	0.008
					(0.031)	(0.022)	(0.022)	(0.014)
Log self-employed					0.007	0.014	-0.001	0.013*
					(0.014)	(0.013)	(0.008)	(0.007)
Log share < €8 in prev. year					0.029***	0.013***	0.017***	0.007***
					(0.007)	(0.004)	(0.004)	(0.003)
Observations	7,035	7,062	7,035	7,062	7,035	7,062	7,035	7,062
Pseudo R^2	0.001	0.000	0.244	0.215	0.019	0.011	0.269	0.235

Note: $r \in \{1, 41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.26: Changes of noncompliance probability and inspection density: marginal effects at means of probit estimation in differences, wages under €8

$\Delta\ u8_{i,r,16,15}$	(1)	(2)
Δ Log $insp_{r,15,14}$	-0.011	-0.011
	(0.015)	(0.014)
Δ Age		0.004
		(0.005)
Δ German		-0.019
		(0.028)
Δ Marginally employed		-0.044***
		(0.015)
Δ Temporary contract		-0.009
		(0.007)
Δ Firm w/ $\geq$ 10 employees		0.004
		(0.008)
Δ Empl. in hotels & restaurants		-0.073***
		(0.024)
Observations	5,690	5,690
Pseudo R^2	0.001	0.036

Note: r $\in \{1, 41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. Taking differences eliminates all time-invariant control variables, such as gender or the regional indicators. Dependent variable is defined to equal 1 if an individual moved from a wage lower than €8.50 in 2015 to at least €8.50 in 2016. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.27: Accommodated refugees as instrument: marginal effects at means of IV probit estimation, wages under €8

$u8_{i,r,2016}$	(1)	(2)	(3)	(4)
Log $\widehat{insp}_{r,2015}$	-0.009	-0.010	-0.028	-0.028**
	(0.014)	(0.011)	(0.018)	(0.013)
Female		0.017***		0.015***
		(0.004)		(0.003)
Age		-0.000		-0.000
		(0.000)		(0.000)
German		-0.011**		-0.016***
		(0.005)		(0.005)
Tert. education		-0.020***		-0.018***
		(0.004)		(0.003)
Marginally employed		0.055***		0.054***
		(0.007)		(0.005)
Temporary contract		0.006		0.006
		(0.006)		(0.005)
Firm w/ $\geq$ 10 employees		-0.034***		-0.030***
		(0.005)		(0.004)
Empl. in hotels & restaurants		0.032***		0.029***
		(0.006)		(0.006)
Log population density			-0.019***	-0.014***
			(0.005)	(0.004)
East			0.004	0.007
			(0.014)	(0.010)
Log GDP p. capita			0.012	-0.001
			(0.028)	(0.021)
Log marginal employment			0.007	-0.002
			(0.018)	(0.011)
Log unemployment			0.014	0.010
			(0.030)	(0.020)
Log unemp. benefit II			0.006	0.004
			(0.025)	(0.015)
Log industry			0.018	0.014
			(0.012)	(0.009)
Log services			0.012	0.013
			(0.026)	(0.018)

Log self-employed			0.017	0.016**
			(0.014)	(0.008)
Log share < €8 in prev. year			0.015***	0.009**
			(0.005)	(0.004)
Observations	7,062	7,062	7,062	7,062

Note: r ∈ {1, 41} ≡ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. (Pseudo-)R^2 cannot be given in IV probit estimation. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.28: First-stage regression statistics

	(1)	(2)	(3)	(4)
Coefficient	-0.327***	-0.324***	-0.336***	-0.336***
p-value	0.0016	0.0014	0.0004	0.0004
F-statistic	11.497	11.747	14.722	14.772
Partial R^2	0.221	0.221	0.219	0.219
Observations	7,062	7,062	7,062	7,062

Note: These are statistics of the first-stage regression following Equation (1.5). They confirm the relevance of the instrument used in the regression yielding the results of Table 1.27. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.29: Relationship between noncompliance and inspection density: marginal effects at means of probit estimation, marginal employees excluded

$u85_{i,r,t}$	(1)		(2)		(3)		(4)	
	2015	2016	2015	2016	2015	2016	2015	2016
Log $insp_{r,t-1}$	0.023**	0.022*	0.018*	0.017*	-0.008	-0.002	-0.004	-0.000
	(0.011)	(0.011)	(0.009)	(0.009)	(0.006)	(0.007)	(0.004)	(0.004)
Female			0.019***	0.019***			0.016***	0.017***
			(0.004)	(0.004)			(0.003)	(0.003)
Age			-0.000	-0.000			-0.000	-0.000
			(0.000)	(0.000)			(0.000)	(0.000)
German			-0.015**	-0.013*			-0.021***	-0.017***
			(0.006)	(0.006)			(0.005)	(0.006)
Tert. education			-0.036***	-0.029***			-0.031***	-0.027***
			(0.007)	(0.006)			(0.004)	(0.004)
Temporary contract			0.030***	0.021***			0.025***	0.019***
			(0.007)	(0.006)			(0.006)	(0.006)
Firm w/ ≥ 10 empl.			-0.042***	-0.041***			-0.035***	-0.035***
			(0.006)	(0.007)			(0.004)	(0.004)
Empl. in hotels & rest.			0.051***	0.046***			0.044***	0.040***
			(0.010)	(0.007)			(0.007)	(0.006)
Log population density					-0.008	-0.002	-0.001	-0.008*
					(0.006)	(0.007)	(0.005)	(0.004)
East					-0.004	-0.013**	0.009	0.012*
					(0.007)	(0.006)	(0.009)	(0.006)
Log GDP p. capita					0.001	0.004	-0.001	-0.001
					(0.014)	(0.009)	(0.019)	(0.015)
Log marginal empl.					-0.002	-0.013	0.002	-0.002

					(0.026)	(0.019)	(0.012)	(0.008)
Log unemployment					-0.010	-0.016	0.045*	-0.002
					(0.019)	(0.012)	(0.026)	(0.022)
Log unemp. benefit II					0.074**	-0.009	-0.029	0.009
					(0.037)	(0.033)	(0.019)	(0.017)
Log industry					-0.047*	0.017	0.017*	0.003
					(0.028)	(0.028)	(0.009)	(0.007)
Log services					0.017	0.002	0.014	0.013
					(0.013)	(0.011)	(0.021)	(0.017)
Log self-empl.					0.019	0.029	0.003	0.001
					(0.029)	(0.026)	(0.008)	(0.007)
Log share < €8.50 in t-1					0.014	0.009	0.032***	0.014***
					(0.014)	(0.011)	(0.008)	(0.005)
Observations	6,611	6,659	6,611	6,659	6,611	6,659	6,611	6,659
Pseudo R^2	0.007	0.005	0.140	0.148	0.049	0.036	0.183	0.182

Note: $r \in \{1, 41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.30: Changes of noncompliance probability and inspection density: marginal effects at means of probit estimation in differences, marginal employees excluded

$\Delta\ u85_{i,r,16,15}$	(1)	(2)
Δ Log $insp_{r,15,14}$	-0.009	-0.010
	(0.023)	(0.023)
Δ Age		-0.002
		(0.006)
Δ German		-0.021
		(0.031)
Δ Temporary contract		-0.006
		(0.014)
Δ Firm w/ $\geq$ 10 employees		0.029**
		(0.012)
Δ Empl. in hotels & restaurants		-0.098**
		(0.044)
Observations	5,392	5,392
Pseudo R^2	0.000	0.011

Note: $r \in \{1, 41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. Taking differences eliminates all time-invariant control variables, such as gender or the regional indicators. Dependent variable is defined to equal 1 if an individual moved from a wage lower than €8.50 in 2015 to at least €8.50 in 2016. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.31: Accommodated refugees as instrument: marginal effects at means of IV probit estimation, marginal employees excluded

$u85_{i,r,2016}$	(1)	(2)	(3)	(4)
$\text{Log}\,\widehat{insp}_{r,2015}$	0.005	-0.000	-0.010	-0.009
	(0.018)	(0.015)	(0.017)	(0.013)
Female		0.019***		0.017***
		(0.004)		(0.003)
Age		-0.000		-0.000
		(0.000)		(0.000)
German		-0.011*		-0.017***
		(0.006)		(0.006)
Tert. education		-0.031***		-0.027***
		(0.006)		(0.004)
Temporary contract		0.022***		0.019***
		(0.006)		(0.006)
Firm w/ $\geq$ 10 empl.		-0.042***		-0.035***
		(0.007)		(0.005)
Empl. in hotels & rest.		0.046***		0.040***
		(0.007)		(0.006)
Log population density			-0.010	-0.010**
			(0.017)	(0.005)
East			-0.015**	0.012*
			(0.007)	(0.007)
Log GDP p. capita			0.004	-0.004
			(0.010)	(0.016)
Log marginal empl.			-0.016	-0.001
			(0.021)	(0.008)
Log unemployment			-0.015	0.001
			(0.012)	(0.024)
Log unemp. benefit II			-0.006	0.007
			(0.036)	(0.018)
Log industry			0.015	0.006
			(0.030)	(0.008)
Log services			0.005	0.017
			(0.011)	(0.018)
Log self-empl.			0.033	0.003
			(0.028)	(0.007)

Log share < €8.50 in t-1			0.010	0.016***
			(0.011)	(0.005)
Observations	6,659	6,659	6,659	6,659

Note: r ∈ {1, 41} ≡ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. (Pseudo-)R^2 cannot be given in IV probit estimation. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.32: First-stage regression statistics, marginal employees excluded

	(1)	(2)	(3)	(4)
Coefficient	-0.325***	-0.323***	-0.354***	-0.355***
p-value	0.0016	0.0015	0.0001	0.0001
F-statistic	11.427	11.660	17.715	17.780
Partial R^2	0.219	0.219	0.236	0.237
Observations	6,659	6,659	6,659	6,659

Note: The first-stage statistics confirm the relevance of the instrument used in the regression yielding the results of Table 1.31. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.33: Relationship between noncompliance and inspection density: marginal effects at means of probit estimation; only full-time employees

$u85_{i,r,t}$	(1)		(2)		(3)		(4)	
	2015	2016	2015	2016	2015	2016	2015	2016
Log $insp_{r,t-1}$	0.018	0.020*	0.014	0.016*	-0.014**	-0.001	-0.006*	0.002
	(0.012)	(0.011)	(0.009)	(0.009)	(0.006)	(0.005)	(0.004)	(0.003)
Female			0.019***	0.013***			0.014***	0.009***
			(0.005)	(0.004)			(0.003)	(0.002)
Age			-0.000	0.000			-0.000	0.000
			(0.000)	(0.000)			(0.000)	(0.000)
German			-0.010*	-0.009			-0.014***	-0.012***
			(0.006)	(0.006)			(0.004)	(0.004)
Tert. education			-0.032***	-0.023***			-0.027***	-0.019***
			(0.007)	(0.006)			(0.004)	(0.004)
Temporary contract			0.026***	0.017***			0.020***	0.012***
			(0.005)	(0.005)			(0.004)	(0.005)
Firm w/ $\geq$ 10 empl.			-0.037***	-0.032***			-0.027***	-0.024***
			(0.006)	(0.006)			(0.004)	(0.004)
Empl. in hotels & rest.			0.035***	0.030***			0.029***	0.026***
			(0.008)	(0.006)			(0.005)	(0.005)
Log population density					-0.014**	-0.001	0.002	-0.001
					(0.006)	(0.005)	(0.004)	(0.003)
East					-0.000	-0.003	0.008	0.011*
					(0.006)	(0.004)	(0.009)	(0.006)
Log GDP p. capita					0.001	0.008	-0.021	-0.004
					(0.015)	(0.009)	(0.017)	(0.015)
Log marginal empl.					-0.031	-0.020	0.010	0.007

					(0.027)	(0.021)	(0.011)	(0.006)
Log unempl.					-0.001	0.004	0.036	-0.011
					(0.019)	(0.012)	(0.022)	(0.020)
Log unemp. benefit II					0.069**	-0.016	-0.025	0.016
					(0.035)	(0.027)	(0.016)	(0.015)
Log industry					-0.047*	0.024	0.018**	0.007*
					(0.026)	(0.021)	(0.008)	(0.004)
Log services					0.019	0.011	0.020	0.010
					(0.012)	(0.007)	(0.016)	(0.013)
Log self-empl.					0.032	0.030	0.001	-0.005
					(0.023)	(0.021)	(0.007)	(0.006)
Log share < €8.50 share in t-1					0.016	-0.002	0.023***	0.012***
					(0.013)	(0.011)	(0.006)	(0.003)
Observations	5,068	5,091	5,068	5,091	5,068	5,091	5,068	5,091
Pseudo R^2	0.006	0.007	0.142	0.148	0.066	0.056	0.198	0.197

Note: $r \in \{1, 41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.34: Changes of noncompliance probability and inspection density: marginal effects at means of probit estimation in differences, only full-time employees

$\Delta\ u85_{i,r,16,15}$	(1)	(2)
Δ Log $insp_{r,15,14}$	-0.000	-0.002
	(0.021)	(0.021)
Δ Age		-0.001
		(0.006)
Δ German		0.007
		(0.010)
Δ Temporary contract		-0.008
		(0.010)
Δ Firm w/ $\geq$ 10 employees		0.024*
		(0.013)
Δ Empl. in hotels & restaurants		-0.092***
		(0.035)
Observations	4,024	4,024
Pseudo R^2	0.000	0.020

Note: $r \in \{1, 41\} \equiv$ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. Taking differences eliminates all time-invariant control variables, such as gender or the regional indicators. Dependent variable is defined to equal 1 if an individual moved from a wage lower than €8.50 in 2015 to at least €8.50 in 2016. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.35: Accommodated refugees as instrument: marginal effects at means of IV probit estimation, only full-time employees

$u85_{i,r,2016}$	(1)	(2)	(3)	(4)
$\text{Log}\,\widehat{insp}_{r,2015}$	0.003	-0.002	0.011	0.005
	(0.018)	(0.015)	(0.017)	(0.010)
Female		0.013***		0.009***
		(0.005)		(0.002)
Age		0.000		0.000
		(0.000)		(0.000)
German		-0.007		-0.012***
		(0.005)		(0.004)
Tert. education		-0.025***		-0.019***
		(0.007)		(0.004)
Temporary contract		0.018***		0.012***
		(0.005)		(0.005)
Firm w/ ≥ 10 employees		-0.034***		-0.024***
		(0.006)		(0.004)
Empl. in hotels & restaurants		0.031***		0.026***
		(0.007)		(0.006)
Log population density			0.011	-0.001
			(0.017)	(0.004)
East			-0.001	0.011*
			(0.006)	(0.006)
Log GDP p. capita			0.008	-0.003
			(0.008)	(0.016)
Log marginal employment			-0.016	0.007
			(0.023)	(0.007)
Log unemployment			0.003	-0.012
			(0.013)	(0.020)
Log unemp. benefit II			-0.020	0.017
			(0.030)	(0.016)
Log industry			0.027	0.007
			(0.023)	(0.005)
Log services			0.008	0.009
			(0.010)	(0.015)
Log self-employed			0.026	-0.005
			(0.024)	(0.006)

Log share < €8.50 in prev. year			-0.004	0.012***
			(0.012)	(0.004)
Observations	5,091	5,091	5,091	5,091

Note: r ∈ {1, 41} ≡ MCO regions. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at MCO level. (Pseudo-)R^2 cannot be given in IV probit estimation. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

Table 1.36: First-stage regression statistics

	(1)	(2)	(3)	(4)
Coefficient	-0.314***	-0.312***	-0.339***	-0.339***
p-value	0.0027	0.0024	0.0003	0.0003
F-statistic	10.239	10.468	15.864	15.994
Partial R^2	0.204	0.205	0.218	0.219
Observations	5,091	5,091	5,091	5,091

Note: The first-stage statistics confirm the relevance of the instrument used in the regression yielding the results of Table 1.35. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v33 (Minimum wage eligible individuals), INKAR, Federal Ministry of Finance; own calculations.

2 Earnings Inequality and Working Hours Mismatch[1]

[1]This is a post-peer-review, pre-copyedit version of an article published in Labour Economics. Due to copyright regulations the final authenticated version is not part of this document but available online at: https://doi.org/10.1016/j.labeco.2022.102184.

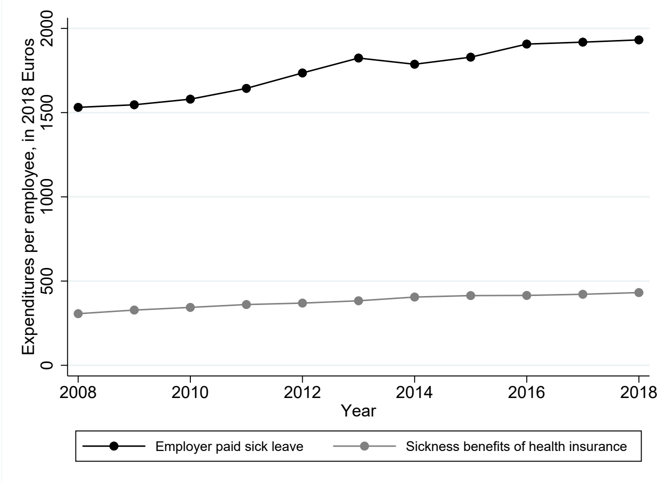
3 Out for Good: Labor Market Effects of Transitory and Persistent Health Shocks

3.1 Introduction

Sudden deteriorations of health pose one of the most severe risks to individual well-being: they directly reduce the quality of life and may require extensive periods of treatment and recovery, limiting the time available for work. The issue of health shocks has wide-ranging consequences, as almost one in three adults in developed countries suffers from two or more chronic illnesses during their lifetime, including hypertension, cancer, and diabetes (OECD, 2019a). For adults between the ages of 50 and 59 who suffer two such illnesses, the probability of being employed is more than 20 percentage points lower than for their healthy counterparts (OECD and European Union, 2016). But illnesses also have many downstream effects beyond the direct effects on the labor market. When a worker has to stop working, their household experiences a downward shift in the budget constraint. This loss has to be offset, whether by the employer, the state, or the family. The costs are high: in Germany, expenditures for sickness-related absences per employee, i.e., wage continuation payments, have steadily increased over the past decades, as Figure 3.1 shows. From 2008 to 2018, real expenditures per employee rose from €1,800 to €2,400, an increase of about 33%. In total, this accounts for employers spending more than € 60 billion on wage continuation payments, while the public health insurance system spends € 14 billion on sickness benefits alone (Federal Statistical Office, 2021b). Combined, the yearly costs for employers and the government are more than twice the yearly spending on unemployment insurance (roughly € 80 billion vs € 35 billion[1] (Federal Statistical Office, 2021c)). While the immediate costs are high, the long-term costs of workers leaving the labor market for good and potentially entering early retirement are even higher. As discussed in Buslei et al. (2019) and Engels et al. (2017), every worker leaving the labor market puts a significant strain on the welfare system, especially in a pay-as-you-go public pension system like Germany's.

In this paper, we quantify the effect of negative health shocks on labor market outcomes and measures of household welfare in Germany using the Socio-Economic Panel (SOEP, Goebel et al. (2019); Schröder et al. (2020)). Our three main con-

[1]The estimated costs of sickness-related absences do not include expenditures for reduced earnings capacity pensions (*Erwerbsminderungsrente*).



Note: Displayed are the real expenditures for sick leave and sickness benefits. Sources: Health expenditure accounts (Federal Statistical Office, 2021b) and Employment Statistics of the Federal Agency for Employment. Both publicly available at the Federal Statistical Office.

Figure 3.1: Real expenditures for health-related absences per employee: employer and state

tributions to the literature lie in measuring and studying the effect heterogeneity of health shocks: First, in a novel approach to measuring health shocks, we use objective measures of health that directly capture working capacity, that is, sick days and hospitalizations[2], to derive health shock indicators. Hence, our indicators are closely connected to labor market behavior and fully incorporate the individual heterogeneity of adverse health events. We validate this by describing the indicators' relationship to previously employed measures of health such as health satisfaction and disease diagnoses, which we find to be strongly correlated. Here, our choice of the SOEP data is vital, as they contain all of these health measures as well as labor market variables. Second, these indicators allow us to distinguish between transitory and persistent health shocks and reveal their differential dynamic impact on labor market outcomes and household welfare. We use machine-learning techniques to differentiate between these two shock types in a data-driven manner. Distinguishing between these shock types has relevance for both policy-making and individual welfare, as rehabilitation and re-entry to the labor market differ vastly between the two types. Third, we document pervasive heterogeneity in effects along several dimensions, finding age to be the most important. Shocks appear to discourage or even prohibit older individuals from participating in the labor market, which is particularly worrisome against the backdrop of a pension system under demographic stress.

[2]We do not consider information on doctor visits because these do not necessarily measure a significant absence from work. A doctor visit may not even correspond to a health event but may only be the result of a routine check-up.

Our pursuit of novel health shock measures is motivated by an important critique of existing approaches: individual heterogeneity may obscure the severity of a given health event regardless of the use of subjective, e.g. health satisfaction or objective health measures, e.g., disease diagnoses (Britton and French, 2020). For example, a diagnosis of cancer 1) may or may not come as a shock and 2) may or may not be severe and long-lasting. Consider that in some cases, a cancer diagnosis may entail chemotherapy with long-term impacts on working capacity, while in other cases, surgery may suffice, and recovery may be fairly quick. Similar arguments apply to subjective health measures, as perceptions of diseases and health states vary from individual to individual (Bound, 1991; Lindeboom and Kerkhofs, 2009; Hosseini et al., 2021b).

Britton and French (2020) and Blundell et al. (2021) model individual heterogeneity as measurement error. In this framework, subjective and objective health measures reflect the true underlying health status or working capacity plus an error term. A central goal of our analysis is to construct a proxy variable that more closely resembles true working capacity.[3] We base our measure on observable behavior: foregoing work or receiving treatment to recover from illnesses. Hence, individual heterogeneity in working capacity is directly integrated into our measure, which prevents measurement error.

The next important empirical challenge is to map the data on our health measures into two latent concepts: transitory and persistent health shocks. Several publications (Hosseini et al., 2021b; Blundell et al., 2021) show that a significant contribution to income risk stems from health risks and, thus, we lean on the life-cycle dynamics of income and labor supply literature (e.g., Blundell et al., 2008, 2016) to motivate our operationalization of health shocks. Thus, just as in the life-cycle dynamics literature, we pursue a binary classification of shocks into transitory and persistent. Further, this binary classification not only harmonizes with the life-cycle dynamics literature, but it is also an intuitive description of the nature of bad health.[4]

To classify our data accordingly, we need a set of assignment rules relating sick days and hospitalizations to the transitory or persistent shock status. *A priori*, it is not clear how to classify individuals based on these variables. Therefore, we rely on data-driven techniques from machine learning, namely clustering. We apply k-means and k-medians algorithms (Friedman et al., 2017) to establish the three groups (persistent shock, transitory shock, control/no shock) within the distributions of sick days and hospitalizations. We use two means of validating that the classification is reasonable and performs well. First, we check whether the group classifications

[3]Whereas papers like Britton and French (2020) and Blundell et al. (2021) use econometric techniques such as instrumental variables estimation to address the issue of measurement error, we pursue a complementary approach by trying to circumvent the issue from the outset.

[4]By intuitive, we mean that some illnesses and injuries such as a bone fracture may lead to a full recovery, whereas others lead to chronic issues with long-term consequences.

capture large fractions of the variation in our health measures. For example, an OLS regression of sick days on the group dummies gives an R^2 of 0.8. Second, we correlate our shock definitions to both objective and subjective measures of health. Individuals in the persistent shock group are more likely to suffer from a severe chronic disease such as cancer, diabetes, or hypertension and to rate their health less favorably than those experiencing a transitory shock or no shock.

Based on this classification, we perform event study analyses in which we estimate the causal effect of experiencing a transitory or a persistent shock on employment, yearly working hours, labor earnings, partner labor earnings, and household net income. Our findings are:

1. There are large and persistent effects on the extensive employment margin for those experiencing a persistent health shock. This is especially severe for older individuals (>50 years), whose employment share drops by 25 percentage points (pp) after experiencing the shock, while the effect is around 10 pp for individuals up to age 50. Those affected by a transitory shock experience a smaller drop in employment (5 pp).

2. Those hit by a persistent shock reduce their labor supply by 700 hours p.a. in the period of the shock, while those hit by a transitory shock reduce their labor supply by 390 hours. After both types of shocks, hours only recover partially to pre-shock levels.

3. Persistent shocks entail a substantial and long-lasting decline of gross labor income, which is reduced by around €6,500 p.a. even three periods after the shock.

4. For partner income we find no evidence of a reaction to either type of shock.

5. Persistent health shocks also reduce household net income. However, the effect size is only 25% of the effect on gross labor income, indicating partial insurance by the family and the tax and transfer system. We do not find a significant effect of transitory shocks on household net income.

The canonical model of health capital (Grossman, 1972) conceptualizes health as a depreciating stock that allows individuals to spend "healthy time", i.e., they have the ability to freely allocate their time between labor and leisure. Many empirical studies, for example Blundell et al. (2021); Hosseini et al. (2021a,b); Capatina (2015); Kemptner (2019), operationalize this concept of health capital by either using survey variables on self-assessed health or by building indices from objective measures like disability classifications or disease diagnoses and pursue structural modelling of health capital.

A second strand of literature also uses these methods of operationalization, but, rather than health capital, models health shocks and estimates their immediate causal impact. For example, García-Gómez et al. (2013) use acute hospitalization records and tax register data to estimate the effect of such shocks on employment and income. Similarly, Schurer (2017) uses the SOEP and information on hospitalizations to examine individual heterogeneity in labor supply responses after a health shock. In the US context, Dobkin et al. (2018) use hospital admissions to investigate the impact on labor market outcomes and beyond, as they also examine medical expenses, credit, and bankruptcy. Our study belongs to this second strand of literature since we also develop concepts of health shocks and investigate their impact on the labor market. However, we combine both sick days and hospitalizations to derive indicators for two types of health shocks, enabling us to give a more comprehensive account of adverse health events. Hospitalizations generally indicate severe health shocks, leaving out more common and less severe health events. Further, some illnesses, although they are severe, do not require for hospitalization. We are able to measure them by combining hospitalizations and sick days. Finally, the distinction between transitory and persistent shocks is important for individual welfare but also for policy-making. Persistent shocks have longer-lasting effects on individuals' labor market outcomes, which must be taken into consideration by policymakers when designing the appropriate mitigation measures, such as the duration and extent of sickness benefits and rehabilitation and re-training programs.

The paper is organized as follows: In Section 3.2, we discuss the particularities of our institutional setting. In Section 3.3, we describe the data basis and derive our health shock classification. In Section 3.4, we delineate the empirical strategy. We present our results in Section 3.5 and discuss and compare them to the existing literature in Section 3.6. Section 3.7 reviews limitations and potential extensions of this paper. Section 3.8 concludes.

3.2 Institutional Background

Germany has created a system of institutions and regulations to alleviate the negative effects of health problems. First, German employees enjoy broad employment protection stemming from the unfair-dismissal act (*Kündigungsschutzgesetz*).While this law does not guarantee employees full protection from termination due to illness, it does stipulate a number of conditions that make it difficult for employers to fire employees for health reasons. Further, under German law, employees have advocates at their place of employment in the form of work councils, which receive notice of all planned terminations and review these decisions. In practice these measures lead to strong protection. The OECD reports that German employment protection ranked seventh on the employment protection index in 2019 among member countries (OECD, 2019b).

Second, employees are entitled to employer-paid sick leave, which covers 100 percent of an employee's salary for up to 6 weeks.[5] Past this period of full replacement, employees become eligible for sickness benefits—paid by the public health insurance system[6]—for the duration of up to 78 weeks including the previous 6 weeks.[7] Sickness benefits do not provide employees with 100% of their regular salary. Generally, they cover 70% of regular gross income but not more than 90% of net income.[8] Once an employee's illness exceeds the 78-week threshold, their main option to receive further benefits is to apply for a partial or full reduction of earnings capacity with the public pension insurance. If individuals are not able to work in any job for at least 3 hours per day, they are granted a full-rate reduced earnings capacity pension. If individuals are able to work at least 3 but less than 6 hours per day, they are granted the half-rate reduced earnings capacity pension, which also permits them to work in a part-time job while receiving the pension. The approval of these pensions is based on assessments by physicians. The amount of the reduced earnings capacity pension depends on how much the individual has paid into the system so far and the pension value of the German public pension system ("Rentenwert"). In 2019, the average reduced earnings capacity pension was €835 before taxes (Rentenverischerung, 2021). Especially for younger individuals who have not yet contributed substantially to the public pension system, the reduced earnings capacity pension will be very low, in some cases even lower than the subsistence minimum defined by social assistance, which amounted to roughly €424 plus rent and heating assistance in 2019 for a single person.

Third, in relation to medical expenses, as German employees have been required since 2009 to have health insurance and, before that, were generally insured by a public health insurance provider, Germans usually do not have to pay out-of-pocket medical expenses to an extent comparable to the United States (Dobkin et al., 2018). In Germany, out-of-pocket medical expenses only occur under special circumstances when patients demand special treatment, e.g., single-patient rooms or treatment by the chief physician, and additional health services, e.g., orthodontic treatments and optometry. Finally, health care prices in Germany are slightly below the OECD average, while the United States ranks eighth among member states (OECD, 2019a).

We provide a brief overview of relevant reforms of the German health care and insurance system in Appendix 3.9.3. Generally, these reforms gradually reduced the generosity of the German health care and insurance system.

To sum up, the German health insurance system covers medical expenses almost completely in stark contrast to the United States. However, job and earnings losses

[5]Between 1996 and 1999, this regulation changed, and sick pay was reduced to 80% of regular salary for those employees who were not protected by a collective bargaining agreement.

[6]Private health insurance providers pay similar amounts, but these contracts are opt-in.

[7]These 78 weeks are counted cumulatively within a period of three years.

[8]For high-income earners, the benefit is capped at 70% of the income ceiling for health insurance contributions, which was €4,537.50 in 2019.

are only partially insured by the employer and the government, making labor market outcomes the relevant variables to study in the German context.

3.3 Data

Our study is based on data from the German Socio-Economic Panel (SOEP), a longitudinal representative household survey, as of 2019, comprising around 30,000 respondents annually (Goebel et al., 2019). The SOEP contains a comprehensive list of socio-economic indicators, detailed labor market information, as well as subjective and objective health measures. For our analyses, we use 27 SOEP waves from the year 1993 to 2018.[9]

We restrict the sample to the working population aged 18 to 65. The sample stretches over the years 1993 to 2017, as some of our variables are retrospectively surveyed. An overview of the number of observations is provided in Table 3.1.

For the definition of our working sample, we exclude spells of mothers. Spells in which women give birth exhibit changes in health and labor market status simultaneously, making them uninformative and potentially contaminating our causal analysis. Further, we exclude the self-employed from our analysis because their access to health care generally differs from the rest of the population in Germany and their income losses are not insured through the state, which alters their incentives to return to work.

Further, in the event-study design, which is our main analytical tool, we restrict the sample to observations that we observe for 7 consecutive years, which enables us to examine 3 relative periods before and after the health shock.

Table 3.1: Observations in the dataset

			Working sample			
	SOEP total	Working age pop.	Treat (trans)	Control (trans)	Treat (pers)	Control (pers)
N	87,171	76,765	1,419	2,646	1,734	2,831
Obs.	575,583	469,186	9,933	20,346	12,138	25,550

Note: N refers to unique individuals in the respective dataset, Obs. refers to person-year observations. Working age population comprises individuals between 18 and 65. The working sample comprises individuals in the transitory shock group, the persistent shock group, and the respective control groups after matching. Source: SOEP v35.

As one might expect, the probability to experience a health shock is not independent of socio-demographic characteristics, which could potentially undermine our identification strategy. Table 3.2 shows descriptive statistics (means and standard

[9] We concentrate on this observation period due to data restrictions. Sick days were not surveyed in the SOEP in 1992, interrupting the time series for one of our essential variables.

deviations) for the control and the two treatment groups. As described in Section 3.4, we use Mahalanobis matching on a set of covariates[10] one period prior to the event taking place and match separately for the transitory and the persistent group, producing two distinct control groups. In Table 3.2, we present the basic descriptives for these two control groups, which show that, compared to the non-matched control group, these two groups are much more similar to their respective treatment groups.

Table 3.2: Descriptive statistics of treatment and control groups

		Age	German	East	Female	Married	Ed 1	Ed 2	Ed 2
Control (all)	Mean	40.68	0.93	0.23	0.47	0.66	0.26	0.43	0.30
	SD	10.16	0.26	0.42	0.50	0.47	0.44	0.49	0.46
Treat (trans)	Mean	43.48	0.92	0.33	0.50	0.71	0.35	0.44	0.20
	SD	10.71	0.26	0.47	0.50	0.45	0.48	0.50	0.40
Control (trans)	Mean	42.06	0.93	0.29	0.49	0.68	0.32	0.44	0.23
	SD	10.11	0.26	0.45	0.50	0.46	0.47	0.50	0.42
Treat (pers)	Mean	45.39	0.92	0.33	0.54	0.71	0.37	0.41	0.21
	SD	10.87	0.28	0.47	0.50	0.45	0.48	0.49	0.41
Control (pers)	Mean	42.95	0.93	0.29	0.51	0.68	0.31	0.43	0.26
	SD	10.27	0.26	0.45	0.50	0.47	0.46	0.49	0.44

Note: Displayed are means and standard deviations of the unmatched control group, as well as the transitory shock group, and the persistent shock group and their respective matched control groups. Ed 1, Ed 2, Ed 3 refers to the respective share with primary, secondary, or tertiary education. Source: SOEP v35.

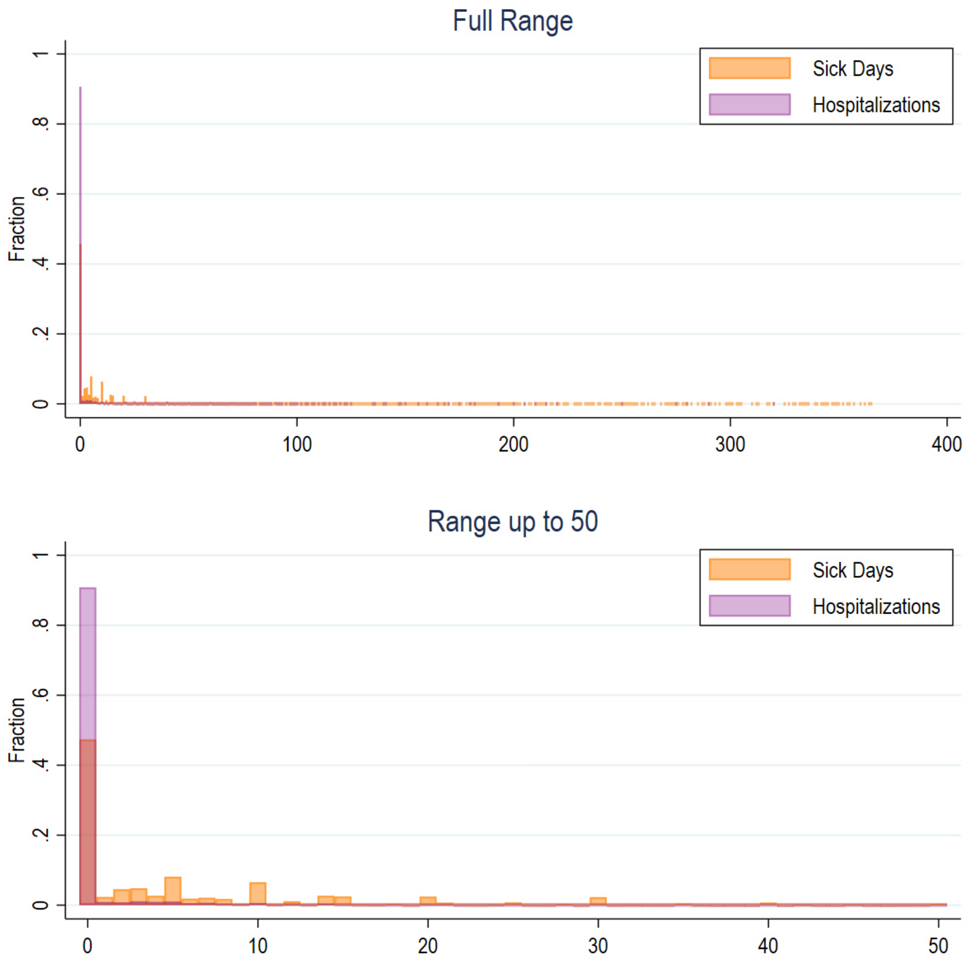
3.3.1 Definition of Health Shocks

Variables The two main variables for the classification of health shocks are *sick days*[11], that is, the number days an employee is absent from work due to illness,[12] and *hospitalizations*, that is, the number of overnight stays in a hospital. The advantage of using these variables for our analysis is that they both capture health behavior related to the labor market. Both variables imply an incapacity to work, while hospitalizations additionally indicate the need for inpatient treatment and thus signal more serious health issues.

[10] The set of covariates is comprised of a dummy for being female, a quadratic in age, a dummy for being over 50 years old, a dummy for having children under the age of 6 in the household, a marriage dummy, three educational categories (primary, secondary, tertiary), the number of average sick days taken prior to two periods before the event timing, as well as full- and part-time work experience.

[11] Note that by definition, sick days are not recorded for the unemployed. Within the scope of our study, this is not a relevant limitation.

[12] In most cases, a sick employee has to receive a doctor's certificate to notify the employer of his or her absence and verify the medical status. Thus, the information is very important to the sick person, making it very likely that it is accurately reported in the survey.



Note: Own calculations based on SOEP v35. Shows histograms of sick days and hospitalizations pooled for all years. Lower panel restricts the range to less than 51 sick days or hospitalizations.

Figure 3.2: Histograms of sick days and hospitalizations

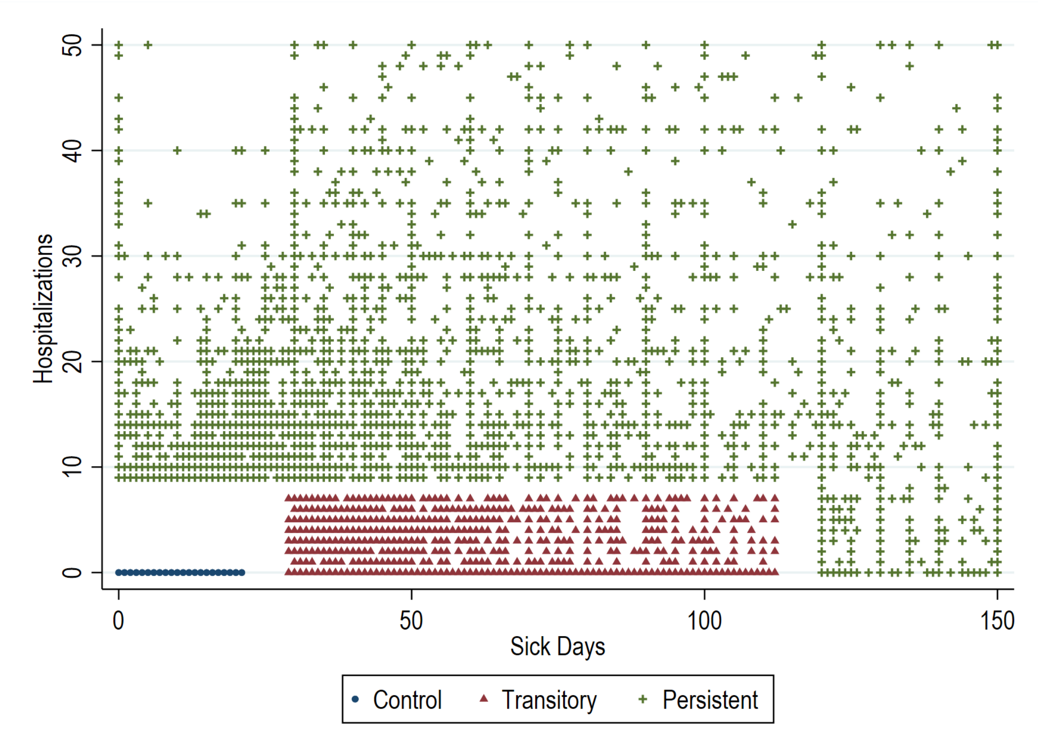
Comparing administrative statistics compiled by the Institute for Employment Research (IAB), average sick days in 1993 were at 12.3, decreased to 8.1 in 2007, and then rose again to 10.6 in 2017 (Wanger et al., 2019). Figure 3.16 in the Appendix shows that the SOEP data track these administrative trends well. Hospitalizations have gone down over time, especially since 2002, the year in which a strict reform on the maximum billable days in the hospital was introduced (see Appendix 3.9.3).

Clustering procedure The construction of health shock indicators based on the information on sick days and hospitalizations poses a key challenge: it is unclear how many distinct groups of health statuses one should differentiate and where the boundaries between these groups, in terms of sick days and hospitalizations, lie. Previously, the literature on health shocks distinguished mainly between healthy individuals and those experiencing *any* kind of health shock. This was true regardless of whether shock measures based on self-assessed health (García-Gómez, 2011; García-Gómez and Nicolás, 2006), disease diagnoses (Fadlon and Nielsen, 2021), or data on hospitalizations (Dobkin et al., 2018; Schurer, 2017; García-Gómez et al., 2013) were used. However, in structural work, Britton and French (2020) and Blundell et al. (2021) distinguish between two types of health shocks, namely *transitory* and *persistent* ones. To assess how many groups we can distinguish based on sick days and hospitalizations, we plot histograms of both variables in Figure 3.2. Both variables are heavily skewed to the left, which comes as no surprise because as the literature documents, most of the variation in ill health is concentrated in just a few unhealthy individuals (Hosseini et al., 2021a). Accordingly, the majority of individuals in our sample have a low number of sick days and hospitalizations, indicating that the group of healthy individuals should be the largest. However, a binary distinction between good health and bad health does not take into account the long tail and, therefore, meaningful distinctions between someone having to leave work for six weeks or six months, for example. To draw meaningful distinctions within the group of sick individuals, we rely on features of the data: we allow a clustering algorithm to reveal where best to set the boundaries between groups. The theoretical underpinning coming from life-cycle earnings models (Blundell et al., 2008, 2016) gives the number of groups we are looking for. We want to identify three distinct groups: the healthy, the transitory shock, and the persistent shock group. Fortunately, the data also largely support this classification: in Figure 3.17 in the Appendix, we show the R^2 from OLS regressions of sick days or hospitalizations on group dummies derived from clustering and sequentially allowing for more and more groups. The figure indicates that only sparse gains with respect to R^2 can be made by introducing more than three groups, as, for example, the R^2 for sick days with three groups is at about 0.8. In our empirical implementation, we face an important trade-off between group size and statistical power: introducing many different groups threatens the accuracy of our effect measurements. Thus, distinguishing between more than three groups is not only theoretically but also empirically unappealing.[13]

To locate the boundaries between the three groups, we apply a two-step clustering procedure to sick days and hospitalizations. We make the *ex-ante* assumption that people spending at least one night in a hospital cannot be assigned to the control (healthy) group. We make this assumption to ensure that the healthy control group is not contaminated by sick individuals. The extremely skewed distribution of

[13]The R^2 of distinguishing between three groups based on hospitalizations is about 0.6. However, meaningful gains in the R^2 would only be achieved when distinguishing between eight groups or more, which is not feasible due to observation numbers.

hospitalizations suggests that this assumption is justified because such a small fraction of individuals (<10%) end up in the hospital (see Figure 3.2). In our two-step procedure, we first cluster with respect to sick days using k-means and differentiate three groups: the precursor group to the healthy/control group, and the two precursor groups to those having potentially experienced either a transitory or a persistent shock. Clustering based on k-means appears to be the appropriate tool for the classification based on sick days, as it reacts more sensitively to outliers and tends to form groups of unequal size, allowing one group (in our case the control group) to be particularly large. Second, we cluster based on non-zero hospitalizations using k-medians and distinguish two groups: the precursor groups to the transitory and persistent shock groups. Here, we choose to cluster with k-medians, which forms groups of more equal size, because we have already ruled out the control group and only want to distinguish between the precursor groups to the transitory and the persistent shock groups. Since we are only clustering in the long right tail of the distribution, this appears appropriate. We call these the precursor groups because we still need further criteria to determine whether these individuals will actually end up in one of the treatment groups.



Note: Shows observations after applying the classification derived from the clustering procedure and excluding observations that fall into the symmetric bands around the clustering thresholds. Many observations occupy the same points, which we do not represent graphically, i.e., the figure is not weighted. For example, the point at (0,0) represents 122,178 observations with zero sick days and zero hospitalizations. Source: SOEP v35.

Figure 3.3: Groups after clustering

To gain higher contrast between the groups and mitigate measurement error, we exclude observations that fall in symmetric bands (±4 sick days and ± 1 hospital night) around the clustering thresholds. Figure 3.3 shows the result of the entire procedure by showing an unweighted scatter plot of all the joint realizations of sick days and hospitalizations in our dataset, while Table 3.3 shows the classification ranges for each group. Note that because of the strict requirement that the control group may not have had any hospitalizations, we exclude observations with a small number of sick days and a comparably small number of hospitalizations.

The clustering procedure provides intuitive classifications. The lower threshold for transitory shocks is at 29 sick days, which is close to the six-week threshold for receiving sickness benefits and the discontinuation of employer-paid sick leave. Further, individuals spending up to seven nights in the hospital are also assigned to the transitory shock group. Nine hospitalizations, the threshold for persistent health shocks, is in line with the average duration of hospitalizations due to various serious illnesses such as cancer (7.6), diabetes (10.2), coronary diseases (7.7), and hypertension-related diseases (7.7) (Federal Statistical Office, 2017b).

The classification procedure leads to the exclusion of observations that experienced positive numbers of hospitalizations (1-7) and low numbers of sick days (0-21). These observations experienced, if anything, a mild health change. Thus, as a robustness check of whether excluding these observations is important for our results, we augment the transitory shock group with these otherwise excluded observations. The results for employment with this alternative classification is shown in Figure 3.23 in the Appendix. The figure shows that results do not meaningfully change due to the alternative classification.

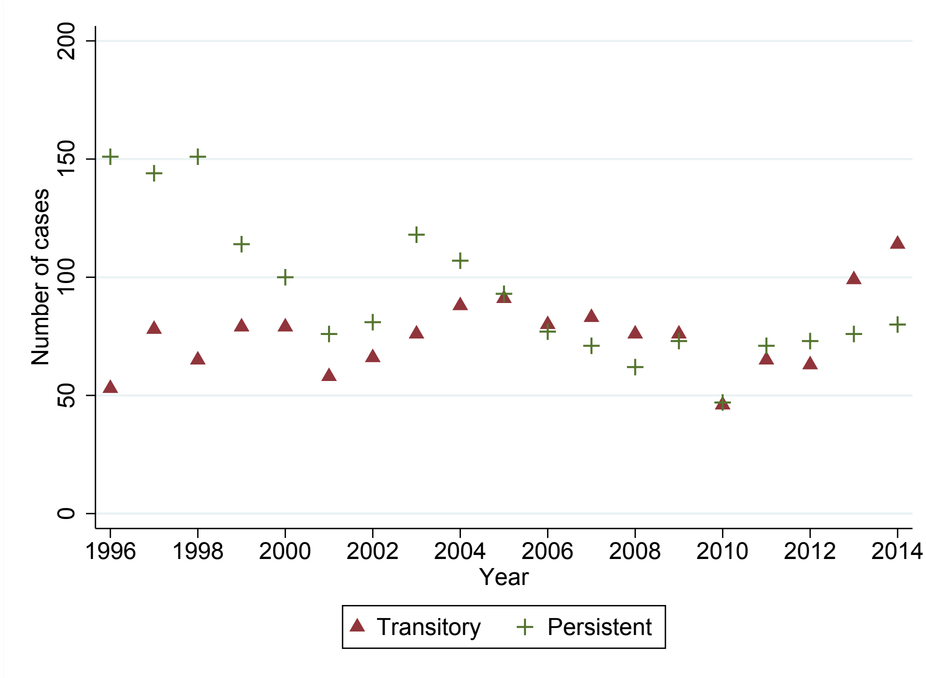
Table 3.3: Ranges for group classification

	Control Group	Transitory Shock Group	Persistent Shock Group
Sick Days	$[0, 21]$	$[29, 112]$	$[120, 365]$
Hospitalizations	$[0]$	$[0, 7]$	$[9, 365]$

Note: Displayed are the ranges in the distributions of sick days and hospitalizations for the control group, the transitory shock group, and the persistent shock group. Thresholds result from the clustering procedure. Source: SOEP v35.

Individual deviation condition Because health (slowly) declines over the life cycle and because health trajectories are bound to be subject to individual heterogeneity, a single, adverse health event—a long absence from the job or a long hospitalization—does not necessarily represent a health shock because these events might be part of a declining trajectory or individual heterogeneity. Shocks are, by definition, sudden deviations from the current trajectory. Thus, we exploit the panel dimension of our data and require that the health events that we classify as shocks are also major deviations from an individual's health trajectory.

Our implementation of this requirement is as follows: We measure the individual's medians and standard deviations of sick days and nights hospitalized for the duration of time that we observe a given person. Only if a health event exceeds the individual's median by more than two standard deviations for either of the measures do we assign this person to their respective shock group. As a robustness check, we lower the standard deviation condition to one standard deviation and show alternative results for employment in Figure 3.21 in the Appendix. The results do not substantially differ from those shown in the main analysis.



Note: Authors' calculations using SOEPv35. The figure only shows observed shocks from 1996 to 2014 because of the consecutive-observation requirement, that is, observing three years before and three years after the shock, which we employ in our main analysis. Further, because information on health and labor market status is retrospective in the SOEP, we need to omit another year.

Figure 3.4: Number of health shocks per year

Occurrences of health shocks With the shock definitions at hand, we verify that 1) we observe a sufficient number of shocks and that 2) the shocks are not strongly clustered in a given year. We show the number of shocks per year in Figure 3.4. To plot the figure, we impose the same sample restrictions as for our main analysis: we impose the requirement of consecutive observations for seven periods. This leaves us with 1,435 cases of transitory and 1,765 cases of persistent shocks. The figure shows that both shock types occur at frequent and similar rates over the observation period. Note that the figure only includes observed shocks from 1996 to 2014 because of the consecutive observation requirement and because information on health and labor market status is retrospective in the SOEP.

In Section 3.5.3.1, we show several benchmarks of our health shock definition against other objective and subjective measures of bad health, which all indicate that our classification strongly and positively covaries with these alternative measures. Since some of these analyses follow our main empirical framework, we present them after the framework is introduced.

3.3.2 Outcome Variables

Our outcome variables fall into two groups: individual-level outcomes, which concern the labor market, and household-level outcomes, which concern the partner and welfare of all members of the household.

Individual level outcomes We consider three outcomes:

1. labor force status, that is, being in regular employment[14];
2. yearly working hours adjusted for sick leave;
3. yearly personal labor income adjusted for sick leave;

The construction of the latter two measures requires us to adjust the existing measure of yearly hours provided in the equivalence file of the SOEP (Grabka, 2020). The existing measure of yearly hours is constructed by combining the SOEP's information on months spent in employment and the regular weekly working hours, but no attempt is made to correct for time spent away from work due to sickness. We use sick days to construct a corrected measure of yearly hours. We calculate

$$h_{i,t} = \tilde{h}_{i,t} - sickdays_{i,t} \times hpd_{i,t}\,, \tag{3.1}$$

where $\tilde{h}_{i,t}$ is the existing hours variable from the equivalence file, $sickdays_{i,t}$ is the number of sick days away from work, and $hpd_{i,t}$ is the average number of hours the individual works per day.[15] In Figure 3.19 in the Appendix we show the distributions of sick-leave-adjusted and unadjusted hours, which are fairly similar, yet the adjusted distribution is uniformly shifted to the left (lower hours).

[14] Regular employment is defined as dependent employment, regardless of the number of hours worked. Not in regular employment are apprentices, interns, or on-the-job trainees. We consider individuals as regularly employed in a given year if they meet the above conditions at any point of the year.

[15] We construct $hpd_{i,t}$ from the recorded hours of work per week. Our assumption is that the individual works five days per week.

To adjust personal labor income, we use the information on sick days, aggregating to months, and then using microsimulation,[16] calculate the replacement income. We then reduce the unadjusted yearly income by the difference between replacement and employment income for the duration of the sickness spell.

Household level outcomes We consider two outcomes:

1. partner labor income
2. household net income

Unlike the other outcomes, we do not need to adjust partner labor income or household net income. Partner labor income is reported directly by the individual's partner, preventing the problem that the individual misreports their partner's income. Household net income is compiled by adding up all income sources and subtracting taxes and social security contributions calculated by microsimulation as detailed in Grabka (2020). We needs-adjust this household net income with the modified OECD scale.

3.4 Empirical Strategy

Our empirical strategy relies on the comparison of individuals who experience a transitory or persistent health shock (treatment groups) with individuals who have not and will not experience a shock (never-treated/control group). We go through a two-step process to facilitate the analysis:

First, we match control units to treatment units one relative period prior to the shock based on a set of socio-demographic characteristics and other variables possibly affecting trend evolution. The matching procedure is based on the Mahalanobis distance (Mahalanobis, 1936) and allows for up to three matches per treated unit.[17] The set of covariates is comprised of a dummy for being female, a dummy for having German citizenship, a quadratic in age, a dummy for being over 50 years old, a dummy for having children under the age of 6 in the household, a marriage dummy, three educational categories (primary, secondary, tertiary), the number of average sick days taken prior to two periods before the event timing, as well as full- and part-time work experience.[18] Thus, we obtain two control groups, one for the transitory and one for the persistent treatment group.

[16] We calculate replacement rates for every year according to the sickness benefits framework of the German health insurance system.

[17] Let x and y be two vectors with observations on several variables. Then, $D_M(x,y) = \sqrt{(x-y)'CV^{-1}(x-y)}$ is the Mahalanobis distance of the two vectors, where CV is the covariance matrix associated with the variables in (x,y). We use the `psmatch2` package in Stata to implement the matching.

[18] Note that for the estimation of partner incomes we additionally include a dummy indicating whether the partner has German citizenship, a quadratic for the partner's age, a dummy indicating

Second, based on these matched groups, we perform event study analyses by running OLS regressions with individual and year fixed effects as well as a set of dummies for the pre- and post-shock relative periods and their interactions with the treatment dummy. The post-shock interaction dummies measure the average treatment effect of a health shock on the treated. The regression equations take the form

$$Y_{it} = \sum_{k=-3}^{3} \gamma_k P_{it}^k + \sum_{k=-3}^{3} \delta_k P_{it}^k \times T_i + \nu_i + \tau_t + \epsilon_{it}, \tag{3.2}$$

where Y_{it} is the outcome of interest for person i in year t, for example, employment or yearly hours, $\left\{P_{it}^k\right\}_{k=-3}^{3}$ is a set of relative period-dummies running from -3 to =3, but excluding $k = -1$,[19] with a shock occurring in period k=0 if the person is in a treatment group, T_i is the respective treatment group dummy, ν_i is an individual fixed effect, τ_t is the year dummy, and ϵ_{it} is an idiosyncratic error. The coefficients of interest are the δ_k-coefficients, which represent the period-specific average treatment effects on the treated (ATT) for the respective treatment group.

Our central identifying assumption is the common trend assumption, which states that in the absence of the shock, the evolution of the outcome for the treatment group would have been the same as for the control group.[20] When this assumption holds, we can interpret the differences in outcomes between control and treatment after the shock as causal effects. A common way to confirm that the assumption holds is to show pre-trends, that is, outcome differences prior to the treatment. We show two pre-treatment periods in all of our analyses.

The event study design addresses the concern of reverse causality—the possibility that a labor market shock (e.g., job separation) can cause health problems (Haan and Myck, 2009; Britton and French, 2020). Further, it rules out that other contemporaneous confounders affect the outcome.

3.5 Results

We show the results of our analysis by plotting the coefficients δ_k for employment, working hours, labor income, partner income, and household net income. For each outcome, we show two plots: one for the treatment coefficients of the transitory shock group, and an analogous one for the persistent shock group. Each coefficient can be interpreted directly as the average period-specific treatment effect, that is, the

whether the partner is employed, and the partner's full-time and part-time experience as further covariates for the matching procedure.

[19] We exclude this period to avoid perfect multicollinearity in the dummy set.

[20] See, for example, Sun and Abraham (2021) and Goodman-Bacon (2021) for recent expositions on the topic.

average difference between the respective treatment and control group in a given period.

3.5.1 Main Analyses

Figure 3.5 shows the effect of experiencing a transitory or a persistent shock on the five outcome variables.

Employment Panel A displays the effects of either type of shock on employment. Transitory health shocks have small, negative, and statistically significant employment effects as the employment rate is reduced by 4 to 5 percentage points (pp) for the treated. Persistent health shocks have much larger employment effects: after persistent health shocks the employment rate of the treated drops by roughly 16 pp.

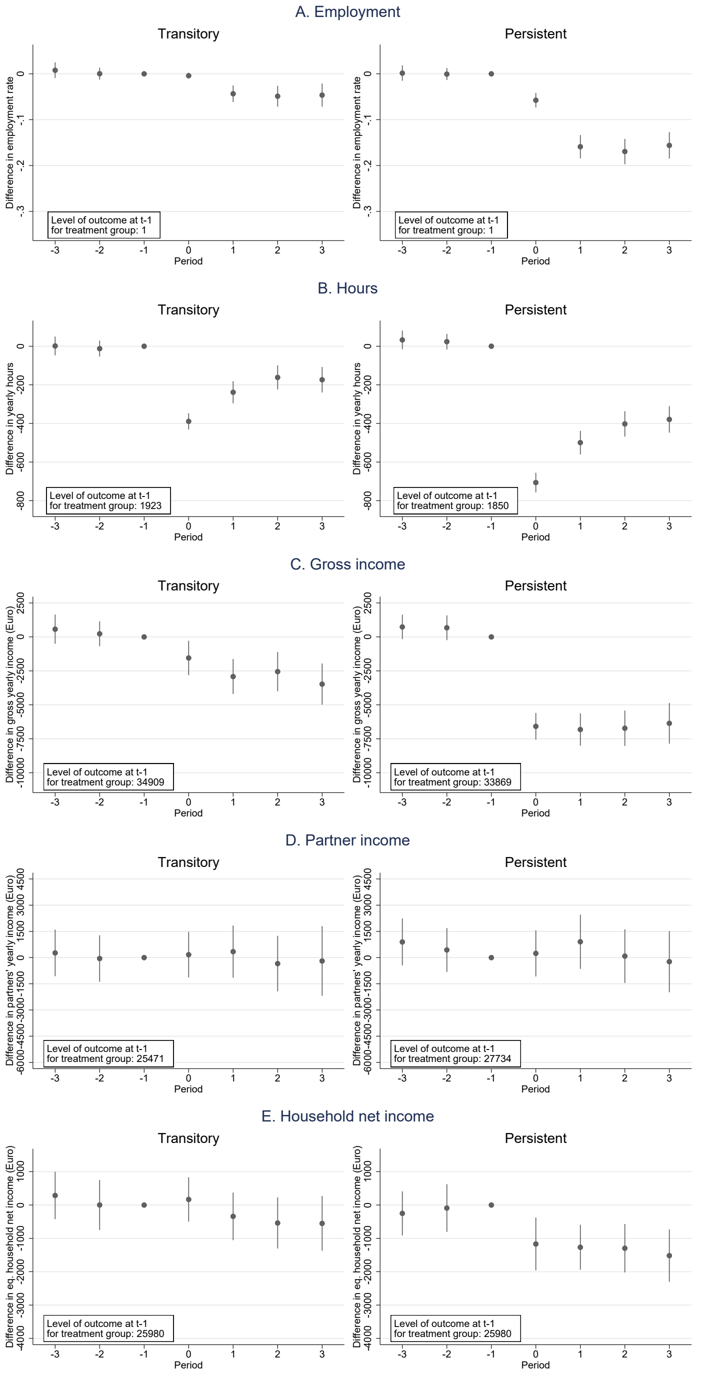
Note that in Figure 3.5, we only distinguish between employed individuals and those not working. Thus, we do not examine whether people register as unemployed or exit the labor force. In Figure 3.20 in the Appendix, we additionally display ATTs of health shocks on a dummy for registered unemployment. These effects are of much smaller magnitude than the effects on employment. This indicates that the majority of the individuals who drop out of employment after experiencing a health shock exit the labor force rather than registering as unemployed.

Yearly hours Panel B shows the effect of experiencing a transitory or a persistent shock on yearly hours.[21] Note that we keep the sample constant, meaning that those who are not employed remain in the sample and work zero hours .[22] Transitory shocks reduce hours in the period of the shock by around 400, which amounts roughly 2.5 months of full-time work. The drop is followed by a quick but only partial recovery: three periods after the shock, yearly hours of the affected are still around 200 hours less compared to the control group. Following persistent shocks, hours drop by around 700 in the period of the shock, which amounts to more than four months of full-time work. In the following periods, hours partially recover, but three periods after the shock the average treatment effect still indicates a reduction of about 400 hours, which translates to around 7.7 hours per week.

In considering these estimates, it remains unclear whether this effect is completely driven by those who drop out of employment and consequently have work zero hours or whether some of the affected individuals remain employed but reduce their hours and switch into part-time arrangements. Figure 3.6 sheds some light on this matter by displaying transition matrices of hours categories between periods -1 and 3 for all treatment and control groups. Transitions between the following

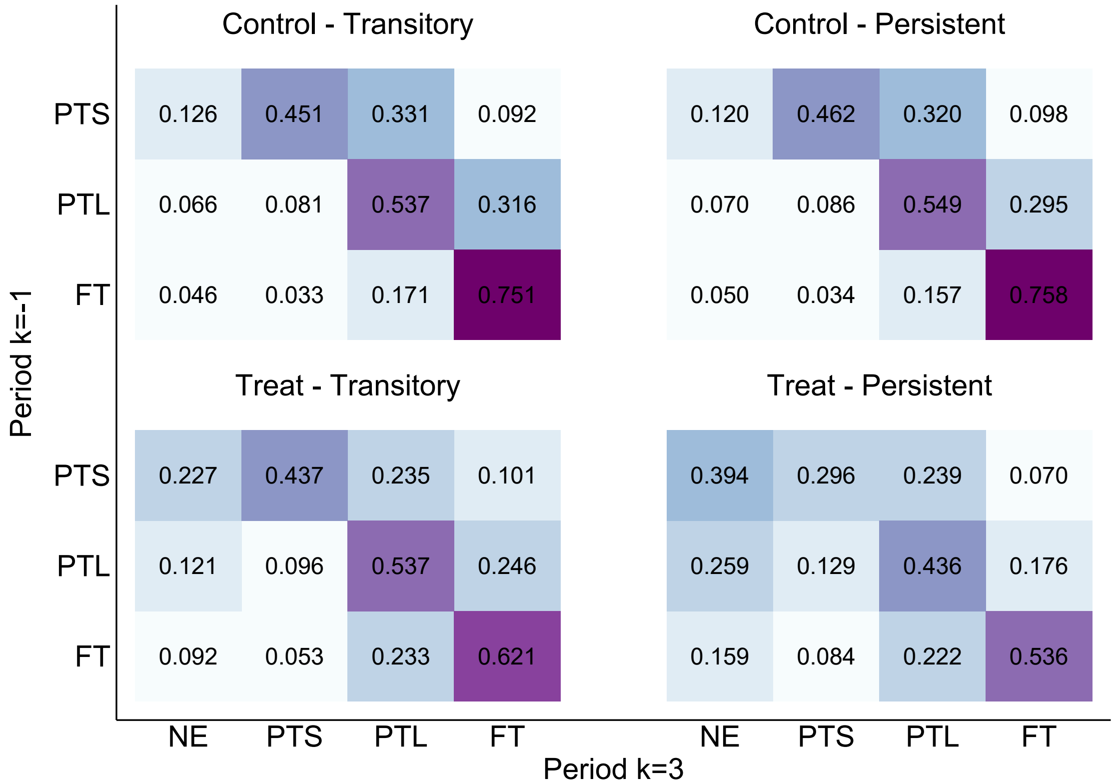
[21] We adjust yearly hours to be consistent with the number of days registered sick. See Section 3.3.

[22] In Figure 3.24 in the Appendix, we also consider a smaller sample consisting of individuals who remained employed over all seven periods to isolate the intensive margin effect.



Note: Shows period-specific coefficients according to Eq.(3.2). Bars give robust 99% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 1,419 treated, 2,870 control; persistent shock: 1,734 treated, 3,072 control. Source: SOEP v35.

Figure 3.5: Main results: ATTs of transitory and persistent health shocks



Note: Shows transition probabilities for yearly hours categories for treatment and control group moving from period -1 to 3. NE means not employed, PTS means less than 1,000 hours per year, PTL means between 1,000 and 2,000 hours, FT is 2,000 or more hours. Source: SOEP v35.

Figure 3.6: Transition matrices of hours categories from period -1 to 3

categories are shown: 1) full-time (FT), defined as equal to or more than 2,000 hours, 2) long part-time (PTL), defined as from 1000 to 2000 hours, 3) short part-time (PTS), defined as less than 1,000 hours, 4) not employed (NE), that is, zero hours.

The central result of the transition matrices is that after both a transitory and a persistent shock, individuals remain in full-time employment less often and more often switch either to being not employed or into one of the part-time categories. In line with the findings from Panel B of Figure 3.5, transitions into the NE category occur more frequently after persistent shocks: for example, 39% of those who experienced a persistent shock and worked in PTS prior to the shock are out of employment three periods after the shock. In contrast, only 12% of the control group switch from PTS to NE. Further, after persistent health shocks, the probability to move from FT to NE more than triples (16% vs. 5%). Transitions from FT into one of the part-time categories are also more frequent after experiencing either type of health shock.

Overall, the transition matrices round out the complete picture of how health shocks affect employment: transitory and persistent health shocks cause some of the affected individuals to reduce their working hours and switch into part-time arrangements. However, in comparison, the extensive margin effect of persistent health shocks is much larger and, as expected, increases with lower labor market attachment, that is, for those already working part-time.

Individual labor income Panel C of Figure 3.5 shows the effect of transitory or persistent shocks on yearly gross labor income. The sample remains constant, meaning those who are not employed earn €0 in our data.[23] In the period of the shock, transitory shocks reduce gross income by about €1,500. This effect increases in the following periods and amounts to roughly €3,500 in the third period (10% drop from baseline). Persistent shocks have larger effects on gross income: in all post-shock periods, gross income of those who experienced a persistent health shock is reduced by about €6,500, that is, a 19% drop from the baseline.

Partner labor income Panel D of Figure 3.5 shows the effect of both shock types on partner labor income, allowing us to investigate potential added-worker effects. These occur when one partner increases their labor supply and, thus, income to compensate for the income loss due the other partner's health shock. We consider all partners for whom the relationship status remained unchanged during all seven periods, and we do not condition on their employment status at the time of the shock. Unlike our results for the individual-level outcomes, we find no significant effects for either of the two shock types. The point estimates are close to zero, but confidence intervals are too large to strongly assert a null effect. Nonetheless, added worker effects are generally not found to be relevant in the related literature (Dobkin et al., 2018; De Nardi et al., 2021).

Household net income Panel E of Figure 3.5 shows the effect of experiencing transitory or persistent shocks on household net income. Estimating the effects on household net income allows us to calculate a pass-through coefficient of the shocks: dividing the effect on household net income by the effect on individual gross labor income gives a pass-through coefficient and is, therefore, a measure of insurance provided by the household and the state.

Transitory shocks have no significant effect on household net incomes. In contrast, persistent health shocks reduce household net income by about €1,150 in the period of the shock. This effect increases to €1,500 three periods after the shock. Since we find no significant effect of transitory shocks, we will not calculate a pass-through

[23] In Figure 3.24 in the Appendix, we also consider a smaller sample consisting only of individuals who remained employed over all seven periods to isolate the income effect of those who remain employed.

coefficient for transitory shocks. For persistent shocks, the effect on gross labor income was a decrease of about €6,500 in the third period, while the decrease in net household income amounted to roughly €1,500. Thus, just about 25% of the gross shock passes through to net household income.

Finally, and considering all outcomes, we do not find significant pre-shock trend deviations in any treatment group, increasing our confidence that our identifying assumptions hold.

3.5.2 Effect Heterogeneity

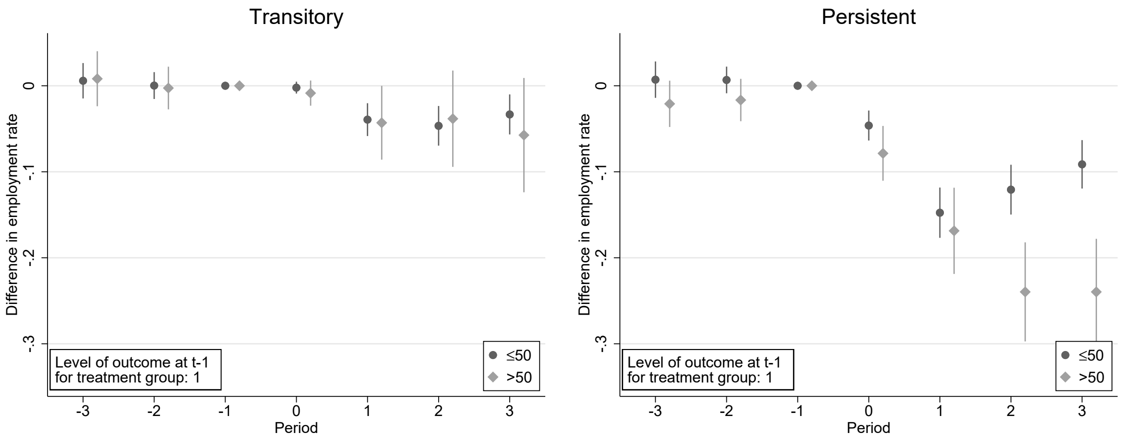
Grossman (1972) was one of the first to model health as depreciating with age. However, not only health itself, but also individuals' labor market reactions to adverse health events differ substantially depending on their age (Blundell et al., 2021; Dobkin et al., 2018). While we have shown average treatment effects for the whole sample of employed individuals so far, examining effect heterogeneity is crucial to understand which demographic groups may be more or less exposed to health shocks. Age is the most obvious margin of heterogeneity. Thus, we repeat our main analysis after splitting the sample into those 50 years of age or younger, which we call the younger group, and those over 50 years of age, which we call the older group.[24] Thereafter, we consider further heterogeneity margins and present sample splits between primary and higher education as well as managers and non-managers. Since we find the strongest effects on the extensive margin of labor supply, we only report the heterogeneity analyses with respect to employment.[25] Additionally, we conducted heterogeneity analyses between men and women, singles and partnered individuals, private and public sector employees as well as between employees at small firms and at large firms. Since we do not find significant differences along any of these dimensions, we report the results in the Appendix in figures 3.25, 3.26, 3.27 and 3.28.

Age Figure 3.7 shows the effects of transitory and persistent health shocks on employment separated into the younger and the older group. There are no significant differences between the effects of transitory shocks on the younger and the older group. Point estimates are very similar and resemble the results for the whole sample. Persistent health shocks have similar effects on both age groups in periods 0 and 1 but diverge thereafter. Both age groups exhibit a reduction of around 15 pp in employment in the first period. However, in the second and third period, the younger group partially recovers, and the effect is only -10 pp in the third period. In

[24]We fix this age heterogeneity in period -1.

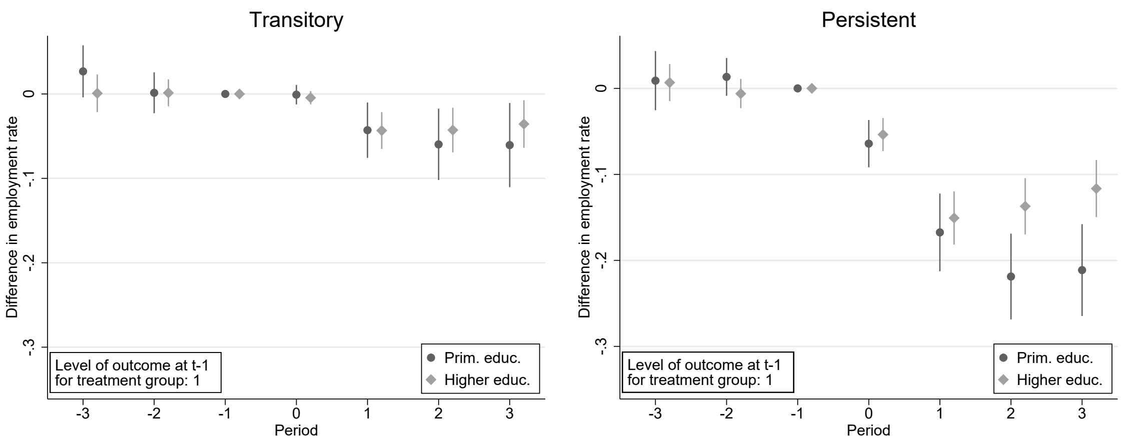
[25]Further analyses with respect to the other outcomes and split by age are reported in Figure 3.22 in the Appendix.

contrast, the effect becomes larger for the older group and amounts to -25 pp in the third period.



Note: Shows period-specific coefficients according to Eq.(3.2) for the treated split by age (≤ 50 vs. > 50). Bars give robust 99% confidence intervals of the respective coefficients. Transitory shock figure: Individuals for Treat(≤50) are 1,020, for Treat(>50) are 399, for Control(≤50) are 2,138, for Control(>50) are 732. Persistent shock figure: Individuals for Treat(≤50) are 1,132 , for Treat(>50) are 602, for Control(≤50) are 2,173, for Control(>50) are 899. Source: SOEP v35.

Figure 3.7: Effect heterogeneity: age



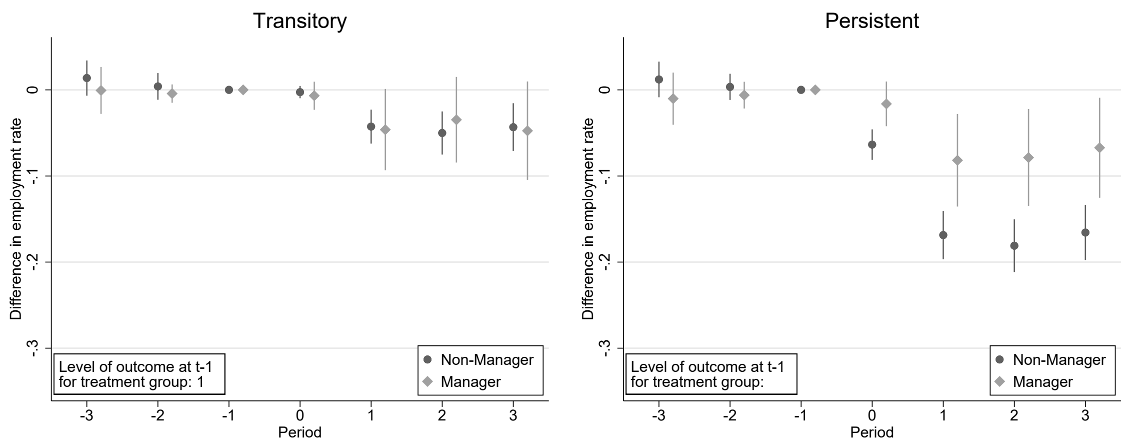
Note: Shows period-specific coefficients according to Eq. (3.2) for the treated split by education (primary vs. higher). Bars give robust 99% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 907 treated/higher education, 512 treated/primary education, 1,904 control/higher education, 946 control/primary education. Persistent shock: 1,065 treated/higher education, 668 treated/primary education, 2,069 control/higher education, 996 control/primary education. Source: SOEP v35.

Figure 3.8: Effect heterogeneity: education

Education Several papers have documented differing health behaviors by education, both before and after adverse health events (Blundell et al., 2021; Britton and

French, 2020). For both transitory and persistent shocks, effects are very similar for both education groups one period after the shock. However, in the long run, those with higher education make a quicker recovery if they experienced a persistent shock. For transitory shocks, the effects remain similar even in the third period. We suspect that the more educated recover more quickly from persistent shocks because they find better ways to manage and treat their illnesses, as other papers document (Blundell et al., 2021).

Managers Managers have an exceptional position in organizational hierarchies and primarily complete tasks that are non-routine and cognitively demanding compared to other workers, who hold positions that are more physically demanding. Thus, three effects may be at play leading to potential effect heterogeneity for this group: First, managers have stronger incentive to work, because of higher wages. Second, managers are essential to a firm and are difficult to replace so that there are strong incentives for the firm to have the manager return after a health shock. Third, since managers perform tasks that are less physically demanding, their recovery and re-entry into their job may be more easily facilitated.



Note: Shows period-specific coefficients according to Eq. (3.2) for treated split by being a manager or not being a manager. Bars give robust 99% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 1,246 treated/non-manager, 173 treated/manager, 2,384 control/non-manager, 466 control/manager. Persistent shock: 1,504 treated/non-manager, 229 treated/manager, 2,513 control/non-manager, 552 control/manager. Source: SOEP v35.

Figure 3.9: Effect heterogeneity: managers

Figure 3.9 shows that after transitory shocks employment differences between managers and non-managers are small and statistically insignificant. The dynamics after persistent shocks exhibit more heterogeneity: in the period of the shock, managers hardly show any change in employment, while non-managers show an immediate employment reduction of close to 10 pp. In the following periods, the effects diverge even more. Managers' employment drops by close to 10 pp three periods after the shock, while non-managers' employment falls by about 20 pp compared to the control group. From these results, we find some support for the hypotheses stated above, as managers appear to exhibit stronger labor market attachment even after a persistent health shock.

3.5.3 Robustness Checks

3.5.3.1 Validation of the Health Shock Classification

Because of the novelty of our health shock definitions, it is crucial to demonstrate that they are strongly linked with other health indicators used in the literature, such as disease diagnoses and self-assessed health. Thus, we show partial correlations of both shock indicators with various disease diagnoses variables. We would expect that both types of shocks are correlated with serious and chronic diseases and that persistent shocks are more strongly associated with these chronic diagnoses compared to transitory shocks.[26]

Table 3.4 shows coefficients from OLS regressions of a dummy equal to 1 if one is ever diagnosed with a chronic illness, on a dummy of ever experiencing a transitory or persistent health shock. The regressions thus purposely disregard the panel dimension of the data and yield basic associations between these variables. The results are congruent with our expectations: associations are positive and statistically significant for both shocks and are stronger for persistent shocks.

Similarly, we would expect that health shocks are associated with reductions in self-assessed health. Figure 3.10 shows the effects of both types of shocks on the health satisfaction measure in our data. In both age groups, health satisfaction drops sharply after either type of shock, but more so after persistent health shocks. After transitory health shocks, health satisfaction fully recovers, and no statistically significant difference remains three periods after the shock. Following persistent shocks, we only observe a partial recovery as health satisfaction remains depressed three periods after the shock.

3.5.3.2 Accounting for Heterogeneous Dynamic Treatment Effects

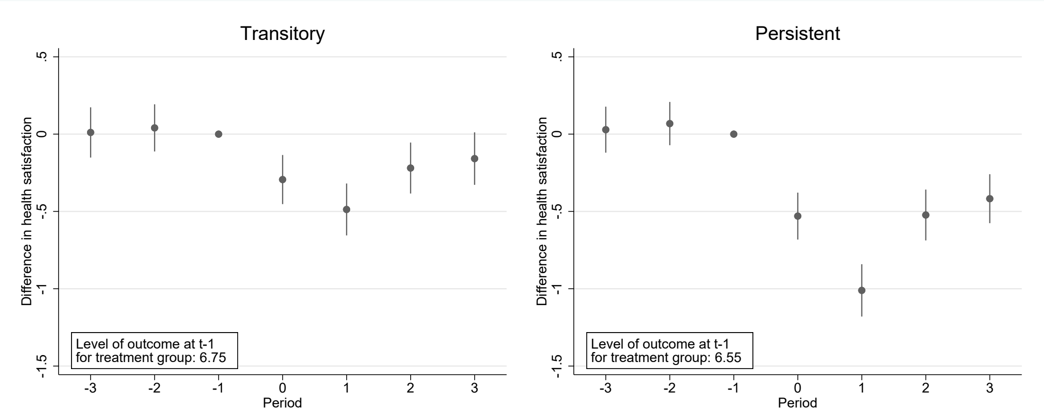
The dynamic treatment effect literature has shown the importance of accounting for heterogeneous treatment effects along cohorts in event study designs (Sun and Abraham, 2021). Cohorts are groups defined by the calendar year that corresponds to the first relative time period with respect to treatment. ATTs in a simple two-way fixed effects design are influenced by the proportion of each cohort in the dataset and, thus, may not give the true overall ATT. Rather, the ATT from two-way fixed effects may correspond to some other linear combination of the cohort-specific ATTs.

[26]Note that chronic disease diagnoses are only surveyed biennially starting in 2009. The correlation measures in Table 3.4 thus only refer to this part of the observation period.

Table 3.4: Validation of shock measures: chronic illnesses

	Transitory shock	Persistent shock
Heart disease	0.022	0.069
	(0.002)	(0.003)
Cancer	0.016	0.057
	(0.002)	(0.002)
Stroke	0.007	0.020
	(0.001)	(0.001)
Obs.	21,570	22,091

Note: The table shows coefficients from an OLS regression of a dummy equal to 1 if one is ever diagnosed with a chronic illness, on a dummy of ever experiencing a health shock. Robust standard errors are shown in parentheses. Obs. refers to observations used in the OLS regression. Source: SOEP v35, own calculations.



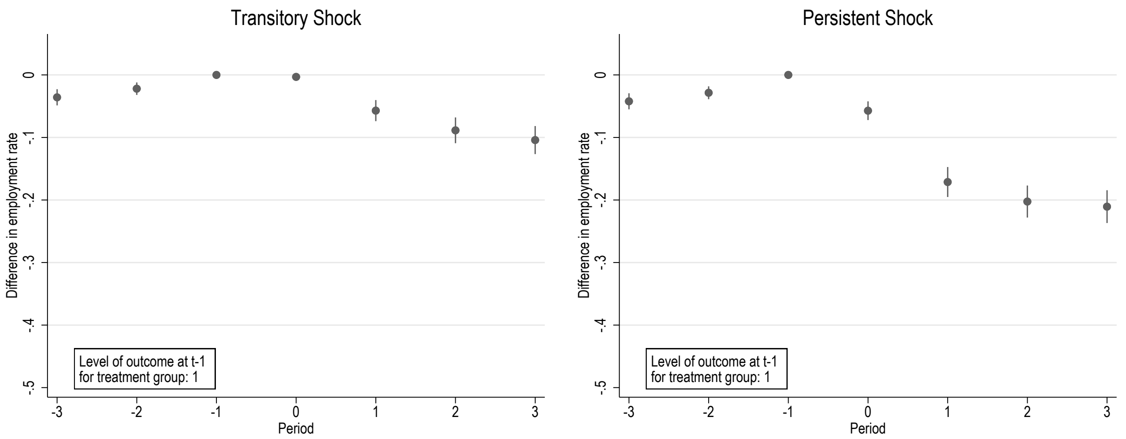
Note: The figure shows period-specific coefficients according to Eq.(3.2) for the treated. Bars give robust 99% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 1,419 treated, 2,870 control. Persistent shock: 1,734 treated, 3,072 control. Source: SOEP v35.

Figure 3.10: Validation of shock measures: health satisfaction

Sun and Abraham (2021) propose an estimator to address this problem, which we use to check the robustness of our results. The `eventstudyinteract` package for Stata provided by Sun enables us to implement the analysis.[27] The package produces estimates of the differences in outcomes along relative time periods compared to never-treated control units. We show these treatment effects with respect to employment in Figure 3.11.

Both figures show qualitatively equivalent trends to those shown in Figure 3.7, while effect sizes after the shocks are also of similar magnitude—for persistent shocks even somewhat stronger than in the main specification. Hence, we can conclude

[27]The package is available at https://economics.mit.edu/grad/lsun20/stata.



Note: Shows period-specific treatment effects estimated using the eventstudyinteract-package. Bars give robust 99% confidence intervals. Number of observations: Transitory shock: 1,419 treated, 2,870 controls. Persistent shock: 1,734 treated, 3,072 controls. Source: SOEP v35.

Figure 3.11: Employment treatment effects—Sun-Abraham estimator (2020)

that our main estimates do not suffer from substantial bias due to heterogeneity in dynamic treatment effects.

3.5.3.3 Intention to Return to Work

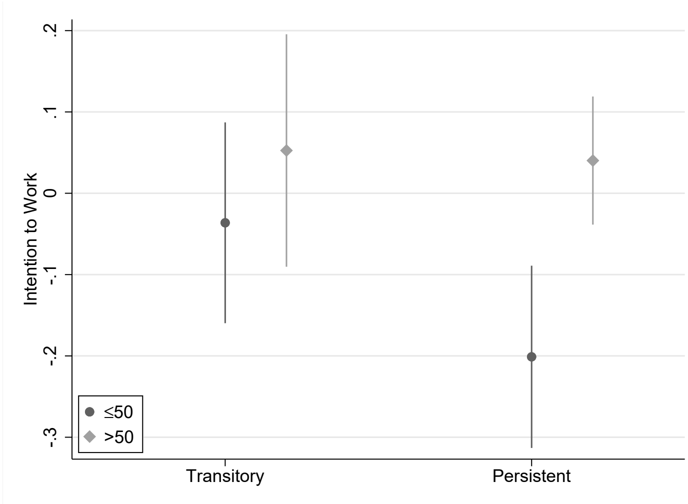
In our main analysis, we find the largest adjustments to health shocks on the extensive margin (employment). To understand the labor market dynamics for individuals after a health shock, we need to investigate whether they intend but fail to reenter the labor market or whether they refrain from reentering in the first place. If the former were the case, there would be a strong public policy case for fostering these individuals' reentry to the labor market by helping to increase their probability of finding a job that is accommodating with respect to both their qualifications and their physical capacity.[28]

To determine whether individuals who dropped out of the labor market after a health shock are actively looking for and cannot find a job or are simply not looking for one, we use a question on the intention to work included in the SOEP. The SOEP asks those not in employment whether they are likely to obtain or resume employment in the future, with the answers falling into four categories: 1) "No, definitely not", 2) "Probably not", 3) "Probably", and 4) "Yes, definitely". For ease of interpretation, we recode these categories into a dichotomous variable, in which categories 1 and 2 are coded as a zero and categories 3 and 4 are coded as a 1.

[28]The German system has made some progress in this direction: when employees return to their old job after an illness, they can file to reenter at reduced capacity and then progressively increase their workload up to full capacity ("Wiedereingliederung", as per Sozialgesetzbuch IX).

This is the intention-to-work indicator that we use in subsequent analyses. We only consider individuals who are out of employment after period 0. As in the main analysis, we distinguish between people from the control group, who have not actually experienced a shock, and those who experienced either a transitory or a persistent shock. Further, we distinguish between those over the age of 50 and those 50 years of age or younger.

Because the number of individuals that are out of work in our treatment groups for every post-shock period is fairly low, we pool all observations post-shock. We report the predictions of the intention-to-work indicator after running OLS regressions for the control and treatment groups by age group. The OLS regressions were separated by age group. Figure 3.12 shows the results.



Note: Shows post-treatment coefficient differences for treated and control units either being older than or up to 50 years of age. Bars give robust 99% confidence intervals of the respective coefficient differences. These coefficient differences are based on two regressions with 2022 individuals up to age 50 and 2006 individuals over the age of 50. Source: SOEP v35.

Figure 3.12: Intention-to-work coefficients after a shock

Individuals up to age 50 in the control group generally intend to return to work, with the mean of the intention-to-work indicator at about 0.87. For those in the control group over the age of 50, the mean is much lower, at about 0.13. Accordingly, when individuals over 50 fall out of employment—regardless of whether this was caused by a health shock or not—there is a very low tendency to seek re-employment overall.

The figure shows that neither age group exhibits a significant difference from the control group in its intention to work after a transitory health shock. The same holds for the older group, even after a persistent health shock. In contrast, the younger group leaving employment after a persistent shock, indicates a significantly lower intention to reenter employment: the coefficient is at -0.20, that is, a 20 pp drop in intention to work compared to the control group.

3.6 Discussion

The main results from our analysis can be summarized as follows: First, there are large extensive margin effects after health shocks, and there exists relevant heterogeneity with respect to age, education, and occupational status. The employment effects are milder for young employees, the more educated, and managers. When we examine the intention to re-join the labor force after a health-related spell of non-activity, young individuals affected by transitory shocks appear to be willing to return to work quickly, while those who experienced persistent health shocks largely do not intend to return to the labor market.

Second, intensive margin adjustments are strong in the period of the shock, but hours partially recover in the periods after the shock. This indicates that most individuals return to their previous working hours choices, while some reduce them and move into a part-time arrangement.

Third, both on the individual and on the household level, incomes decrease after a persistent shock. However, transitory shocks have no significant effect on household net income. Pass-through of a persistent shock is only 25% of the direct effect on individual labor income, indicating significant insurance provided by state and family.

Benchmarking to related literature We can benchmark our results to other empirical studies on labor market effects of health shocks. Similar to our approach, studies like Fadlon and Nielsen (2021), Dobkin et al. (2018) or García-Gómez et al. (2013) estimate reduced forms with differing measures of health events and find differing severity of effects on labor market outcomes.[29]

Fadlon and Nielsen (2021) analyze the effects of fatal and non-fatal health shocks such as strokes and heart attacks on household labor supply and income in Denmark. For fatal health shocks, they find that widows increase their labor supply and obtain higher individual earnings after the death of their husbands. For non-fatal health shocks, they find that the labor force participation of sick individuals drops by 12 pp and annual earnings decrease by around €4,700, that is, an 18% drop from the baseline.[30] Fadlon and Nielsen (2021) do not distinguish between transitory and persistent health shocks, and yet their results for non-fatal shocks can be compared to our shock definitions: the 18% drop in annual earnings is congruent with the 19% drop we report for persistent health shocks. Like us, they find no adjustment of partner labor supply or income in the case of non-fatal shocks. Shock pass-through to household net income is 50%, owing in part to the insurance mechanisms and the

[29] Further studies estimating reduced-form models of health events are Meyer and Mok (2019) and Smith (2004).

[30] Fadlon and Nielsen (2021) estimate a drop of 35,467 Danish crowns, which in September 2021 translates to around €4,700.

safety net provided by the Danish tax and transfer system, which strongly parallels the German one.[31]

Dobkin et al. (2018) study the economic consequences of hospital admissions by adults aged 50 to 59 in the United States.[32] Three years after a hospital admission, they find a drop in employment of 11 pp and reduced labor earnings of around €9,300, a 24% drop from the baseline.[33] Dobkin et al. (2018) do not find significant effects on spousal earnings. Further, household net income does not significantly change after the shock, as estimates are imprecise. However, the point estimates for net household income indicate a drop of about €6,900 per year. About 10% of the raw impact of the shock on earnings is compensated for by social security disability insurance payments (€745).[34]

García-Gómez et al. (2013) analyze the effect of acute hospitalizations on employment and labor income in the Netherlands. They find that employment drops by 7 pp two years after the shock, while personal post-tax and transfer income is reduced by €1,000.[35] Similar to us, they also find slightly stronger effects on employment for individuals older than 50 (1 pp more than the average effect.). Effects on household net income—a drop of about €1,500—are larger than those on individual labor income because the spouse's probability to stay employed is reduced by 1.5 pp three years after the shock.

The magnitude of the estimated employment effects differs among the aforementioned studies because they use different concepts of health shocks and examine different countries and institutional settings. The employment drop we estimate after persistent shocks is around 16 pp, making it slightly larger than the effects in papers above. Our estimated effects on labor income are strikingly close to the effects shown in Fadlon and Nielsen (2021) for Denmark. This appears intuitive, as Denmark exhibits a fairly similar setting in terms of labor market conditions, social security, and health insurance systems.

[31]The Danish health insurance system is fairly comparable to its German counterpart, although slightly less generous. Health insurance is funded through municipal income taxation at a flat rate, which, to the contributor to the system, is like paying into the German public health insurance system. Employers pay wage continuation for 30 days, as in Germany, and employees with prolonged absences receive sickness benefits thereafter for up to 22 weeks within a year. Sickness benefits are slightly less generous than in Germany. For further details, see Online Appendix E of Fadlon and Nielsen (2021).

[32]The authors also report results for other age groups, such as those aged 60 to 64. However, we choose this age group as it compares well with our definition of the older group.

[33]Dobkin et al. (2018) report reduced earnings of $11,071, which in September 2021 was equivalent to around €9,300.

[34]Dobkin et al. (2018) report reduced household net income of $8,161, which in September 2021 was equal to around €6,900. The authors find social security disability insurance payments of $881, which amounts to €745. The implied pass-through of the gross income shock to net is 0.73, and thus much larger than in the German or Danish context.

[35]As the average effective tax and contribution rate in the Netherlands is around 38%, we can make a ballpark estimate that the effect on gross income is around €1,600.

Evaluating the results Overall, the magnitude of the effects of health shocks depends on country-specific particularities of the social security system. Our findings of large negative effects of health shocks on employment in Germany are worrying, especially against the backdrop of Germany's aging society and its comprehensive social security system. Old-age pensions and other social security benefits are financed through contributions from the actively working population. With large demographic groups such as the baby-boomers nearing or past retirement age, the public pension system is facing substantial financial challenges (Rürup, 2002). To ensure the sustainability of the system, more contributions—whether through a larger workforce, a more productive workforce, or a workforce that retires later—would be crucial (Buslei et al., 2019).

In this respect, our results on employment effects should raise public concern, as people, even in the highly productive age range of 18-50, tend to drop out of the labor force after experiencing a persistent health shock. Additionally, a substantial fraction of these younger individuals does not seek to return to work and will not contribute to the social security system in the most productive phase of their lives. While we document individuals' low intentions to re-join the workforce, the ultimate reasons *why* they do not look for a subsequent job remain unknown. The most obvious reason is diminished capacity to work. Further, one might suspect that individuals judge their prospects of finding an appropriate and well-paid job to be low. In both cases, there is room for a public policy response. While the former case calls for improved rehabilitation measures, the latter points towards a need for retraining and more efficient matching of individuals recovering from health shocks with jobs that suit their capacity for work (Mehnert et al., 2013; Rick et al., 2012). Our results also suggest substantial potential for improved reintegration of the formerly sick into the labor market. Compared to the large effects on the extensive employment margin, relatively few affected individuals move into part-time arrangements. Hence, at least for many, labor supply seems to be a binary decision: either full-time work or none at all. In contrast, working capacity ranges from a complete inability to carry out tasks in the workplace to only minor impairments that slightly limiting working time. Thus, the rigidity of the labor market with respect to working schedules may lead to unused productive capacity. Pencavel (2016) reviews the reasons for desired and actual hours mismatch among workers, stating that these mismatches may stem from employers' hours mandates, which in turn reflect the firms' price and production environment. One possibility why firms demand full-time hours is that part-time work implies more start-up or quasi-fixed costs (more office space, transaction costs when sharing tasks), while another issue may be that employers require the joint presence of several inputs (e.g., two skill-types of labor) and therefore restrict workers' hours choices (Deardorff and Stafford, 1976). Hence, the binary employment decision we observe for sick individuals may be deeply rooted in the production environment of the firm, which makes the flexibilization of work infeasible.

Activating the unused working capacity of the formerly sick, which helps to ensure the sustainability of the social security system, may hinge not just on the intention to work on the supply side, but also on the incentives of the demand side to offer working arrangements in line with employees' working capacity. Further, an intensified information asymmetry problem between job applicant and potential employer may exist since it may be very difficult for the employer to assess the type of the formerly sick job applicant. Thus, it may be prudent for policymakers to consider how these incentives on the demand side and signaling problems can be influenced. Exploring whether reactivation policies should focus on the demand, or the supply side is a promising avenue for future research.

Our results with respect to the effects of health shocks on income imply long-lasting earnings penalties for those experiencing persistent shocks. Encouragingly, these penalties do not affect household net income one to one. The insurance, however, seems not to come from partner labor supply, as we find no effect on partner income. Shock pass-through is only about 25% of the raw shock, leaving net income—a prime determinant of household welfare—much less affected than gross income.

3.7 Qualifications and Extensions

In this study we have introduced a novel approach to measure transitory and persistent health shocks and have shown the causal effect of either shock type on individual-level and household-level labor market outcomes. Unfortunately, due to data constraints, we are not able to distinguish between physical and mental health shocks. The SOEP contains summary measures of physical and mental health (Andersen et al., 2007) which would theoretically allow for such a distinction. However, the physical and mental health summary measures were first incorporated in the questionnaire in 2002 and are surveyed only every other year. Therefore, the number of observed health shocks that we could examine with respect to mental or physical health is very limited. With more data available, a future extension of the paper could distinguish between physical and mental health shocks and separately assess their impact on labor market outcomes. To distinguish physical and mental health issues is important because different types of health issues pose different challenges for the recovery and the reintegration into the labor market thereafter.

Moreover, while the heterogeneity analyses include a comparison of the effects for managers with those for other employees, there is room to take into account further occupational characteristics. Employees jobs do not only play a role for their probability to experience a health shock, but also for the rehabilitation and reintegration into the labor market after a health shock. An extension of the paper could link the analysis of health shocks with the research on job task profiles (Autor et al., 2003). It is plausible that the tasks that employees perform in their jobs are decisive for their ability to quickly return to their old job after experiencing a health shock.

Another extension could concentrate on the partner's reaction to a health shock. In our analyses, we found no evidence of an added-worker effect. The added-worker effect describes partners expanding their labor supply to compensate for the income that was lost because of the health shock (De Nardi et al., 2021). At the same time, instead of expanding labor supply, partners may also be forced to decrease labor supply because they have to care for their sick partner or are more involved in other household activities. The null-effect that we find could in fact be a result of these two counteracting mechanisms. It would be interesting to examine the partner's reaction to a health shock in more detail and investigate whether the observed null-effect is just an average of two counteracting effects or if partners labor supply is indeed completely unaffected by health shocks.

3.8 Conclusion

This paper investigates the impact of transitory and persistent health shocks on labor market outcomes in Germany. To define health shocks, we follow a novel approach, relying on health-related restrictions in working capacity: sick days and hospitalizations. We exploit these variables in a two-step clustering approach, which we use to assign individuals to groups according to whether they were affected by transitory shocks, persistent shocks, or no shock at all. We cross-validate this classification with other objective and subjective health measures, finding it to be strongly predictive of bad health, regardless of the measure.

Using this novel classification, we applied an event study analysis to German SOEP data from 1993 to 2018. Our main findings are that those experiencing health shocks reduce their employment: those affected by transitory shocks by about 5 pp with respect to the baseline period, and those affected by a persistent shock by about 16 pp. Age heterogeneity is important: while younger workers (≤ 50) may return to work, older individuals (> 50) who are hit by a persistent health shock show little sign of recovery. Three periods after the shock, the employment rate of these individuals is about 25 pp smaller compared the control group. For the intensive margin, we find sharp adjustments in the period of the adverse health event for both shock types, but a partial recovery thereafter. Individual labor income decreases, especially after persistent health shocks, with no sign of catching up to the control group. These income penalties are strongly insured: only a quarter of the income loss passes through to household net income.

In trying to understand the long-lasting employment effects of health shocks, we compared the willingness to work of those out of employment after having experienced either type of shock to the control group. Remarkably, young individuals having experienced a persistent health shock are 20 pp less likely to intend to return to work than the control group. This finding points to potential for improved public policies that would help these workers to return to productive employment.

3.9 Appendix

3.9.1 SOEP Questions

150. What about hospital stays in the last year - were you admitted to a hospital for at least one night in 2017?

Yes.................. ☐ No........................ ☐ ➡ *Question **152!***

151. How many nights total did you spend in the hospital last year, that is, in 2017?

☐ nights

And how often did you have to go to the hospital in the year 2017?

☐ times

152. Were you on sick leave from work for more than 6 weeks at one time last year?

Yes, once.............. ☐

Yes, several times... ☐

No......................... ☐

Does not apply, I was not employed in 2017.......... ☐ ➡ *Question **155!***

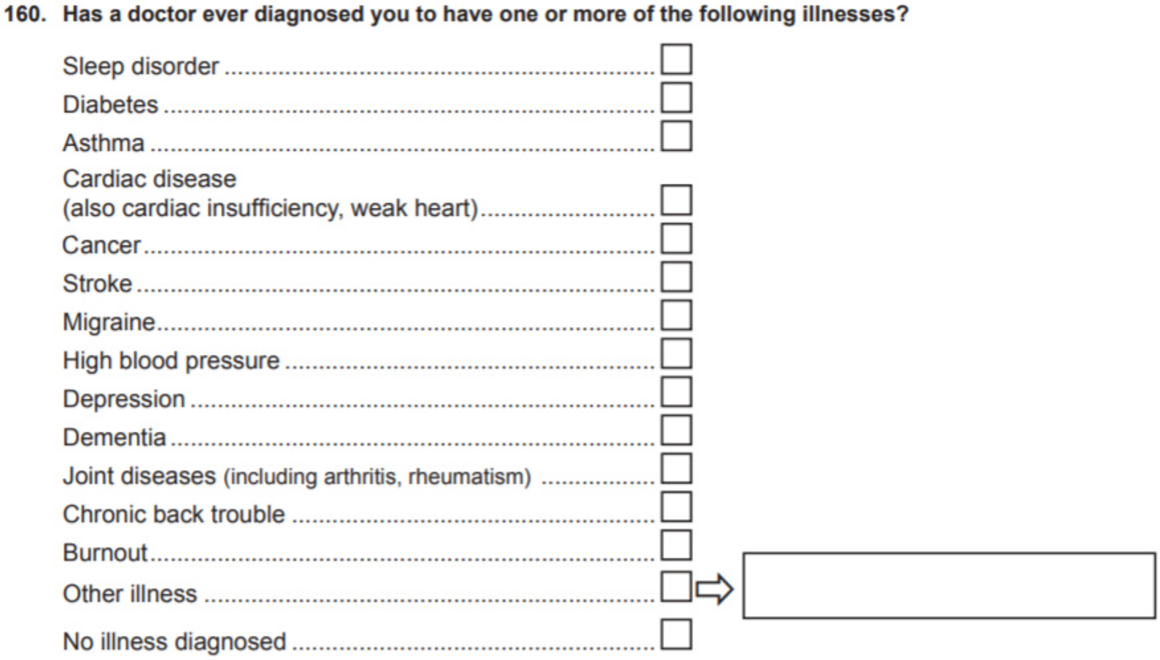
153. How many days were you unable to work in 2017 due to illness?

☞ *Please state the total number of days, not just the number of days for which you had an official note from your doctor.*

None..................... ☐

A total of .... ☐ days

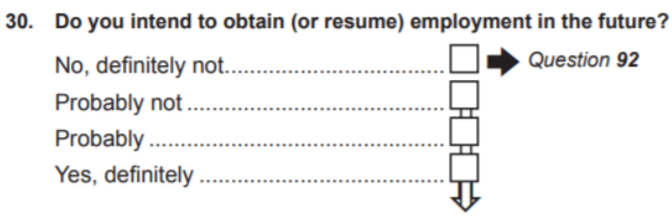
Figure 3.13: Health questions in the SOEP



160. Has a doctor ever diagnosed you to have one or more of the following illnesses?

- Sleep disorder
- Diabetes
- Asthma
- Cardiac disease (also cardiac insufficiency, weak heart)
- Cancer
- Stroke
- Migraine
- High blood pressure
- Depression
- Dementia
- Joint diseases (including arthritis, rheumatism)
- Chronic back trouble
- Burnout
- Other illness
- No illness diagnosed

Figure 3.14: Questions on chronic illnesses in the SOEP

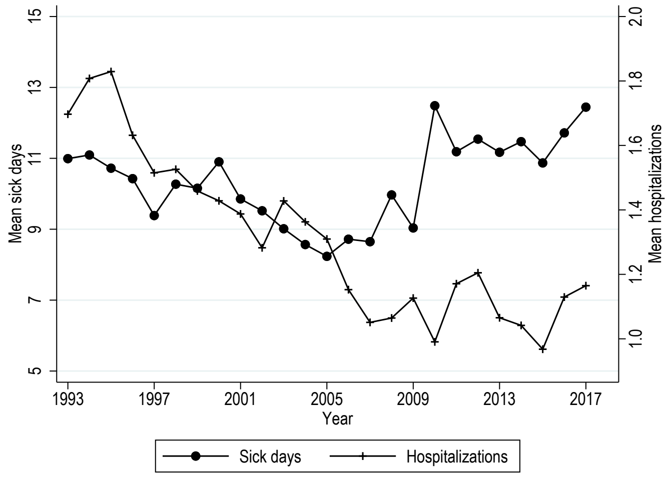


30. Do you intend to obtain (or resume) employment in the future?

- No, definitely not → *Question* ***92***
- Probably not
- Probably
- Yes, definitely

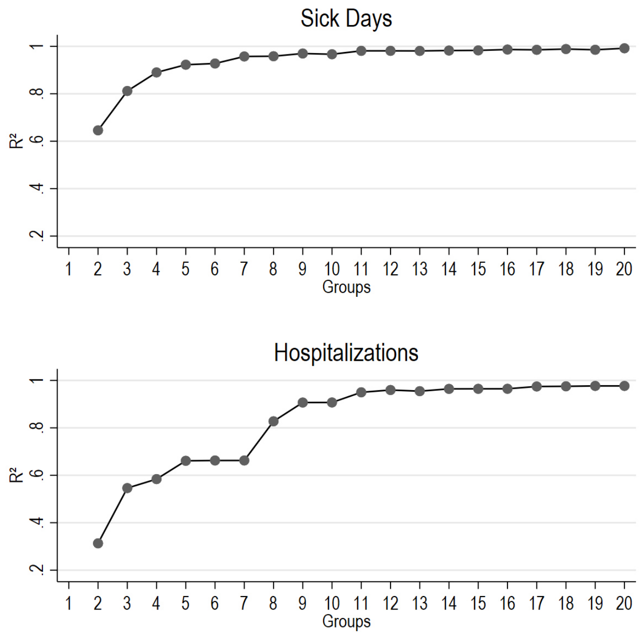
Figure 3.15: Willingness to work in the SOEP

3.9.2 Additional Tables and Figures



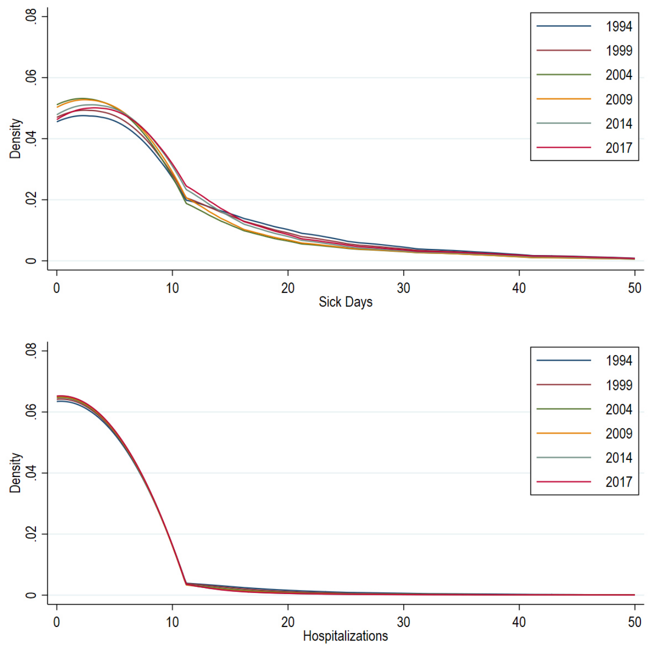
Note: Authors' calculations using SOEPv35. Based on working age population (18-65).

Figure 3.16: Means of sick days and hospitalizations



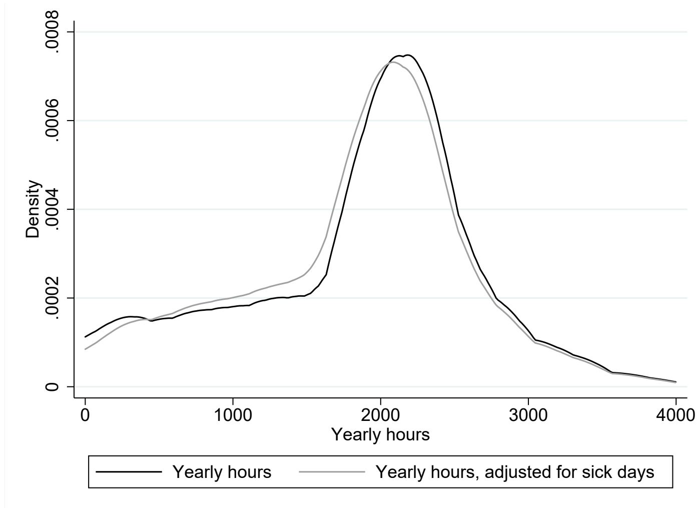
Note: Own calculations based on SOEP v35. The figure shows the R^2 for regressions of sick days and hospitalizations on cluster group dummies.

Figure 3.17: Elbow plot: R^2 along cluster group number



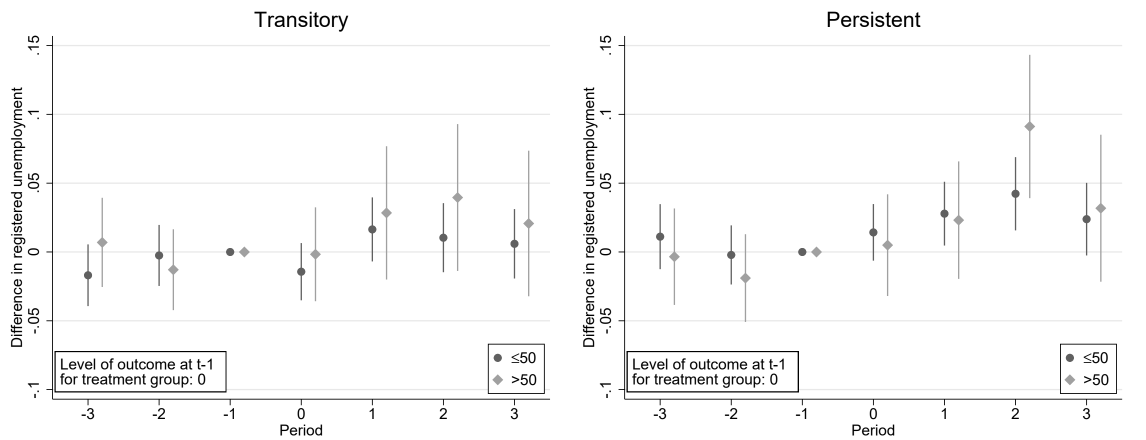
Note: Own calculations based on SOEP v35. Shows kernel densities of sick days and hospitalizations for a selected number of years.

Figure 3.18: Time series of kernel densities for sick days and hospitalizations



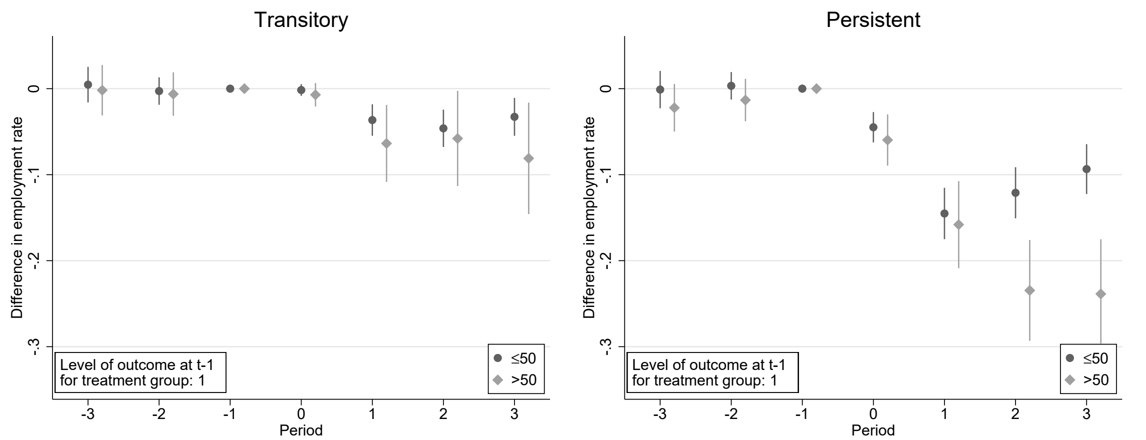
Note: Own calculations based on SOEP v35. The figure shows kernel densities for yearly hours for the working population.

Figure 3.19: Pooled distribution of yearly hours: adjusted vs. unadjusted



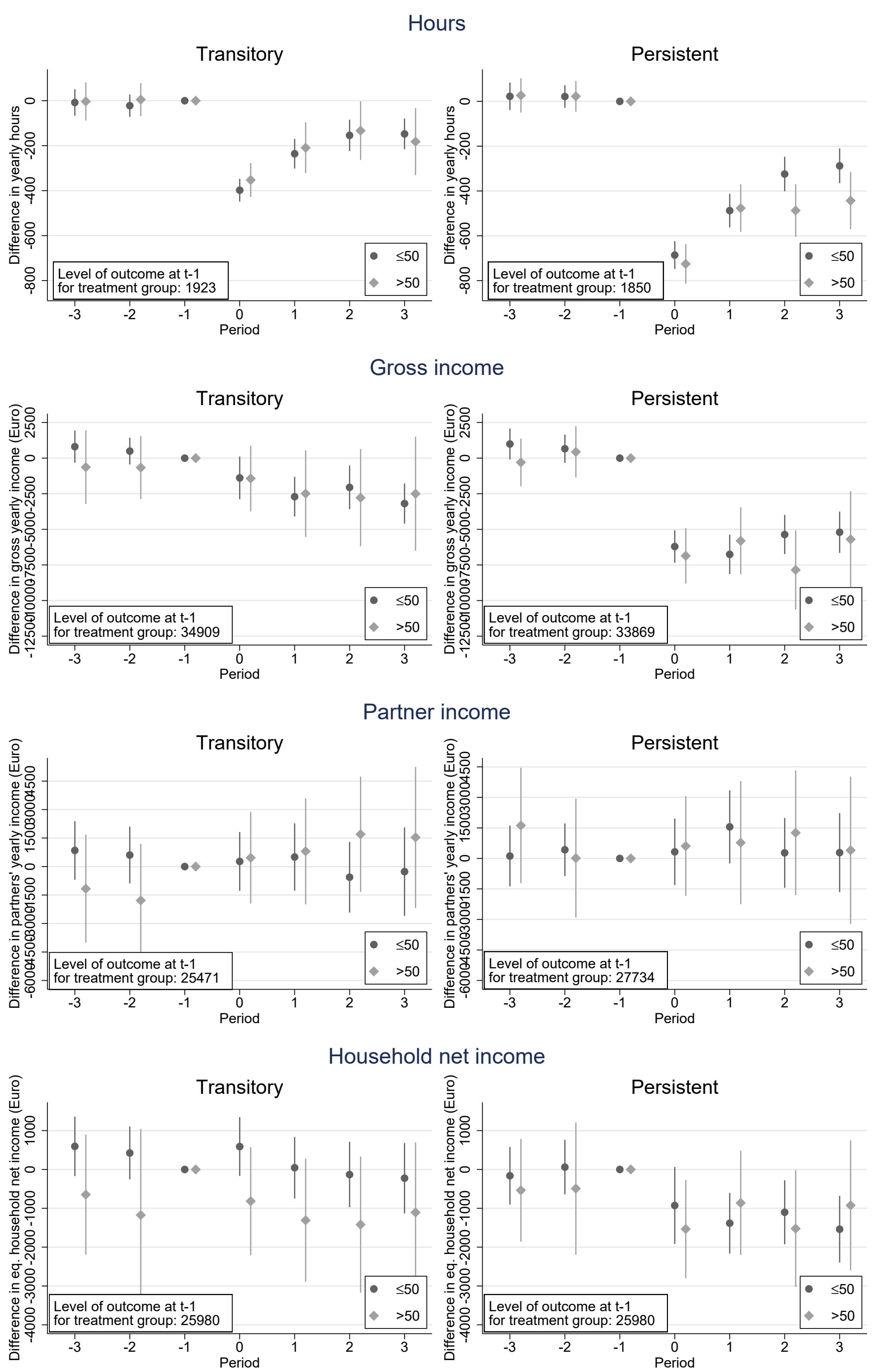
Note: Shows period-treatment-specific coefficients according to Eq.(3.2) for treated and controlled units either over or up to 50 years of age. Bars give robust 99% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 1,020 treated ≤50, 399 treated >50, 2,138 control ≤50, 732 control >50. Persistent shock: 1,132 treated ≤50, 602 treated >50, 2,173 control ≤50, 899 control >50. Source: SOEP v35.

Figure 3.20: Registered unemployment after health shocks



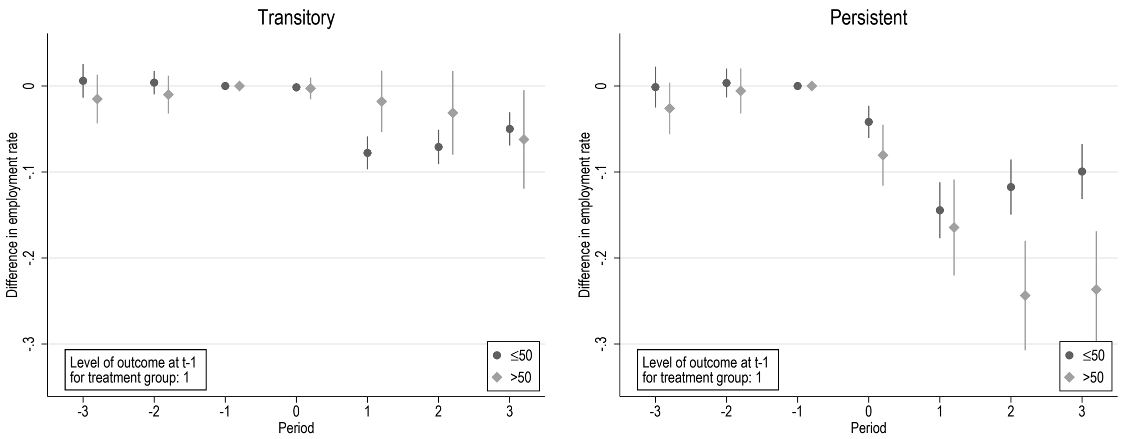
Note: Shows period-treatment-specific coefficients according to Eq.(3.2) for treated and controlled units either over or up to 50 years of age. Bars give robust 99% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 737 treated ≤50, 436 treated >50, 2,275 control ≤50, 1,135 control >50. Persistent shock: 856 treated ≤50, 569 treated >50, 2,104 control ≤50, 1,065 control >50. Source: SOEP v35.

Figure 3.21: Employment after health shocks—1SD definition



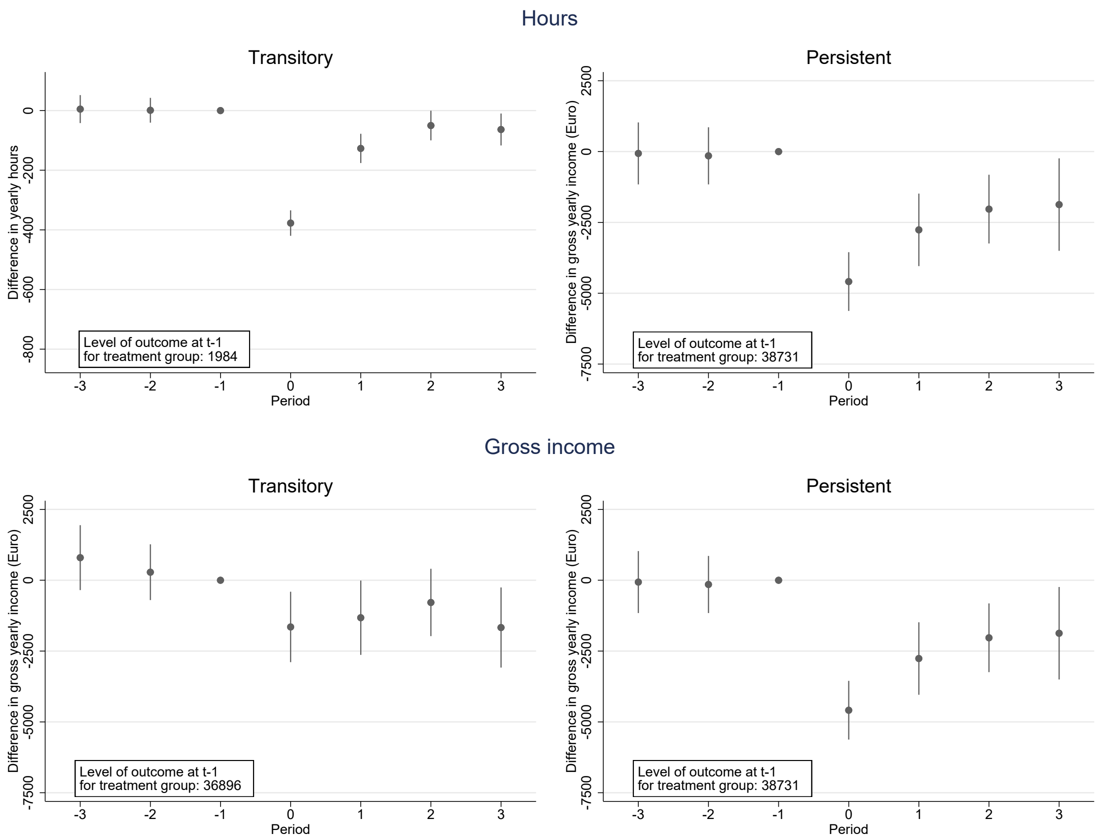
Note: Shows period-treatment-specific coefficients according to Eq.(3.2) for treated and controlled units either over or up to 50 years of age. Bars give robust 99% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 1,020 treated ≤50, 399 treated >50, 2,138 control ≤50, 732 control >50. Persistent shock: 1,132 treated ≤50, 602 treated >50, 2,173 control ≤50, 899 control >50. Source: SOEP v35.

Figure 3.22: Effect heterogeneity: age



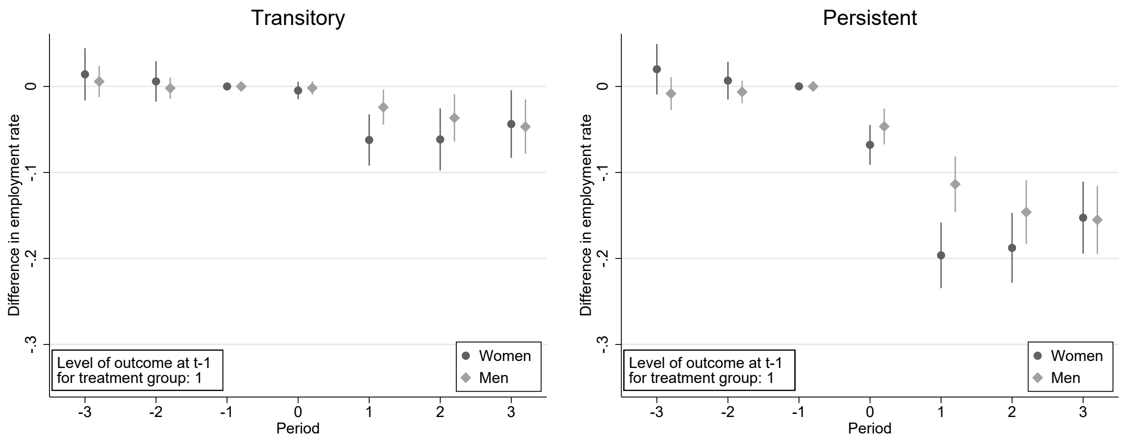
Note: Shows period-treatment-specific coefficients according to Eq.(3.2) for treated and controlled units either over or up to 50 years of age. Bars give robust 99% confidence intervals of the respective coefficients. Transitory shock group has been expanded with individuals with 0 to 29 sick days and 1 to 7 hospital nights. Number of observations: Transitory shock: 1,771 treated ≤50, 537 treated >50, 2,784 control ≤50, 778 control >50. Persistent shock: 916 treated ≤50, 517 treated >50, 1,745 control ≤50, 754 control >50. Source: SOEP v35.

Figure 3.23: Employment after health shocks—excluded units in transitory Group



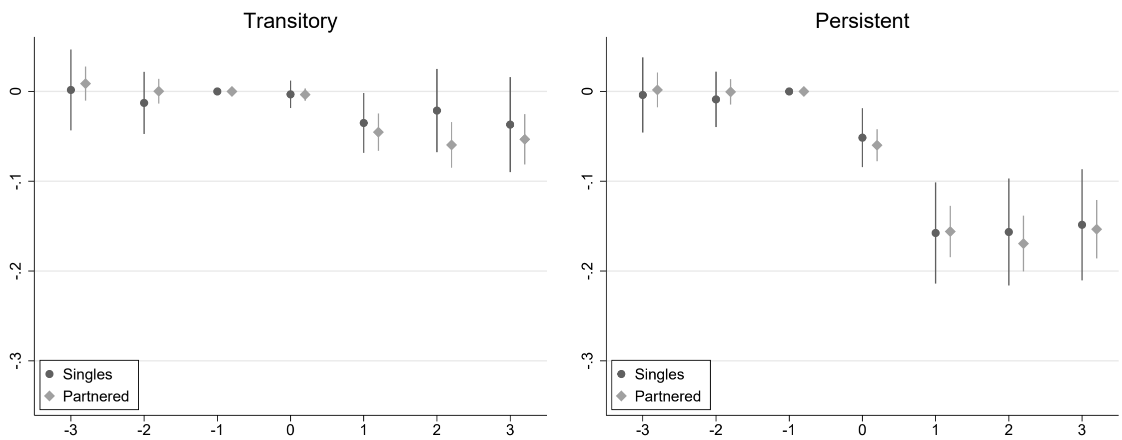
Note: Shows period-treatment-specific coefficients according to Eq.(3.2) for treated and controlled units either over or up to 50 years of age. Bars give robust 99% confidence intervals of the respective coefficients. Transitory shock figure: number of treated individuals: 1,145, control group: 2,374. Persistent shock figure: number of treated individuals: 945, control group: 2,022. Source: SOEP v35.

Figure 3.24: Hours and incomes of those remaining employed



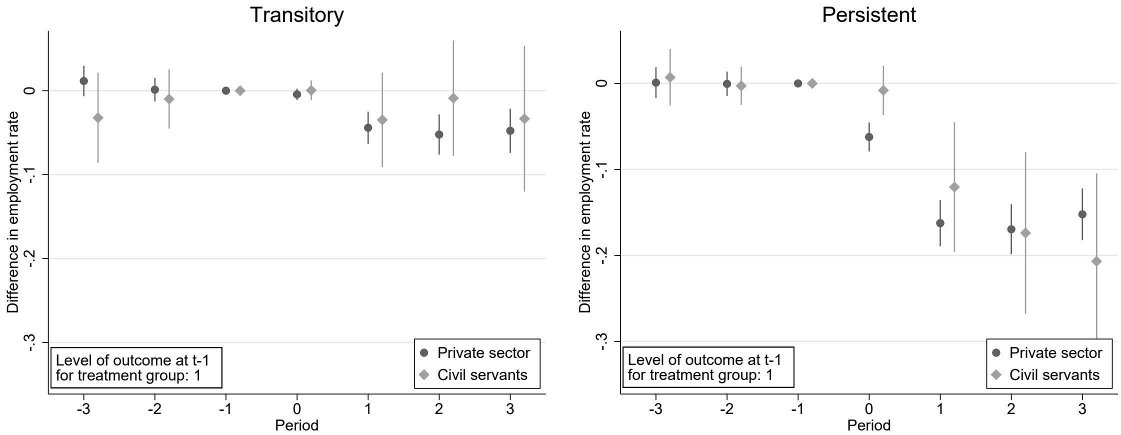
Note: Shows period-treatment-specific coefficients according to Eq.(3.2) for treated and controlled units either over or up to 50 years of age. Bars give robust 99% confidence intervals of the respective coefficients. Source: SOEP v35.

Figure 3.25: Effect heterogeneity: gender



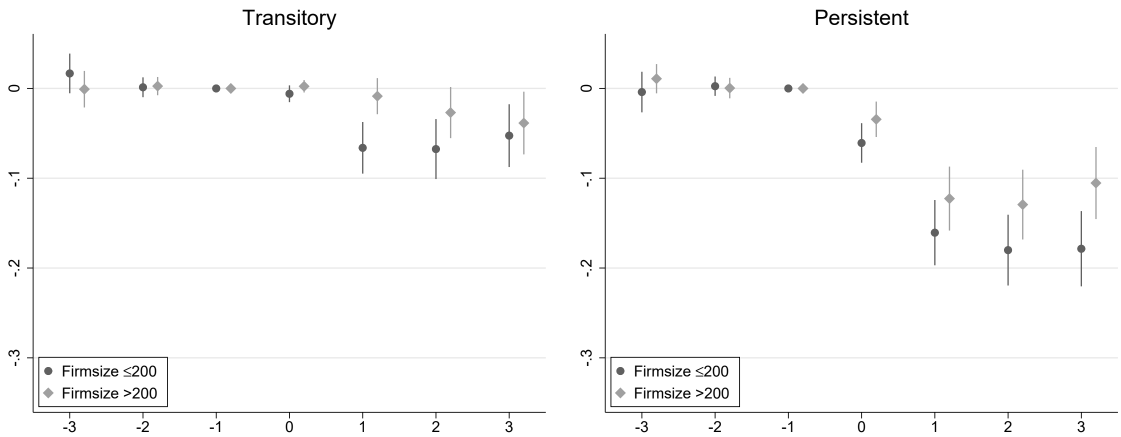
Note: Shows period-treatment-specific coefficients according to Eq.(3.2) for treated and controlled units either over or up to 50 years of age. Bars give robust 99% confidence intervals of the respective coefficients. Source: SOEP v35.

Figure 3.26: Effect heterogeneity: singles



Note: Shows period-treatment-specific coefficients according to Eq.(3.2) for treated and controlled units either over or up to 50 years of age. Bars give robust 99% confidence intervals of the respective coefficients. Source: SOEP v35.

Figure 3.27: Effect heterogeneity: civil servants



Note: Shows period-treatment-specific coefficients according to Eq.(3.2) for treated and controlled units either over or up to 50 years of age. Bars give robust 99% confidence intervals of the respective coefficients. Source: SOEP v35.

Figure 3.28: Effect heterogeneity: firm size

3.9.3 Overview of Reforms of the German Health Care and Insurance System

Over the duration of our sample, several reforms of the German health care system were introduced. Generally, these reforms were intended to reduce the expenditures of the system. In 1993, the "Gesundheitsstrukturgesetz" was introduced to allow Germans to freely choose between their health insurance providers. Further, copays for pharmaceuticals and hospitalizations were increased. In 1996, the "Beitragsentlastungsgesetz" additionally raised copays for pharmaceuticals and cut coverage of some health-related products like eyeglass frames. The "GKV-Neuordnungsgesetz" lowered the replacement rate of sickness benefits from 80% of gross but not more than 100% of net earnings to 70% and 90%, respectively. In 2002, the "Beitragssatzsicherungsgesetz" lowered the flat rates for doctors, clinicians, and hospitalizations, leading to earlier discharge after hospitalization. In 2007, the "GKV-WettbewerbsstÃ€rkungsgesetz" introduced compulsory health insurance for all Germans and established basic insurance contracts that have to be offered regardless of preexisting conditions. In 2011, the "Gesetz zur Neuordnung des Arzneimittelmarktes" slightly increased the contribution rates for public health insurance providers, and the health insurance providers were given more power in bargaining for lower pharmaceutical prices.

4 Job Tasks and Workers' Health

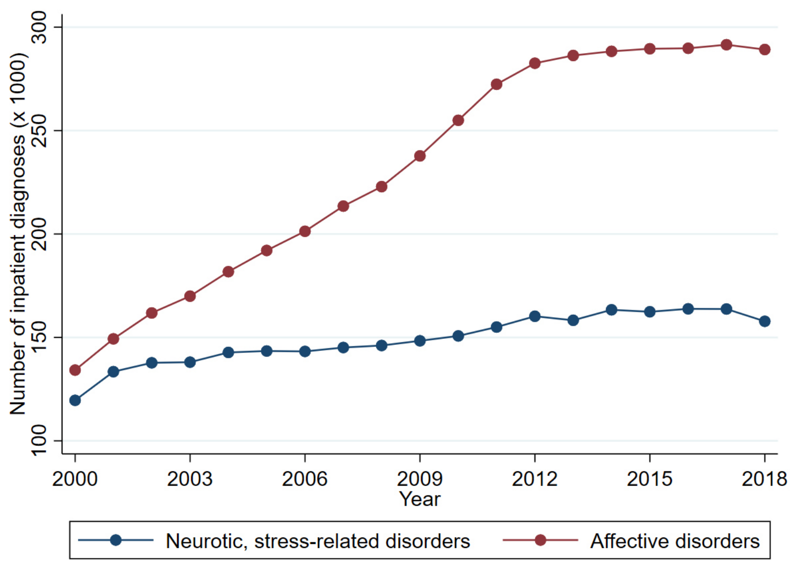
4.1 Introduction

The relationship between employment and workers' health is of major interest to researchers as well as policymakers. For one thing, bad health can impair individuals working capacity, impede employment, cause welfare losses for the affected individuals and households, and pose challenges for the design of public health insurance systems (Beckmannshagen and Koenig, 2022; Blundell et al., 2021; Galama and Kapteyn, 2011; Grossman, 1972; Haan and Myck, 2009). At the same time, job characteristics and working conditions have been found to impact workers' health (Belloni et al., 2022; Case and Deaton, 2005; Cottini and Lucifora, 2013; Danna and Griffin, 1999; Fletcher et al., 2011). One central job characteristic is the task content of occupations, which has been changing rapidly due to automation and digitization. While these changes have not yet led to the famously predicted widespread "technological unemployment" (Keynes, 1930), it has certainly affected labor markets and the working environment of workers. According to the *deroutinization* hypothesis, occupations comprising routine tasks are more exposed to the risk that tasks and processes which were formerly executed by workers can be automated through the advancement of information and communications technology (ICT) (Acemoglu and Autor, 2011; Autor et al., 2003).

The resulting decrease in demand for routine task intensive labor has been observed in many developed economies (Mahutga et al., 2018; Longmuir et al., 2020). Deroutinization has been studied extensively in the context of its consequences for employment and the distribution of wages. For example, Autor et al. (2006) have shown that the technologically-driven demand shifts contributed to wage polarization in the US, Goos et al. (2009) found the same link in Europe, while Dustmann et al. (2009) and Spitz-Oener (2006) examined how technological advancement and the resulting reduction in the demand of routine-intensive job tasks drove wage polarization in Germany.

If the task content of jobs undergoes such fundamental changes, workers' health can be affected. The link between job tasks and workers' health has not received much attention by researchers, although various effects appear plausible: Different job tasks could alter the frequency or severity of on-the-job accidents and could impact the everyday physical strain of work. Workers could become dulled by performing the same repetitive cognitive routine task over and over again. At the same time, abstract tasks could also impact the mental strain of work. Further, secondary effects of automation on health are conceivable: if workers feel that their job tasks are easily automatable, and thus, fear that their job could be in jeopardy, this could also affect their mental health (Abeliansky and Beulmann, 2019). As it is unclear to what extent these potential channels are pronounced, it is difficult to make an unambiguous *ex ante* prediction of the overall effect of routine task intensity on workers' health.

Labor market trends affect workers' lives, potentially in dramatic fashion. For example, in the US, automation and globalization, i.e., outsourcing, have been identified as contributing factors to the descend of the working class (Acemoglu and Restrepo, 2020; Case and Deaton, 2021). Workers' hardship together with institutional factors has contributed to an increase of "deaths of despair" through suicide, drug overdose, or alcoholism, which ultimately result in a decline of life expectancy (Case and Deaton, 2021). While this dramatic demographic trend is not observable in Germany[1], there are other worrying developments, especially regarding mental health. Figure 4.1 displays the yearly numbers of mental-health-related inpatient diagnoses in Germany from 2000 to 2018 based on hospital statistics. The number of affective disorder diagnoses, which contain, for example, depression and bipolar disorders, more than doubled from around 134,000 in the year 2000 to around 289,000 in 2018. Most of this increase occurred until 2012. Also, the number of neurotic, stress-related disorder diagnoses, such as stress reactions and anxiety disorders, increased steadily. The number grew from around 120,000 in 2000 to almost 160,000 in 2018.



Note: Displayed are the yearly number of inpatient diagnoses of affective disorders (ICD codes F30-F39) and neurotic, stress-related disorders (ICD codes F40-F48). Source: Hospital statistics, publicly available at Federal Statistical Office (2021a).

Figure 4.1: Number of mental-health-related inpatient diagnoses 2000-2018

[1] Dauth et al. (2021) have shown that the usage of robots in the German labor market has led to re-allocation of workers into more stable and higher-paying jobs.

To some degree, the hike in mental health related diagnoses may have increased due to a grown societal awareness for mental health issues. At the same time, it is important to know whether the increase in mental diseases is in any way connected to developments on the labor market described above or whether deroutinization impacts workers' health in other manners. Technological advancements will continue to influence the working environment as digitization and automation proceed through artificial intelligence and the usage of robots. If this had a negative impact on workers' mental health, the healthcare system should be prepared to make access to psychotherapy faster and easier. Already, the average waiting period to start psychotherapy in Germany is at 20 weeks in 2018 (Bundespsychotherapeutenkammer, 2018). Besides the obvious consequences for the population's well-being, bad health also entails large economic costs. For example, bad health is a major reason for aging workers to leave the labor market before reaching the regular retirement age, which puts pressure on the public pension system (Beckmannshagen and Koenig, 2022; Blundell et al., 2021; Haan and Myck, 2009). Thus, for the financial sustainability of the welfare state as well as for the design of occupational safety measures and healthcare accessibility, it is important to understand if and how health is affected by one of the secular trends of modern labor markets: the automation of job tasks driven by technological advancement.

The aim of this paper is to quantify the causal effect of occupational routine task intensity on workers' mental and physical health. The main contribution lies in linking the research on deroutinization of job tasks through technological advancement to a detailed assessment of workers' health outcomes. By combining two longitudinal data sources, the German Socio-economic panel (SOEP) (Goebel et al., 2019) and the Occupational Information Network (O*NET), I separately assess how the routine task intensity of occupations affects the mental and physical health of male and female workers. To understand how occupational task content affects workers' health is of major importance because deroutinization is not a finished process of the past. Instead, technological advancement will progress, and automation will continue to affect workers' job tasks in the future (Acemoglu and Restrepo, 2018, 2022).

To operationalize the occupational task content, I rely on the routine task intensity index (RTI), an established measure to describe whether an occupation is intensive in routine tasks (Autor and Dorn, 2013; Autor et al., 2015; Goos et al., 2014). Based on the O*NET database I construct a time series of the RTI for 339 occupations, which are classified according to the ISCO-88 classification. Based on the ISCO-88 classification each employed respondent in the SOEP is then assigned the corresponding RTI score of her occupation. The SOEP also contains the mental component summary score (MCS) and the physical component summary score (PCS), which are used as dependent variables to assess individuals' health.

Estimating a causal effect of job task intensity on health outcomes is not trivial. Simple OLS estimations will be biased because of simultaneity and endogeneity induced by unobservable characteristics. The issue of simultaneity may arise, because workers potentially adjust their occupational tasks according to their physical and mental condition. Omitted variable bias may occur because of selection based on unobservable characteristics. For example, individuals that particularly care about their health may pursuit a career with a specific task profile aligned with this preference. To overcome these endogeneity issues, I apply an instrumental variable approach in the spirit of Bartik (1991). In this framework, identification stems from exogenous differences in regional exposure to economy-wide shocks. In my case, I instrument a worker's occupational routine task intensity with the local employment growth rate in manufacturing which is predicted by interacting the initial regional share of workers in manufacturing and the yearly country-wide growth rate of the manufacturing sector (Goldsmith-Pinkham et al., 2020). I provide several test statistics to verify the instrument's relevance and thoroughly discuss its exclusion restrictions.

With the instrument in place, I find a significant negative effect of routine task intensity and workers' physical health when considering the whole German workforce. Accordingly, an increase by one standard deviation in the RTI leads to a decrease in the PCS by 0.3 standard deviations. For mental health there is no significant overall effect. To put this into context, a one standard deviation increase in the RTI is equal to the difference in routine task intensity between a social care worker (-0.99) and a butcher (0.004) in 2018. People with moderate obesity[2] on average have lower PCS by half a standard deviation compared to people with normal weight. The 0.3 standard deviation effect I find for the PCS is not equal but close to that difference.

While even the effect on the workforce as a whole is meaningful, it neglects the pervasive heterogeneity, particularly between male and female workers. For women, a one standard deviation higher occupational routine task intensity decreases the MCS by 0.6 standard deviations, while there is no significant effect on physical health. For men, a one standard deviation higher occupational routine task intensity increases the MCS by 0.7 standard deviations but lowers the PCS by 0.6 standard deviations. The gender-specific effects are due to the fact that the routine tasks of female workers substantially differ from the routine tasks of male workers. For men, a high routine task intensity is more likely linked to performing manual routine tasks in production or operator occupations, entailing physical strain. For women, routine tasks more likely present cognitive routine tasks in clerical or service occupations entailing mental strain (Acemoglu and Autor, 2011).

Accordingly, the continuing deroutinization of jobs has specific and distinct effects on female and male workers' health. A decrease of routine tasks on the job for women leads to better mental health and has no impact on physical health. For men, fewer routine tasks lead to better physical but worse mental health. This result is buttressed in several robustness checks and adds to the literature on the

[2]Moderate obesity refers to a body mass index between 30 and 39.9.

effects of technological change on labor markets and workers. Understanding how technological advancement and automation affect labor markets and society as whole is crucial, as the rise in "deaths of despair" (Case and Deaton, 2021) in the US is just one example that underlines how far-reaching the consequences of labor market trends can be. To anticipate and potentially prevent similar societal developments, causal evidence on the health effects of automation—as provided in this study—is indispensable.

The remainder of the paper is structured as follows: In Section 4.2, I describe the data basis and present descriptive statistics of the main variables. In Section 4.3, I explain the empirical strategy based on an instrumental variable approach. I present the results in Section 4.4 and discuss and compare them to the existing literature in Section 4.6. Section 4.7 reviews limitations and potential extensions of this paper. Section 4.8 concludes.

4.2 Data

Analyzing the health effects of occupational routine task intensity comes with high data requirements. For the empirical strategy applied in this paper, one needs individual-level data with information on occupations, health, and regions individuals live in. Moreover, one needs information on occupational task profiles to measure the routine task intensity of occupations. Lastly, for the Bartik-style instrument, one needs aggregate information on regional industry shares and industry growth rates. To meet these requirements, the present paper makes use of three data sources: 1) The German Socio-economic Panel (Goebel et al., 2019), 2) the Occupational Information Network (O*NET), which is funded by the US Department of Labor, and 3) publicly available administrative data from the National Accounts by the Federal States of Germany (Arbeitskreis „Volkswirtschaftliche Gesamtrechnungen der Länder", 2021).

The German Socio-economic Panel The SOEP is a longitudinal representative household survey, as of 2019, comprising around 30,000 respondents annually. It contains a comprehensive list of socio-economic characteristics, labor market information, and several variables summarizing the respondents' health (Goebel et al., 2019).

Measuring health is challenging as both measures of health satisfaction and objective information on diagnoses can suffer from measurement error (Bound, 1991). An alternative approach is to measure individuals' health status by surveying their quality of life. Since 2002 the SOEP administers the Short Form (12) Health Survey (SF-12).[3] The SF-12 surveys eight scales regarding mental and physical health.

[3]See Figure 4.6 for an overview of the surveyed items.

Based on these scales, the Physical Component Summary Scale (PCS) and the Mental Component Summary Scale (MCS) are constructed (Andersen et al., 2007; Ware et al., 1996). Both variables are well established in the epidemiological and in the economic literature (Salyers et al., 2000; Marcus, 2013; Schiele and Schmitz, 2016). PCS and MCS are computed by means of z-transformation and in the raw SOEP data have a mean of 50 and a standard deviation of 10 for the whole survey population in the year 2004. For the empirical analyses of this paper, both health summary scores are standardized and rescaled to have mean of zero and a standard deviation of 1 for the working population in every year. Accordingly, a PCS of 1 would indicate better physical health by one standard deviation relative to the average working population in that year.

Particularly relevant for this study is the information on the employed respondents' occupations based on the International Standard Classification of Occupations (ISCO). The SOEP provides 4-digit occupation codes based on the ISCO-88 classification for the years until 2017 and on its successor, the ISCO-08 classification for the years from 2013 onwards. This paper uses the ISCO-88 classification because it is available for the larger part of the observation period. For 2018, I compute ISCO-88 codes by applying a crosswalk from the ISCO-08 classification[4] and forward imputing the ISCO-88 code from 2017 if no occupation changes occurred.[5]

The O*NET database Based on the ISCO-88 codes, each employed SOEP respondent is matched to her occupational task profile, which is taken from the O*NET data. The O*NET database is developed under the sponsorship of the US Department of Labor and contains a rich set of variables that describe work and worker characteristics, including skill requirements and job tasks for almost 1,000 occupations.

I use yearly O*NET data to construct time-series of job task information.[6] In the O*NET data occupations are coded according to the ONET-SOC classification. Hardy et al. (2018) provide a crosswalk that translates the ONET-SOC classification to ISCO-88 classification codes. I apply this crosswalk to make the O*NET data linkable to the SOEP. A challenge is that the O*NET data is not updated regularly across all occupations. For some occupations there are no data in the early years. In this case I impute the job content information with the first available observation for that occupation.[7] Also, if there is information for more specific occupations (e.g., 4-digit ISCO codes) but not for the more broadly defined 3-digit or 2-digit occupations, I impute the occupational information with the most general subcategory, if the broad

[4] I apply the user-written Stata package `iscogen` (Jann, 2019).

[5] The information whether the occupation was changed is available in the SOEP data.

[6] I use the data releases 5.0, 6.0, 8.0, 10.0, 12.0, 13.0, 14.0, 15.0, 16.0, 17.0, 18.0, 19.0, 20.0, 21.0, 22.0, 23.0.

[7] This is the case for 49 occupations in 2004 to 2006 and 26 occupations between 2007 and 2011. As a result, 2.2 percent of my main regression sample (workers in the SOEP) have imputed occupational information.

categories are assigned to individuals in the SOEP. For example, the information for "office clerks" (coded 4100) is imputed with the information for the subordinate "other office clerks" (coded 4190). Table 4.5 in the Appendix gives an overview of the sample size and the share of imputed values for each year. Lastly, 3-year moving averages of each job content item are constructed. This is done to correct for the irregular updates of task content across occupations within the O*NET database. As a result, I obtain a stable time series of job content items and their gradual changes over time for the period 2004 to 2018.[8]

Based on this time series, I follow Autor et al. (2003) and Acemoglu and Autor (2011) and construct three composite measures that contain the task inputs of 1) abstract, 2) routine, and 3) manual tasks for each ISCO-88 occupation. Based on these three composite task measures, an index summarizing the routine task intensity of each occupation is constructed (Autor and Dorn, 2013; Goos et al., 2014):

$$RTI_{o,t} = \ln(rou_{o,t}) - \ln(man_{o,t}) - \ln(abs_{o,t}), \tag{4.1}$$

in which $RTI_{o,t}$ is the routine task intensity index (RTI) for each occupation in every year, and $rou_{o,t}$, $man_{o,t}$, and $abs_{o,t}$ are the respective composite task measures capturing routine tasks, manual tasks, and abstract tasks, for each occupation in every year.[9] Using the relative weights of the SOEP working population, the RTI is standardized to have mean zero and a standard deviation of one in 2004. Accordingly, a RTI value of one indicates that an occupation is one standard deviation more intensive in routine tasks than the occupation of the average worker of the German workforce in 2004. The RTI thus reduces the dimensionality of task profiles of occupations and summarizes their content in one continuous index. For example, in 2004 pastoral workers exhibit the lowest RTI score (-2.23), indicating a high degree of abstract and non-routine tasks, while data entry operators exhibit the highest RTI score (2.97), indicating a high degree of routine tasks. The RTI will be used as the main explanatory variable because it captures the task content of a worker's job in a one-dimensional continuous variable. It also makes this study easily comparable to other studies on routine task intensity of jobs as the RTI is well established in the literature (Autor and Dorn, 2013; Autor et al., 2015; Goos et al., 2014).

Combining the US-based O*NET data on occupational tasks with the German SOEP raises the question if it is valid to assume that workers in the same occupations in Germany and the United States perform the same tasks. One alternative data source containing descriptions of occupational tasks in Germany is the German Qualification and Career Survey and its successor the BiBB/BAuA Employment Sur-

[8]Note that none of my results hinges on the imputations or corrections of the data, as shown in tables 4.12 and 4.14 in the Appendix.

[9]Note that for the computation of the RTI, the composite task measures are not standardized but comprise the sum of the respective task items as characterized in Acemoglu and Autor (2011).

vey.[10] For the period under investigation in this paper, the BiBB/BAuA Employment Survey only comprises three waves from 2006, 2012, and 2018. This makes it impossible to construct a yearly time-series of occupation tasks as is done based on the yearly O*NET waves. Further, the analyses of Cedefop (2013) show that the O*NET data are highly correlated—with correlation coefficients mostly around 0.8—with similar European surveys and thus can be applied to describe occupational tasks in European labor markets as well.

National Accounts data The main analysis of this paper is based on an instrumental variable approach, as explained in Section 4.3. For the instrumental variable approach, I rely on administrative data from the National Accounts of the Federal States (Arbeitskreis „Volkswirtschaftliche Gesamtrechnungen der Länder", 2021). From these data I use two variables: 1) The share of workers working in the manufacturing sector at the county level (NUTS-3), and 2) the gross value added of the manufacturing sector. Based on the gross value added I calculate growth rates for every 2-year period from 2004 to 2018.

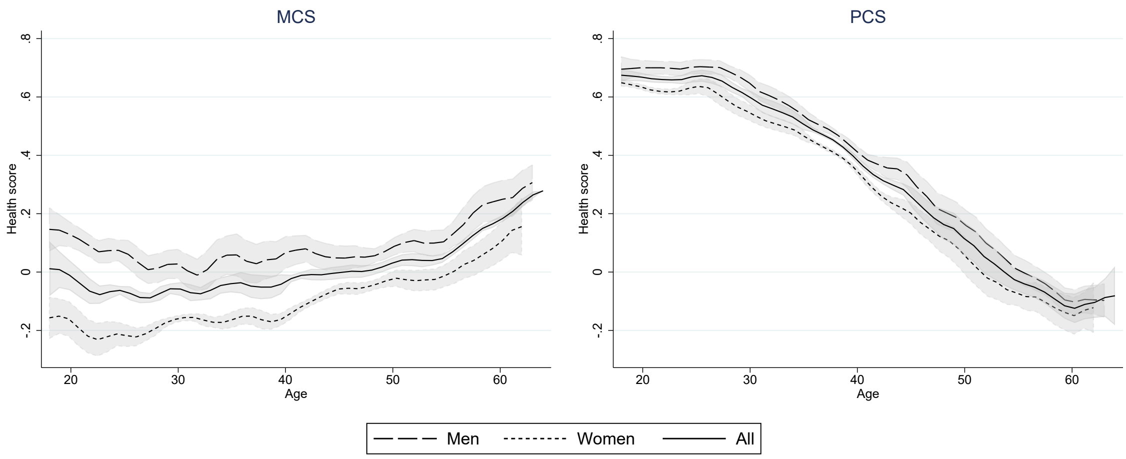
Sample restrictions The two health summary scores are surveyed biennially starting in 2002. The O*NET data are available since 2003. As a result, this study covers the time period 2004 to 2018. As the health scores are only available in every second year, the regression analysis is based on a pooled cross-section covering eight years in that period. The analysis sample contains all employed individuals between the ages of 18 and 65. Accordingly, I obtain 89,157 observations of 33,329 different individuals.

Data description The central dependent variables of this paper are the two health summary scores measuring workers' health: MCS and PCS. Measuring individuals' health is a great challenge due to individual heterogeneity in the perception of one's physical and mental state (Bound, 1991; Lindeboom and Kerkhofs, 2009). Both, data on subjective health satisfaction as well as data on diagnoses suffer from this issue. The SF-12 is a survey instrument that aims to address these issues by asking a battery of questions regarding the health-related quality of life. Based on these questions the mental and physical health summary scores are constructed as described in Andersen et al. (2007).

Figure 4.2 shows how these summary scores capture mental and physical health over the life cycle. It displays means of both health scores by gender and age. The left panel of the figure shows that the mental health summary score is u-shaped over the life cycle. Women have worse mental health in younger ages but start to catch up in their 40s. In the right panel the physical health summary score is displayed. In line with evidence from other countries, physical health in Germany is declining

[10] For applications see Spitz-Oener (2006) or Dustmann et al. (2009).

over the life cycle (Hosseini et al., 2021b). The decline begins around the age of thirty and is almost linear from there on. Throughout most of the life cycle the average PCS is slightly lower for women than for men but the difference diminishes in the late 50s.

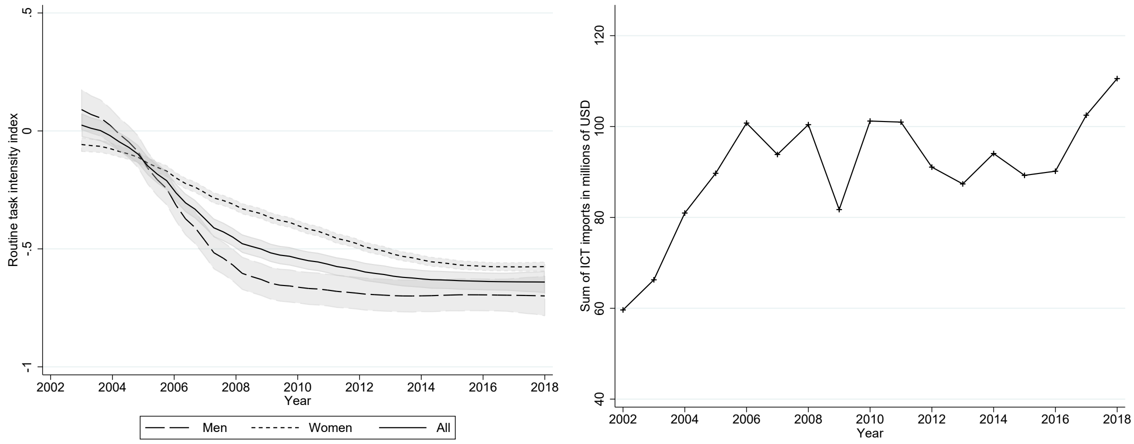


Note: Displayed are local polynomial smooth plots (bandwidth=1) including 95-percent confidence intervals of average health summary scores by gender over the life cycle. Values are standardized to have mean 0 and a standard deviation of 1 in each survey year. Values are only displayed if there are at least 1,000 observations for each gender and every specific age. Figures are based on a pooled sample of all individuals between 18 and 65 in the SF-12 survey years 2004 to 2018. Own calculations based on SOEP v36.

Figure 4.2: Mental and physical health over the life cycle

The central explanatory variable analyzed in this paper is the routine task intensity index. By using the RTI index I follow Autor and Dorn (2013), Autor et al. (2015), and Goos et al. (2014) in utilizing the singular summary measure of the importance of routine, abstract or manual tasks within occupations. An alternative approach is the clustering of occupations into groups that are particularly intensive in either routine, abstract or service tasks and then examining the relative growth of these occupations within the workforce (Acemoglu and Autor, 2011; Longmuir et al., 2020). Irrespective of the approach—occupation classes or RTI—there is extensive evidence for many developed economies that technological progress and automation led to a decline in the relative demand for routine-intensive jobs (Acemoglu and Autor, 2011; Autor et al., 2003, 2006; Autor and Dorn, 2013; Autor et al., 2015; Goos et al., 2009; Longmuir et al., 2020). As this trend of technologically driven deroutinization of jobs has also been shown in Germany (Dustmann et al., 2009; Spitz-Oener, 2006), the RTI of the German workforce constructed in this study should resemble the finding of a decrease in average task intensity among workers.

The left panel of Figure 4.3 shows the development of the average RTI of the whole German workforce over time, and separately for male and female workers. By definition, the mean for the whole workforce is at zero in 2004. In 2018 the



Note: Left panel: local polynomial smooth plots (bandwidth=1) including 95-percent confidence intervals of average RTI within the German workforce from 2003 to 2018. Right panel: yearly imports of ICT goods into Germany in millions of USD from 2002 to 2018. The import data are official statistics which is why they are displayed without smoothing and confidence intervals. Despite not being in the observation period of the remainder of the paper, the years 2002 and 2003 are included in this figure if available to show the full picture of the trends. Source: SOEP v36, O*NET, UNCTADstat, publicly available.

Figure 4.3: Average RTI of the German workforce and ICT imports

average RTI is 0.64 standard deviations smaller relative to 2004, indicating that workers in 2018 are less often performing routine tasks than in 2004. The large part of this decrease occurs until 2008, while the decline slows down thereafter. The average RTI is higher for men than for women at the beginning of the observation period. But male workers also experience a larger decrease in routine task intensity, resulting in a slightly lower RTI for men than for women in 2018. The right panel of the figure provides a potential explanation for the steep decrease of average routine task intensity during the 2000s. It shows the sum of ICT goods that are imported to Germany over time. Imports of ICT goods sharply increased from 60 million in 2002 to 100 million in 2006 and stayed on this rather high level in the following years. The large inflow of ICT goods in the decade from 2000 to 2010 tracks the observed decrease in average routine task intensity in the German labor market. This suggests a possible link between the drop in occupational routine task intensity and the dissemination of ICT in the working environment.

4.3 Empirical Strategy

Estimating the effect of occupational routine intensity on mental and physical health comes with two major challenges. The first issue is simultaneity. Not only do job characteristics influence workers' health, at the same time the health condition of an individual can also affect or limit the individual's labor supply and occupational choice (Beckmannshagen and Koenig, 2022; Blundell et al., 2021). For example, if workers in a bad physical health condition choose occupations without any physical strain, an OLS estimation would underestimate the positive effect of practicing that occupation on health. The second issue is omitted variable bias: unobserved individual characteristics, for example a preference for a healthy lifestyle, may influence both, the career choice and thus the occupational tasks performed by a worker, and her health behavior. An OLS estimation would ignore this selection and overestimate the positive effect on health, crediting the worker's good health only to the occupational task profile but not taking into account the good health behavior.

To address these challenges, I apply an instrumental variable approach in the spirit of Bartik (1991). In this setting, two facts are exploited for identification: a) that workers in the manufacturing sector are more likely to perform routine tasks in their jobs and are thus particularly exposed to a technologically-driven deroutinization of their job tasks (Acemoglu and Autor, 2011; Autor et al., 2003), and b) that the share of workers in the manufacturing sector differs substantially between regions in Germany. I thus use the differential regional exposure to the economy-wide automation *shock* to estimate the effect of occupational routine intensity on workers' health.

The Bartik-style instrument, $z_{i,r,t}$, is constructed as follows:

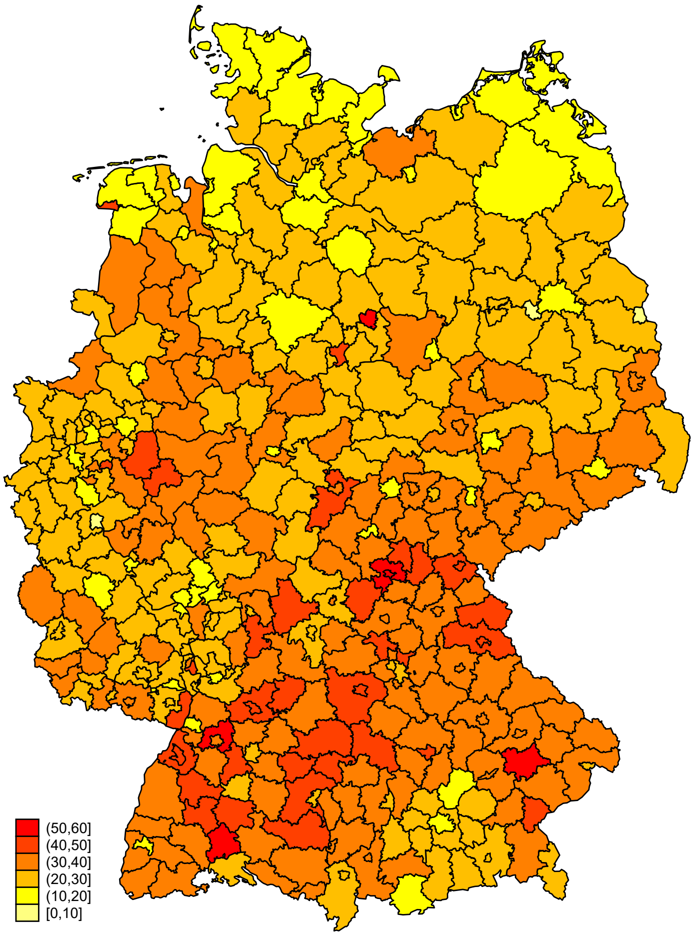
$$z_{r,t} = shm_{r,t04} \times gm_t, \tag{4.2}$$

with $shm_{r,t04}$ containing the share of workers in the manufacturing sector on the regional level in the initial period, the year 2004, and gm_t comprising the country-wide growth rate of the manufacturing sector for every 2-year period.

In every period t, every individual living in region r is assigned with the product of the region-specific share of workers in the manufacturing sector in the initial period 2004 and the *country-wide* growth rate of the manufacturing sector.[11] Accordingly, the identification stems from the regional heterogeneity in exposure—characterized by the manufacturing share—to the ongoing automation *shock*, which is measured by the country-wide growth rate of manufacturing.

Figure 4.4 shows the regional shares of workers in the manufacturing sector in the initial period of 2004. The underlying regional units are on the NUTS-3 level, i.e., the 401 counties within Germany. The figure shows that there is substantial variation in manufacturing shares between regions. Shares vary from 8 percent in Potsdam, to 59 percent in the county of Dingolfing-Landau, which is home to the largest BMW production site in Europe. For the construction of the instrument, these regional

[11]The regional information in the SOEP refers to individuals' residence, not their work location. However, the median of SOEP respondents' commuting distance at 10 km, p90 at 40 km. The average county size is 892 km^2. Thus, it seems appropriate to use the instrument based on the county of individuals' residence as long commuting distances are rare. In Section 4.5, the regional unit is changed to commuting zones, which does not alter the main results.



Note: Displayed are the county level shares of workers in the manufacturing sector in percent. Source: National accounts data provided by Arbeitskreis „Volkswirtschaftliche Gesamtrechnungen der Länder" (2021).

Figure 4.4: County-level shares of workers in manufacturing in 2004

shares are multiplied with the country-wide growth rates of the manufacturing sector which vary from 1 percent in 2009-2010 to 9 percent in 2015-2016.

The Bartik instrument is used to estimate the effect of the job task content, operationalized by the RTI index, on both health summary scores in a 2SLS approach. In the first stage,

$$RTI_{i,t} = \alpha_1 + \beta_1 z_{r,t} + \gamma_1 \mathbf{X} + \tau_t + \epsilon_{i,t}, \tag{4.3}$$

the main explanatory variable, $RTI_{i,r,t}$ is regressed on the Bartik instrument $z_{r,t}$, a set of control variables $\mathbf{X}$, period dummies τ_t and an idiosyncratic error $\epsilon_{i,t}$. Vector $\mathbf{X}$ includes a gender dummy, a quadratic in age, and a dummy indicating whether an individual's father obtained a college degree. These control variables are chosen because they are strictly exogenous to the occupational routine task intensity of an individual. Based on this first stage regression, values for $\widehat{RTI}_{i,t}$ are predicted and used as the main explanatory variable in the second stage regression:

$$H_{i,t} = \alpha_2 + \beta_2 \widehat{RTI}_{i,t} + \gamma_2 \mathbf{X} + \delta_i + \tau_t + \varepsilon_{i,t}, \tag{4.4}$$

in which $H_{i,t}$ represents the two main variables of interest, $mcs_{i,t}$ and $pcs_{i,t}$, $\mathbf{X}$ are the same control variables as in the first stage, δ_i and τ_t are personal and time fixed effects, while $\varepsilon_{i,t}$ is an idiosyncratic error term.[12]

For the β_2-coefficient to consistently estimate the effect of occupational routine task intensity on the respective health score, two assumptions must hold. First, the instrument must be *relevant*, meaning that the Bartik instrument must have sufficient predictive power for $RTI_{i,t}$, the routine task intensity index. This is validated by examining parameters of the first stage regression. Based on the high Kleibergen and Paap (2006) F-statistic of 92.1 it can be ruled out that the Bartik instrument is weak. It has strong predictive power for routine task intensity. As expected, a higher value of the shift-share instrument for a region is associated with a higher routine task intensity of a worker living in that region. The second assumption is the *exclusion* restriction, which states that the instrument must be independent of the error term of the second stage $\varepsilon_{i,t}$. As discussed in Goldsmith-Pinkham et al. (2020), this assumption could be particularly problematic if the model is estimated in levels because the regional share of workers in manufacturing is determined in equilibrium resulting from many factors on the supply and the demand side. However, including fixed effects in the model—thus examining changes rather than levels—alters the assumption so that changes in the instrument must be exogenous to changes in the error term. Accordingly, in this case identification is based on the assumption that regional manufacturing shares are exogenous to changes in individuals' health. This assumption would be problematic if workers systematically moved to regions with growing or shrinking manufacturing shares because of changes in their health status, which is highly unlikely. Thus, including fixed effects in the model substantially strengthens the argument that the exclusion restriction holds and the IV approach consistently estimates the causal effect of routine task intensity on both health measures.

4.4 Results

Main analysis Table 4.1 displays the main regression results. For both outcome variables the results for the whole sample and separately for male and female workers are displayed. The results are based on Equation (4.4), which I estimate in 2SLS to utilize the shift-share instrument and at the same time include fixed effects

[12]Note that in the second stage the time-invariant control variables in $\mathbf{X}$, such as father's education, drop out because personal fixed effects are included.

in the second stage. In Table 4.6 in the Appendix, I also present regression results of more simplistic specifications, which will be biased for the reasons discussed above.

Table 4.1: Instrumental variable regressions of health scores on routine-task intensity, full sample and by gender

	MCS			PCS		
	All	Men	Women	All	Men	Women
RTI	-0.098	0.737*	-0.605**	-0.314**	-0.618**	-0.132
(s.e.)	(0.193)	(0.404)	(0.241)	(0.154)	(0.310)	(0.182)
Obs.	89,157	44,944	44,213	89,157	44,944	44,213

Note: All regressions include a gender dummy, a quadratic in age, and a dummy whether an individual's father obtained a college degree and time fixed effects. In the second stage personal fixed effects are included, so the time-invariant controls drop out. Standard errors are bootstrapped (1,000 runs). Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v36, O*NET, National accounts data.

For mental health, the regression yields a very small negative but insignificant coefficient when considering the whole workforce. The absence of a significant effect for the whole sample is a result of opposing effects of routine task intensity on mental health for men and women. For men, an increase in routine task intensity by one standard deviation leads to a higher MCS by 0.7 standard deviations. For women, an increase in routine task intensity by one standard deviation leads to a lower MCS by 0.6 standard deviations. Both effects are statistically significant, the effect for women at the 5 percent significance level, the effect for men only at the 10 percent significance level.

Looking at physical health, I find a significant negative effect of routine task intensity on the PCS when considering the whole workforce. Accordingly, an increase in the RTI by one standard deviation leads to a lower PCS by 0.3 standard deviations. This effect is mostly driven by men. When only considering men, an increase in the RTI by one standard deviation leads to a lower PCS by 0.6 standard deviations. For women, the point estimate is also negative, however insignificant.

The first stage statistics presented in Table 4.2 strengthen argument that the implemented IV strategy is valid for the whole sample as well as after the sample split by gender. The Bartik instrument's coefficient is positive and significant at the 1 percent significance level. Thus, higher regional employment growth in manufacturing is predictive of a higher occupational routine task intensity among workers in that region. In all three cases, the Kleibergen and Paap (2006) statistic is sufficiently high and yielding a p-value smaller than 0.01. This is strong evidence that the relevance assumption holds, which is essential for identification.

Table 4.2: First-stage regression statistics

	All	Men	Women
β Bartik IV	0.033***	0.025***	0.042***
(s.e.)	(0.003)	(0.005)	(0.005)
Obs.	89,157	44,944	44,213
KP_F	92.095	27.529	68.104
p_UID	0.000	0.000	0.000

Note: Table displays relevant statistics of the first stage regression of the routine-task intensity index on the Bartik instrument which is used in Table 4.1. KP_F refers to the Kleibergen and Paap (2006) F statistic, p_UID refers to the p-value of the underidentification LM statistic. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v36, O*NET, National accounts data.

As described in Table 4.1, when examining the whole working population, the effect of the RTI on the MCS was not significantly distinguishable from zero. However, the small and insignificant estimate is the result of opposing effects on men and women balancing out each other. For men, performing more routine-intensive tasks leads to better mental health, while for women the opposite is true. At the same time, performing more routine-intensive tasks leads to worse physical health only for men.

Statistical significance of an effect does not necessarily imply its economic relevance. Both, dependent and explanatory variables of this study are standardized measures making the interpretation of the regression results difficult to grasp. To facilitate the interpretation, one can return to the descriptive figures presented in Section 4.2 and make some back-of-the-envelope calculations. According to the right panel of Figure 4.2, a PCS difference of 0.3 standard deviations approximately corresponds to the average difference between the physical condition of a 40-year-old person and a 50-year-old person. It is also close to the average difference found between individuals with normal weight and those with moderate obesity (Doll et al., 2000).[13] Ware et al. (1994) provide further examples which facilitate the interpretation of differences in MCS and PCS. For example, they show that, on average, a depression is reflected in a lower MCS by 1.3 standard deviations.

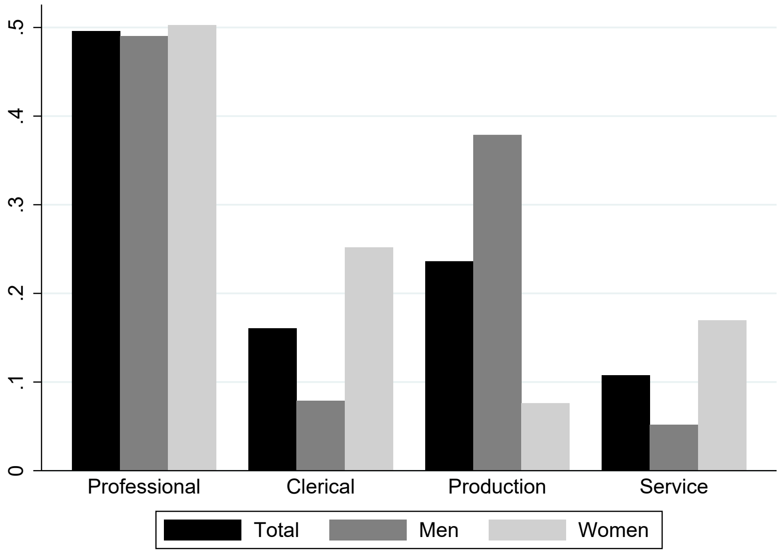
The left panel of Figure 4.3 shows that for women the average RTI score decreased by about one half of a standard deviation during the observation period, for men the decrease was about two thirds of a standard deviation. Multiplying the average decreases with the significant effects in Table 4.1 gives a rough estimate how deroutinization of jobs tasks between 2004 and 2018 affected women's and men's health. On average the MCS of male workers decreased by half a standard deviation,

[13] Moderate obesity refers to a body mass index between 30 and 39.9 and thus describes a person that is 180 cm tall and weighs between 100 and 130 kg.

while the MCS of female workers increased by 0.2 standard deviations due to the decreasing routine task intensity. The PCS of male workers increased by 0.4 standard deviations due to the decline of occupational routine task intensity. Comparing the size of these effects with the mentioned MCS and PCS differences entailed by obesity or depression shows that the effects are meaningful and the revealed differences in the health effects of routine task intensity between male and female workers are large.

This finding raises the question why gender is such an important determinant for the relationship between routine task intensity and health. One obvious reason could be that the types of routine and non-routine tasks that men and women perform differ. The RTI index summarizes an occupation's routine-task intensity in one continuous variable. This concise measure is well-suited as the main explanatory variable for this study, which primarily focuses on routine task intensity. At the same time, using the RTI neglects the nuances of job task profiles as it cannot distinguish between different types of routine tasks. In contrast, grouping workers in occupational classes as in Acemoglu and Autor (2011) enables a more detailed view into job task profiles. Acemoglu and Autor (2011) distinguish between a) professional, technical, managerial occupations that are intensive in non-routine cognitive tasks; b) clerical and sales occupations that are intensive in routine cognitive tasks; c) production and operators occupations that are intensive in routine and non-routine manual tasks; and d) service occupations that are intensive in non-routine manual tasks.

Figure 4.5 shows the shares of workers in the four described occupational classes within the whole workforce and differentiated by gender. About half of male and female workers works in professional occupations. In clerical and sales occupations women are overrepresented, as 25 percent of female workers work in such occupations versus only 8 percent of male workers. In contrast, men are overrepresented in production and operators occupations, with 38 percent of male workers working in such occupations versus only 8 percent of female workers. Lastly, the share of women (men) working in service occupations is at 17 (5) percent. Accordingly, for women, higher routine task intensity as indicated by a high RTI score, more likely means working in clerical or sales occupations and thus performing routine cognitive tasks. For men, a high RTI score is associated with working in production and operators occupations and thus performing routine manual tasks. The different health effects of a higher RTI for men and women are rooted in the fact that they are performing different types of routine tasks. For men, performing routine tasks more likely means performing manual and physical routine tasks. For women, however, performing routine tasks more likely means performing cognitive routine tasks. Consequently, as shown in Table 4.1, a higher RTI leads to worse physical, but better mental health for men, while it leads to worse mental health for women.



Note: Displayed are the 2004 shares of workers in occupational classes following the classification by Acemoglu and Autor (2011) by gender. Source: SOEP v36, O*NET.

Figure 4.5: Shares in occupational classes by gender

Symptoms and Channels The SF-12 survey asks respondents twelve questions to measure eight domains of health. Based on these eight domains or *subscales* the MCS and the PCS are computed (Andersen et al., 2007).[14] To understand in what way the health of male and female workers is affected by performing a job with high routine task intensity, I examine the eight subscales of the SF-12 survey, by using them as dependent variable and repeating the analysis following equations (4.3) and (4.4). All subscales are standardized to have a mean of zero and a standard deviation of 1 in each year. The coefficients yield insights about which symptoms workers are experiencing conditional on the routine task intensity of their job.

Table 4.3 displays the regressions results for each subscale separately for men and women as well as a brief description how the respective health domain is surveyed. The first four subscales are the main factors for the PCS and the bottom four subscales for the MCS.[15] The negative effect on routine task intensity on the PCS for men, which was displayed in Table 4.1, is mostly driven by a higher routine task intensity leading to physical health problems that limit usual everyday role activities. Accordingly, an increase in the RTI by one standard deviation significantly decreases the score of the role-physical scale by around half a standard deviation. Moreover, the positive effect of routine task intensity on the MCS for men is driven

[14] See Figure 4.6 in the Appendix for an overview of the specific questions. See Andersen et al. (2007) and Ware et al. (1996) for an overview how the subscales are computed on single survey items and then used to compute MCS and PCS.

[15] See Andersen et al. (2007) for the precise factor loadings of the eight subscales in PCS and MCS.

by increased vitality. A one standard deviation increase in the RTI accordingly leads to an increase of the vitality score by around half a standard deviation. However, this effect is only significant at the 10-percent significance level. For women, there are significant negative effect of routine task intensity on the role physical and on the bodily pain scale. Accordingly higher routine task intensity leads to more limitations in everyday activities and to more severe physical pain. However, these negative effects do not translate into a an overall significant effect on the PCS score, when taking into account the insignificant coefficients for the other subscales. The negative effect of routine task intensity on the MCS among female workers is mostly driven by a higher RTI leading to worse social functioning and worse mental health, i.e., feeling down and gloomy rather than calm and relaxed.

Table 4.3: Regression of the SF-12 subscales on routine-task intensity

		Men	Women
Phys. functioning	Problems climbing several stairs, lifting heavy	-0.284	-0.130
	objects, demanding everyday activities.	(0.178)	(0.102)
Role physical	Achieving less or being limited in everyday	-0.525**	-0.334**
	activities due to physical health problems.	(0.238)	(0.137)
Bodily pain	Having severe physical pain.	0.089	-0.447***
		(0.222)	(0.139)
General health	Overall current health status.	-0.265	-0.163
		(0.203)	(0.124)
Vitality	Feeling energetic.	0.467*	0.110
		(0.261)	(0.128)
Social functioning	Being limited socially, in contact with	-0.024	-0.498***
	friends, family, or acquaintances	(0.218)	(0.157)
Role emotional	Achieving less or being limited in everyday	-0.014	-0.175
	activities due to mental health problems.	(0.202)	(0.137)
Mental health	Feeling down and gloomy, not feeling	0.330	-0.546***
	calm or relaxed.	(0.239)	(0.142)
Observations		44,944	44,213

Note: Displayed are the coefficients of the RTI analogous to the 2SLS estimation described in equations (4.3) and (4.4), but using the eight SF-12 subscales as dependent variables. All subscales are standardized to have mean zero and a standard deviation of 1. All regressions include a gender dummy, a quadratic in age, and a dummy whether an individual's father obtained a college degree and time fixed effects. In the second stage personal fixed effects are included, so the time-invariant controls drop out. Standard errors are bootstrapped (1,000 runs). First stage statistics are shown in Table 4.2. Sources: SOEP v36, O*NET, National accounts data.

One channel that could drive job-related mental health issues are worries stemming either from potential job loss or the personal economic situation. In particular, worries about job security have been identified as a potential channel through which technological advancement and automation could affect workers' mental health (Abeliansky and Beulmann, 2019; Avdic et al., 2021). The SOEP asks respondents whether they are very concerned, somewhat concerned, or not concerned at all about their job security and about their own economic situation. Based on both questions I create dummy variables, which take the value of 1 if a person is either somewhat or very concerned regarding the respective issue and 0 otherwise. I examine the two variables as dependent variables, using the empirical framework of the main analysis, to investigate whether these worries are growing with routine task intensity and, therefore, represent a channel through which mental health may be affected. I thus estimate linear probability models of the probability of being worried about job security or the own economic situation on an individual's routine task intensity.

Table 4.4: Channels for mental health impact: worries

	Job security		Econ. situation	
	Men	Women	Men	Women
RTI	0.373*	0.149	0.328	-0.043
(s.e.)	(0.211)	(0.118)	(0.203)	(0.103)
Obs.	44,944	44,213	44,944	44,213

Note: Shown are the coefficients analogous to the 2SLS IV estimation described in equations (4.3) and (4.4), but using a dummy variable indicating whether an individual is somewhat or very concerned about job security or her own economic situation as explanatory variable. All regressions include a quadratic in age, and a dummy indicating whether an individual's father obtained a college degree, as well as time fixed effects. In the second stage personal fixed effects are included, so the time-invariant controls drop out. Standard errors are bootstrapped (1,000 runs). First stage statistics are shown in Table 4.2. Sources: SOEP v36, O*NET, National accounts data.

The results are presented in Table 4.4 separately for men and women. None of the coefficients are statistically significant at the 5-percent level. For men there is a positive effect, significant at the 10-percent level. Accordingly, an increase of the routine task intensity by one standard deviation would lead to an increased probability of being worried about their job security by 37 percentage points. So, while there is some indication that male workers in routine-intensive jobs are more likely to worry about their job security, this does not seem to translate into worse mental health since, for men, a higher RTI leads to a higher MCS (see Table 4.1). For women, occupational routine task intensity is not significantly associated with the probability of being worried about job security or the own economic situation. Thus, I do not find support for the hypothesis that the perceived threat of losing the job to automation and the economic consequences thereof drive the negative effect of routine task intensity on mental health among female workers.

4.5 Robustness checks

To assess the robustness of my results, I conduct several robustness checks. For the sake of brevity, I will not present all robustness checks in detail in the main body of the paper, but instead give an overview of the analyses I conducted, refer to the relevant tables in the Appendix, and discuss how they relate to the main results.

Using the top quintile of the RTI index as explanatory variable. The explanatory variable of this study, the RTI, is a standardized, continuous variable. The main specification imposes a functional form that assumes a linear effect along the RTI distribution. To see whether this assumption is justified and additionally make sure that the estimated effects are not exclusively driven by extreme values in the distribution of the RTI, I partition the RTI distribution. More specifically, I generate a dummy variable that takes the value one if an individual's RTI is in the top quintile of RTI scores by year. Then, I use this indicator of being among the top 20% of routine task intensity as the explanatory variable in the empirical framework of the main analysis. Being in the top quintile of routine task intensity should yield effects in the same direction of the main results based on the continuous RTI. The results of this robustness check are displayed in Table 4.7 in the Appendix. For men, the effects are very similar to the results in the main analysis. Being among the top quintile of occupational routine task intensity leads to an increase in the MCS by 0.7 standard deviations, but to a decrease of the PCS by 0.6 standard deviations. For women, the point estimates are larger but very imprecise and insignificant. However, the signs of the point estimates still point in the same direction as the effects estimated in the main analysis. As Table 4.8 shows, the first-stage statistics indicate that the relevance assumption also holds for this robustness check.

Altering the regional analysis unit. The empirical strategy applied in this paper relies on a regional shift-share instrument (Bartik, 1991). The idea is to use the regionally different exposure to the country-wide automation *shock* to instrument workers' occupational task intensity. In the main specification, I use the share of workers in the manufacturing sector on the NUTS-3 level, the 401 German districts, to construct the instrument. As shown in Table 4.2 the instrument does well in predicting an individuals' routine task intensity.

To assess the robustness of my results, and make sure they are not contaminated by regional spillovers, e.g., through commuting between neighboring counties, I alter the regional unit which the applied instrumental variable strategy is based on. Instead of using the share of workers in manufacturing per county, I use the larger units of "regional labor markets" (Kosfeld and Werner, 2012). The 141 regional labor markets are constructed based on German commuting structure and distances. I compute the instrument based on the share of workers in manufacturing in regional labor markets multiplied with the yearly country-wide growth rate of manufacturing and then use this instrument to predict individuals' routine task intensity. The results based on this modified instrument are shown in Table 4.9. The results are similar to the results of the main analysis and the estimated effects only differ in magnitude. Table 4.10 shows that the instrument is relevant even on the regional labor market difference.

Examining alternative health indicators. The SOEP data contain further health indicators, for example, subjective health satisfaction, the number of sick days in a year and the number of doctor visits in the past three month. Significant effects on health should also be reflected in these measures. However, they blur the distinction between mental and physical health, and it is unclear which one of the two is reflected in a specific health indicator more than the other. If the alternative health indicators reflect a combination of mental and physical health one could expect the results to be noisier as mental and physical health are affected quite differently by routine task intensity (see Table 4.1). For women, the indicators should point towards worse health as there is a significant negative effect of routine task intensity on women's mental health and an insignificant relationship with respect to physical health. For men, the picture is more complicated as routine task intensity affects mental and physical health in opposite directions.

The results of the 2SLS regressions using alternative health indicators as dependent variables are displayed in Table 4.11 in the Appendix. For health satisfaction and sick days, the point estimates tend towards worse health and are more strongly pronounced for women, but insignificant for both genders. For men, doctor visits significantly increase with higher routine task intensity, while there is no such effect for women. This could indicate that doctor visits are more closely related to physical than to mental health, as there is also a significant negative effect of routine task intensity on men's physical health and no significant effect on women's physical health.

Ultimately, it is difficult to draw very strong conclusions from the robustness checks based on alternative health measures. This is mainly due to the fact that it is not possible to disentangle physical and mental health using health satisfaction, sick days, or doctor visits, which makes the coefficients imprecise. However, under the assumption that doctor visits are more likely to occur in the case of physical health problems, examining alternative health measures yields results which are very much in line with the gender-specific results of the main analysis. The effects on mental and physical health go in opposite directions. Therefore, they offset each other and are not significantly reflected in health measures that capture both mental and physical health. Only the significant negative effect on men's physical health is reflected in an increased tendency to take up doctor visits.

Excluding imputed values. As described in Section 4.2, I implement corrections to the data in order to address deficiencies in the data collection process. First, in the case of missing occupational task content information, I use the task content assigned to closely related occupations as imputations for the missing values. Second, due to the irregularly updated information across occupations, I compute 3-year moving averages of each job task item, based on which the RTI index is constructed. These corrections appear valid to improve the quality of the database. At the same time, the main results of the analysis should also hold if only the original, uncorrected data are used.

I repeat the main analysis of the paper, but once without using imputed RTI values, and once without computing 3-year moving averages of the RTI. Table 4.12 shows the results of the same empirical strategy as in the main analysis but without using any imputed data, which reduces the analysis sample by 2.2 percent. All significant effects point in the same direction as the main results. Compared to the main results, the point estimates are of larger magnitude, but also the standard errors are larger. The first-stage statistics in Table 4.13 show that the instrument is still relevant even with the reduced sample. Table 4.14 displays the results based on the raw RTI time-series without building moving averages. All estimated effects of this robustness check are very close to the main analysis results, both in size and precision. Also, Table 4.15, shows the first-stage statistics indicating that instrument relevance is given.

Overall, the wide-ranging robustness checks all support the main analysis results and give no reason to question the estimated effects of routine task intensity on mental and physical health.

4.6 Discussion

While this is not the first study examining the association between job content and workers' health, only very few were able to distinguish between mental and physical health. Most studies focus on more general health measures such as health satisfaction which makes them difficult to compare to the present study. For example, Case and Deaton (2005) find that manual workers' health is more rapidly declining over the life cycle even after controlling for selection into manual occupations. Fletcher et al. (2011) find that harsh working conditions negatively affect self-assessed health and that this effect is stronger for women. Similarly, Belloni et al. (2022) find that women's mental health is particularly affected by working conditions.

Other studies analyze specifically the link between automation and health. For example, Nazareno and Schiff (2021) find that automation and the usage of artificial intelligence is associated with less stress but worse general health for workers. Abeliansky and Beulmann (2019) analyze the usage of industrial robots in the manufacturing sector and find that it negatively affects workers' mental health. They argue that the negative effect on mental health is caused by growing worries about job security when robot intensity is increasing. Patel et al. (2018) even find that a higher automation risk is associated with worse general, physical, and mental health.

The present study adds to this strand of literature and underlines the importance of taking into account heterogeneity when it comes to the health effects of automation. In contrast to the studies mentioned above, I can separately assess the gender-specific physical and mental health effects of automation. When considering the whole German workforce, a higher occupational routine task intensity leads to worse physical health, while there is no significant impact on mental health. Thus, overall, one could conclude that deroutinization only has a positive impact on physical health. At the same time, this general statement regarding the workforce as a whole is flawed, as it neglects heterogeneity. When the analysis is conducted separately for men and women, I find large gender-specific differences in the relationship between routine task intensity and health. For men, higher routine task intensity leads to worse physical but better mental health. For women higher routine task intensity leads to worse mental health, while there is no significant effect on physical health. The reason behind this finding is that performing *routine* tasks in many cases means different things for male and female workers. As shown, men are more likely to be working in production or operators occupations, while women are overrepresented in clerical and service occupations. Accordingly, for men, routine tasks more likely are manual tasks, while the opposite is true for women.

My findings complement the work by Black and Spitz-Oener (2010), who show that technological advancement has impacted men's and women's work and their tasks differently, which has contributed to the decrease of the gender wage gap in Germany. According to my results, the differential impact of technological advancement has not only affected men's and women's wages differently. I show that also the health effects of deroutinization are gender-specific. The gender-specific health effects of routine task intensity are of particular importance in the broader context of the continuing deroutinization through technological advancement. In the likely case that technological advancement further drives automation and, thus, the deroutinization of job tasks, the impact on male and female workers' health will continue to diverge if their occupational segregation remains. According to my estimates, deroutinization positively impacts men's physical health as it alleviates the physical strain of physical routine tasks. At the same time men's mental health suffers from the decreasing routine task intensity. For female workers fewer routine tasks leads to better mental health, as the routine cognitive tasks they often perform reduce social functioning and worsen mental health.

The health effects of automation need to be closely assessed. On top of the direct effects on people's well-being, the health effects can also be understood as secondary economic effects of automation. Beyond the direct impact of automation on employment and the distribution of income, positive health effects for men could mean fewer injuries and fewer health-related labor market exits, which would positively impact the public insurance system (Beckmannshagen and Koenig, 2022). At the same time, the negative effects on mental health for men are alarming. Men are reluctant to seek professional help of psychotherapists in case of emotional or relational problems (Addis and Mahalik, 2003), and these behavioral differences are associated with men's lower life expectancy (Courtenay, 2000). Even if men decide to seek professional help, in Germany they face an average waiting period of 20 weeks before psychotherapy starts (Bundespsychotherapeutenkammer, 2018). At that point mental health issues might already have become much more serious, which has severe consequences for the well-being of affected individuals as well as for society as a whole.

4.7 Qualifications and Extensions

This study provides unique evidence of the physical and mental health effects of occupational routine task intensity and emphasizes the role of gender heterogeneity in the context of automation. Analyzing physical and mental health separately for men and women is facilitated by combining the German SOEP data and the US O*NET data. As described in Section 4.2, combining these two data sources comes at the cost of a loss in precision. The original O*NET data contain occupational information based on the O*NET-SOC classification, which then is translated into the ISCO-88 classification used in this study. To make sure that this translation is suitable, and the O*NET data accurately describe the task content of German workers, an extension of this paper could validate the trends of deroutinization across and within occupations based on data from the BiBB/BAuA Employment Survey. At this point the BiBB/BAuA Employment Survey only consists of three waves within the period under investigation in this study. In an extension one would ideally include an additional future wave of the BiBB/BAuA Employment Survey and accordingly extend the investigation period.

The research question of this paper sets a logically consistent boundary for the sample under investigation. As this paper focuses on the health effects of routine task intensity among workers, the analysis sample consists of the active workforce in Germany. It is not within the scope of my analyses to assess the health of individuals who lose their job because of automation. Still, compositional changes in the workforce during the investigation period, such as increased labor market participation of women, could potentially influence the results of my analyses. An extension of this paper could address the issue of selection into employment by implementing a Heckman (1979) selection correction. The implementation of the selection correction hinges on the exclusion restriction, i.e., on finding at least one instrument that influences whether an individual is employed or not but is not correlated with the individual's health. While the implementation of the selection correction is challenging, it would underpin the robustness of the results if they were not substantially altered after the selection correction.

Further, the results of this study are partially not very intuitive due to the nature of the two main dependent variables, both of which are standardized summary scores. Still, when interpreting the results and comparing with other applications of the MCS and PCS, it is clear that the measured effects are meaningful. To make the health impacts of routine task intensity even more tangible, one could additionally use disease diagnoses as outcome variables and estimate the effect of occupational routine task intensity on the probability to be diagnosed with a depression or a physical condition. The SOEP contains information on some diagnoses, but unfortunately the timing of a diagnosis is unclear. Thus, it is not straightforward to apply the same empirical framework that is used in the main analysis by simply substituting the dependent variables with disease diagnoses.

In an extension of the paper, the main explanatory variable, the RTI, could be reviewed as well. My choice to operationalize workers' task content by using the RTI was made after a careful consideration of the advantages and disadvantages of the measure. Using the one-dimensional measure of routine task intensity facilitates the implementation of my empirical framework and the causal identification strategy based on the Bartik-style instrument. Also, focusing on a singular measure of routine task intensity seems adequate in a study that investigates the health effects of automation. It is the repetitiveness of tasks that is pivotal for their automation potential and not the type of routine tasks (Acemoglu and Autor, 2011). At the same time, the types of routine tasks that workers perform seem to be crucial for the effect on their health. The fact that men perform more manual routine tasks, while women perform more cognitive routine tasks likely explains the gender-specific effects found in this study. Future work should take into account the complexity of occupations in more detail by examining the health effects of specific task types instead of only focusing on the routine task intensity.

4.8 Conclusion

This paper examines the impact of occupational routine task intensity on workers' mental and physical health. An understanding of the relationship between routine task intensity and health is vital since through technological progress and automation, the demand for routine task intensive labor is decreasing. By using a Bartik-style instrument based on the regional shares of workers in manufacturing and the manufacturing growth rate, endogeneity issues in the estimation of the effect of routine task intensity on health outcomes are addressed. The results indicate a small but significant effect of routine task intensity on physical health when the whole German workforce is under consideration. Accordingly, an increase in the RTI by one standard deviation leads to a decrease in the PCS by 0.3 standard deviations, while there is no significant effect on mental health.

However, only considering the workforce as a whole neglects the pervasive heterogeneity between men and women. When examining female workers separately, higher routine task intensity leads to significantly worse mental health, while there is no significant effect on physical health. For men, a higher routine task intensity leads to better mental health but worse physical health. The gender-specific effects of occupational routine task intensity are likely due to the fact that male and female workers are performing different types of routine tasks on their jobs. Men are more often performing manual routine tasks in production or operators occupations, while women are more often performing cognitive routine tasks in clerical occupations. Thus, the deroutinization of job tasks driven by technological advancement has different consequences for the mental and physical health of male and female workers. For male workers, deroutinization has positive effects on physical health, as manual tasks exerting physical strain are automated and replaced. At the same time, there are substantial negative effects on men's mental health. Women's mental health is positively impacted by deroutinization. As technological advancement progresses, these interrelations between job tasks and workers' health are important to recognize when occupational safety measures and the availability of adequate healthcare services are discussed.

4.9 Appendix

4.9.1 SOEP Questionnaire

105. How would you describe your current health?

- Very good ☐
- Good ☐
- Satisfactory ☐
- Poor ☐
- Bad ☐

106. When you have to climb several flights of stairs on foot, does your health limit you greatly, somewhat, or not at all?

- Greatly ☐
- Somewhat ☐
- Not at all ☐

107. And what about other demanding everyday activities, such as when you have to lift something heavy or do something requiring physical mobility: Does your health limit you greatly, somewhat, or not at all?

- Greatly ☐
- Somewhat ☐
- Not at all ☐

108. During the last four weeks, how often did you:

	Always	Often	Some-times	Almost never	Never
• feel rushed or pressed for time?	☐	☐	☐	☐	☐
• feel down and gloomy?	☐	☐	☐	☐	☐
• feel calm and relaxed?	☐	☐	☐	☐	☐
• feel energetic?	☐	☐	☐	☐	☐
• have severe physical pain?	☐	☐	☐	☐	☐
• feel that due to physical health problems					
– you achieved less than you wanted to at work or in everyday activities?	☐	☐	☐	☐	☐
– you were limited in some way at work or in everyday activities?	☐	☐	☐	☐	☐
• feel that due to mental health or emotional problems					
– you achieved less than you wanted to at work or in everyday activities?	☐	☐	☐	☐	☐
– you carried out your work or everyday tasks less thoroughly than usual?	☐	☐	☐	☐	☐
• feel that due to physical or mental health problems you were limited socially, that is, in contact with friends, acquaintances, or relatives?	☐	☐	☐	☐	☐

Note: Displayed are the questions of the SF-12 module as contained in the SOEP questionaire. Source: SOEP v36.

Figure 4.6: SF-12 Questions in the SOEP

4.9.2 Additional Tables and Figures

Table 4.5: Overview of yearly sample size and imputations

	Main analysis sample		No imputations		
	Ind. obs.	Distinct occ.	Ind. obs.	Distinct occ.	Share imputed obs.
2004	10,496	279	9777	245	0.074
2006	9,466	283	8,851	252	0.069
2008	9,155	284	8,773	266	0.044
2010	8,677	276	8,359	259	0.038
2012	9,885	271	9,882	269	0.000
2014	14,370	279	14,370	279	0.000
2016	12,972	281	12,972	281	0.000
2018	14,136	314	14,136	314	0.000

Note: Displayed are yearly number of observations and distinct occupations for the main analysis sample, the *raw* sample without imputations, and the share of imputed observations. The table is computed by using the stata command `distinct` Cox and Longton (2008). Sources: SOEP v36, O*NET.

Table 4.6: Alternative empirical strategies

	MCS			PCS		
	(1)	(2)	(3)	(1)	(2)	(3)
Whole sample						
RTI	-0.005	0.069	-0.098	-0.046***	-0.528***	-0.314**
(s.e.)	(0.005)	(0.155)	(0.193)	(0.004)	(0.139)	(0.154)
Obs.	89,157	89,157	89,157	89,157	89,157	89,157
Men						
RTI	-0.005	0.189	0.737*	-0.054***	-1.498***	-0.618**
(s.e.)	(0.006)	(0.284)	(0.404)	(0.005)	(0.366)	(0.310)
Obs.	44,944	44,944	44,944	44,944	44,944	44,944
Women						
RTI	-0.008	-0.002	-0.605**	-0.040***	0.033	-0.132
(s.e.)	(0.007)	(0.181)	(0.241)	(0.006)	(0.149)	(0.182)
Obs.	44,213	44,213	44,213	44,213	44,213	44,213

Note: All regressions include a gender dummy, a quadratic in age, and a dummy whether an individual's father obtained a college degree and time fixed effects. Specification (1) is a standard OLS estimation. Specification (2) is based on a 2SLS estimation using the bartik-style instrument, but without using personal fixed effects. Specification (3) is based on the 2SLS estimation resembling the main analysis (including personal fixed effects). When personal fixed effects are included, the time-invariant controls drop out. Standard errors are robust in (1) and (2), and bootstrapped (1000 runs) in (3). Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v36, O*NET, National accounts data.

Table 4.7: Robustness check: top quintile of routine task intensity as explanatory variable

	MCS			PCS		
	All	Men	Women	All	Men	Women
Q5_RTI	-0.189	0.706**	-2.362	-0.609**	-0.593**	-0.515
(s.e.)	(0.374)	(0.344)	(2.689)	(0.298)	(0.264)	(0.791)
Obs.	89,157	44,944	44,213	89,157	44,944	44,213

Note: Shown are the coefficients analogous to the 2SLS IV estimation described in equations (4.3) and (4.4), but using a dummy variable indicating whether an employee is among the top quintile of routine task intensity in a year as explanatory variable. All regressions include a gender dummy, a quadratic in age, and a dummy whether an individual's father obtained a college degree and time fixed effects. In the second stage personal fixed effects are included, so the time-invariant controls drop out. Standard errors are bootstrapped (1000 runs). First stage statistics are shown in Table 4.8. Sources: SOEP v36, O*NET, National accounts data.

Table 4.8: First-stage statistics—top quintile of routine task intensity as explanatory variable

	All	Men	Women
β Bartik IV	0.033***	0.025***	0.042***
(s.e.)	0.003	0.005	0.005
Obs.	89,157	44,944	44,213
KP_F	92.095	27.529	68.104
p_UID	0.000	0.000	0.000

Note: Table displays relevant statistics of the first stage regression of a dummy indicating whether an individual is among the top quintile of routine-task intensity on the Bartik instrument which is used in Table 4.7. KP_F refers to the Kleibergen and Paap (2006) F statistic, p_UID refers to the p-value of the underidentification LM statistic. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v36, O*NET, National accounts data.

Table 4.9: Robustness check: regional labor markets as regional unit

	MCS			PCS		
	All	Men	Women	All	Men	Women
RTI	0.165	1.406***	-0.654**	-0.202	-0.695**	0.119
(s.e.)	(0.222)	(0.533)	(0.294)	(0.178)	(0.354)	(0.224)
Obs.	89,157	44,944	44,213	89,157	44,944	44,213

Note: Shown are the coefficients analogous to the 2SLS IV estimation described in equations (4.3) and (4.4), but using the IV is constructed based on regional labor markets instead of districts as the regional unit. All regressions include a gender dummy, a quadratic in age, and a dummy whether an individual's father obtained a college degree and time fixed effects. In the second stage personal fixed effects are included, so the time-invariant controls drop out. Standard errors are bootstrapped (1,000 runs). Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v36, O*NET, National accounts data.

Table 4.10: First-stage regression statistics—regional labor markets

	All	Men	Women
β Bartik IV	0.039***	0.032***	0.047***
(s.e.)	0.005	0.006	0.007
Obs.	89,157	44,944	44,213
KP_F	74.221	26.900	48.726
p_UID	0.000	0.000	0.000

Note: Table displays relevant statistics of the first stage regression of the routine-task intensity index on the Bartik instrument (here based on regional labor markets) which is used in Table 4.9. KP_F refers to the Kleibergen and Paap (2006) F statistic, p_UID refers to the p-value of the underidentification LM statistic. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v36, O*NET, National accounts data.

Table 4.11: Robustness check: alternative health indicators by gender

	Health sat.		Sick days		Doctor visits	
	Men	Women	Men	Women	Men	Women
RTI	-0.018	-0.116	5.316	13.692	2.860**	-0.406
(s.e.)	(0.315)	(0.199)	(14.977)	(9.994)	(1.399)	(0.793)
Obs.	44,579	43,871	34,768	34,528	44,120	43,961

Note: Shown are the coefficients analogous to the 2SLS IV estimation described in equations (4.3) and (4.4), but using alternative health indicators as dependent variable. Note that lower observation numbers are due to missing values in the respective dependent variable. All regressions include a gender dummy, a quadratic in age, and a dummy whether an individual's father obtained a college degree and time fixed effects. In the second stage personal fixed effects are included, so the time-invariant controls drop out. Standard errors are bootstrapped (1000 runs). First stage statistics are shown in Table 4.2. Sources: SOEP v36, O*NET, National accounts data.

Table 4.12: Robustness check: no RTI imputations

	MCS			PCS		
	All	Men	Women	All	Men	Women
RTI	-0.534	2.209*	-2.284***	-0.967*	-1.786*	-0.443
(s.e.)	(0.647)	(1.274)	(0.842)	(0.514)	(0.984)	(0.623)
Obs.	87,120	43,458	43,662	87,120	43,458	43,662

Note: Shown are the coefficients analogous to the 2SLS IV estimation described in equations (4.3) and (4.4), but without using any imputed RTI values. All regressions include a gender dummy, a quadratic in age, and a dummy whether an individual's father obtained a college degree and time fixed effects. In the second stage personal fixed effects are included, so the time-invariant controls drop out. Standard errors are bootstrapped (1,000 runs). Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v36, O*NET, National accounts data.

Table 4.13: First-stage regression statistics—no RTI imputations

	All	Men	Women
β Bartik IV	0.010***	0.008***	0.013***
(s.e.)	0.001	0.002	0.002
Obs.	87,120	43,458	43,662
KP_F	81.200	27.411	55.995
p_UID	0.000	0.000	0.000

Note: Table displays relevant statistics of the first-stage regression of the routine-task intensity index excluding imputed values on the Bartik instrument which is used in Table 4.14. KP_F refers to the Kleibergen and Paap (2006) F-statistic, p_UID refers to the p-value of the underidentification LM statistic. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v36, O*NET, National accounts data.

Table 4.14: Robustness check: time-series of RTI without moving averages

	MCS			PCS		
	All	Men	Women	All	Men	Women
RTI	-0.098	0.736*	-0.615**	-0.317**	-0.618**	-0.134
(s.e.)	(0.194)	(0.405)	(0.245)	(0.155)	(0.309)	(0.185)
Obs.	89,157	44,944	44,213	89,157	44,944	44,213

Note: Shown are the coefficients analogous to the 2SLS IV estimation described in equations (4.3) and (4.4), but without constructing a smooth time series of the RTI index by computing moving averages. All regressions include a gender dummy, a quadratic in age, and a dummy whether an individual's father obtained a college degree and time fixed effects. In the second stage personal fixed effects are included, so the time-invariant controls drop out. Standard errors are bootstrapped (1,000 runs). Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v36, O*NET, National accounts data.

Table 4.15: First-stage regression statistics—no moving averages

	All	Men	Women
β Bartik IV	0.033***	0.025***	0.042***
(s.e.)	0.003	0.005	0.005
Obs.	89,157	44,944	44,213
KP_F	89.902	26.099	68.076
p_UID	0.000	0.000	0.000

Note: Table displays relevant statistics of the first-stage regression of the routine-task intensity index without building 3-year moving averages on the Bartik instrument which is used in Table 4.14. KP_F refers to the Kleibergen and Paap (2006) F-statistic, p_UID refers to the p-value of the underidentification LM statistic. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Sources: SOEP v36, O*NET, National accounts data.

List of Tables

List of Figures

Bibliography

Aaronson, D. and E. French (2007). Product market evidence on the employment effects of the minimum wage. *Journal of Labor Economics 25*(1), 167–200.

Aaronson, D. and B. J. Phelan (2017). Wage Shocks and the Technological Substitution of Low-wage Jobs. *The Economic Journal 129*(617), 1–34.

Abeliansky, A. L. and M. Beulmann (2019). Are they coming for us? Industrial robots and the mental health of workers. cege Discussion Papers No. 379.

Acemoglu, D. and D. H. Autor (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In *Handbook of Labor Economics*, Volume 4, pp. 1043–1171. Elsevier B.V.

Acemoglu, D. and P. Restrepo (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review 108*(6), 1488–1542.

Acemoglu, D. and P. Restrepo (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy 128*(6), 2188–2244.

Acemoglu, D. and P. Restrepo (2022). Demographics and Automation. *The Review of Economic Studies 89*(1), 1–44.

Adda, J., C. Dustmann, and K. Stevens (2017). The career costs of children. *Journal of Political Economy 125*(2), 293–337.

Addis, M. E. and J. R. Mahalik (2003). Men, Masculinity, and the Contexts of Help Seeking. *American Psychologist 58*(1), 5–14.

Ahlfeldt, G. M., D. Roth, and T. Seidel (2018). The regional effects of Germany's national minimum wage. *Economics Letters 172*, 127–130.

Altonji, J. and C. Paxson (1992). Labor supply, hours constraints, and job mobility. *Journal of Human Resources 27*(2), 256–278.

Amlinger, M., R. Bispinck, and T. Schulten (2014). Niedriglohnsektor: Jeder Dritte ohne Mindestlohn? Ausnahmen vom geplanten Mindestlohn und ihre Konsequenzen. WSI Report, No. 12.

Andersen, H. H., A. Mühlbacher, M. Nübling, J. Schupp, and G. G. Wagner (2007). Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP Version of SF-12v2. *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts- und Sozialwissenschaften 127*, 171–182.

Angrist, J. D. and J.-S. Pischke (2014). *Mastering 'metrics: The path from cause to effect*. Princeton University Press.

Arbeitskreis „Volkswirtschaftliche Gesamtrechnungen der Länder" (2021). Bruttoinlandsprodukt, Bruttowertschöpfung in den kreisfreien Städten und Landkreisen der Bundesrepublik Deutschland 1992 und 1994 bis 2019. Reihe 2, Kreisergebnisse Band 1. Im Auftrag der Statistischen Ämter der 16 Bundesländer, des Statistischen Bundesamtes und des Bürgeramtes, Statistik und Wahlen, Frankfurt a. M.

Ashenfelter, O. and R. S. Smith (1979). Compliance with the minimum wage law. *Journal of Political Economy 87*(2), 333–350.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the US Labor Market. *American Economic Review 103*(5), 1553–1597.

Autor, D. H., D. Dorn, and G. H. Hanson (2015). Untangling Trade and Technology: Evidence from Local Labour Markets. *The Economic Journal 125*(584), 621–646.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006). The Polarization of the U.S. Labor Market. *American Economic Review 96*(2), 189–194.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics 118*(4), 1279–1333.

Avdic, D., S. C. de New, and D. A. Kamhöfer (2021). Economic downturns and mental wellbeing in Germany. *European Economic Review 140*, 103915.

Bachmann, R., W. Dürig, H. Frings, L. S. Höckel, and F. M. Flores (2017). Minijobs nach Einführung des Mindestlohns – Eine Bestandsaufnahme. *Zeitschrift für Wirtschaftspolitik 66*(3), 209–237.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Başlevent, C. and H. Kirmanoğlu (2014). The impact of deviations from desired hours of work on the life satisfaction of employees. *Social Indicators Research 118*(1), 33–43.

Bassanini, A. and E. Caroli (2015). Is work bad for health? The role of constraint versus choice. *Annals of Economics and Statistics/Annales d'Économie et de Statistique 119/120*, 13–37.

Basu, A. K., N. H. Chau, and R. Kanbur (2010). Turning a blind eye: Costly enforcement, credible commitment and minimum wage laws. *The Economic Journal 120*(543), 244–269.

Baum, C. F., M. E. Schaffer, and S. Stillman (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal 7*(4), 465–506.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pp. 13–68. Springer.

Beckmannshagen, M. and J. Koenig (2022). Out for Good: Labor Market Effects of Transitory and Persistent Health Shocks. SSRN Working Paper 4036482.

Beffy, M., R. Blundell, A. Bozio, G. Laroque, and M. To (2019). Labour supply and taxation with restricted choices. *Journal of Econometrics 211*(1), 16–46.

Bell, D., S. Otterbach, and A. Sousa-Poza (2012). Work hours constraints and health. *Annals of Economics and Statistics/Annales d'économie et de statistique 105/106*, 35–54.

Bellmann, L., M. Bossler, M. Dütsch, H.-D. Gerner, and C. Ohlert (2016). Folgen des Mindestlohns in Deutschland: Betriebe reagieren nur selten mit Entlassungen. IAB-Kurzbericht, 18/2016.

Belloni, M., L. Carrino, and E. Meschi (2022). The impact of working conditions on mental health: Novel evidence from the UK. *Labour Economics 76*, 102176.

Benassi, C. (2011). The implementation of minimum wage: Challenges and creative solutions. Global Labour University Working Paper, No. 12.

Bertrand, M. and S. Mullainathan (2001). Do people mean what they say? Implications for subjective survey data. *American Economic Review 91*(2), 67–72.

Bick, A. (2016). The quantitative role of child care for female labor force participation and fertility. *Journal of the European Economic Association 14*(3), 639–668.

Biewen, M., B. Fitzenberger, and J. de Lazzer (2018). The role of employment interruptions and part-time work for the rise in wage inequality. *IZA Journal of Labor Economics 7*(1), 10.

Biewen, M., B. Fitzenberger, and M. Seckler (2020). Counterfactual quantile decompositions with selection correction taking into account Huber / Melly (2015): An application to the German gender wage gap. *Labour Economics 67*, 101927.

Biewen, M. and D. Plötze (2019). The Role of Hours Changes for the Increase in German Earnings Inequality. *Jahrbücher für Nationalökonomie und Statistik 239*(2), 277–304.

Bispinck, R., H. Dribbusch, and T. Schulten (2010). German collective bargaining in a European perspective: continuous erosion or re-stabilisation of multi-employer agreements? WSI-Diskussionspapier No. 171.

Black, S. E. and A. Spitz-Oener (2010). Explaining women's success: Technological change and the skill content of women's work. *Review of Economics and Statistics 92*(1), 187–194.

Blitzstein, J. K. and J. Hwang (2015). *Introduction to probability (2nd ed.)*. Chapman and Hall/CRC.

Bloemen, H. G. (2008). Job search, hours restrictions, and desired hours of work. *Journal of Labor Economics 26*(1), 137–179.

Blundell, R., J. Britton, M. C. Dias, and E. French (2021). The Impact of Health on Labor Supply Near Retirement. *Journal of Human Resources*, 1217–9240.

Blundell, R. and B. Etheridge (2010). Consumption, income and earnings inequality in Britain. *Review of Economic Dynamics 13*(1), 76–102.

Blundell, R., J. Ham, and C. Meghir (1987). Unemployment and Female Labour Supply. *The Economic Journal 97*, 44–64.

Blundell, R., R. Joyce, A. N. Keiller, and J. P. Ziliak (2018). Income inequality and the labour market in Britain and the US. *Journal of Public Economics 162*, 48–62.

Blundell, R. and T. MaCurdy (1999). Labor supply: A review of alternative approaches. *Handbook of Labor Economics 3*, 1559–1695.

Blundell, R., L. Pistaferri, and I. Preston (2008, 12). Consumption inequality and partial insurance. *American Economic Review 98*(5), 1887–1891.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016, 2). Consumption inequality and family labor supply. *American Economic Review 106*(2), 387–435.

Böheim, R. and M. P. Taylor (2004). Actual and preferred working hours. *British Journal of Industrial Relations 42*(1), 149–166.

Bosch, G. and S. Lehndorff (2005). *Working in the service sector*. Taylor & Francis.

Bossler, M. and H. D. Gerner (2020). Employment Effects of the New German Minimum Wage: Evidence from Establishment-Level Microdata. *ILR Review 73*(5), 1070–1094.

Bound, J. (1991). Self-Reported Versus Objective Measures of Health in Retirement Models. *Journal of Human Resources 26*(1), 106–138.

Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement error in survey data. *Handbook of Econometrics 5*, 3705–3843.

Britton, J. and E. French (2020). Health and Employment amongst Older Workers. *Fiscal Studies 41*(1), 221–250.

Bundespsychotherapeutenkammer (2018). Ein Jahr nach der Reform der Psychotherapie-Richtlinien. Wartezeiten 2018.

Burauel, P., M. Caliendo, A. Fedorets, M. M. Grabka, C. Schröder, J. Schupp, and L. Wittbrodt (2017). Mindestlohn noch längst nicht für alle – Zur Entlohnung anspruchsberechtigter Erwerbstätiger vor und nach der Mindestlohnreform aus der Perspektive Beschäftigter. *DIW Wochenbericht 84*(49), 1109–1123.

Burauel, P., M. Caliendo, M. M. Grabka, C. Obst, M. Preuss, C. Schröder, and C. Shupe (2020). The Impact of the German Minimum Wage on Individual Wages and Monthly Earnings. *Jahrbücher für Nationalökonomie und Statistik 240*(2-3), 201–231.

Burda, M. C. and J. Hunt (2011). What Explains the German Labor Market Miracle in the Great Recession? *Brookings Papers on Economic Activity 2011*(1), 273–319.

Buslei, H., P. Gallego-Granados, J. Geyer, and P. Haan (2019). Rente mit 67: Der Arbeitsmarkt für Ältere wird entscheidend sein. *DIW Wochenbericht 86*(16/17), 275–283.

Caliendo, M., A. Fedorets, M. Preuss, C. Schröder, and L. Wittbrodt (2017). The short-term distributional effects of the German minimum wage reform. IZA Discussion Paper No. 11246.

Caliendo, M., A. Fedorets, M. Preuss, C. Schröder, and L. Wittbrodt (2018). The Short-Run Employment Effects of the German Minimum Wage Reform. *Labour Economics 53*, 46–62.

Caliendo, M., C. Schröder, and L. Wittbrodt (2019). The Causal Effects of the Minimum Wage Introduction in Germany – An Overview. *German Economic Review 20*(3), 257–292.

Capatina, E. (2015). Life-cycle effects of health risk. *Journal of Monetary Economics 74*, 67–88.

Carrillo-Tudela, C., A. Launov, and J.-M. Robin (2021). The fall in german unemployment: A flow analysis. *European Economic Review 132*, 103658.

Case, A. and A. Deaton (2005). Broken down by work and sex: How our health declines. In D. A. Wise (Ed.), *Analyses in the Economics of Aging*, Volume 1, pp. 185–212. University of Chicago Press.

Case, A. and A. Deaton (2021). *Deaths of Despair and the Future of Capitalism*. Princeton University Press.

Cedefop (2013). Quantifying skill needs in Europe. Occupational skills profiles: methodology and application. European Center for the Development of Vocational Training. Research Paper No. 30.

Chang, Y.-M. and I. Ehrlich (1985). On the economics of compliance with the minimum wage law. *Journal of Political Economy 93*(1), 84–91.

Checchi, D., C. García-Peñalosa, and L. Vivian (2016). Are changes in the dispersion of hours worked a cause of increased earnings inequality? *IZA Journal of European Labor Studies 5*(1), 1–34.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011). Adjustment costs, firmresponses, and micro vs. macro laborsupply elasticities: Evidence fromdanish tax records. *Quarterly Journal of Economics 126*(2), 749–804.

Clemens, J. P. and M. Strain (2020). Understanding "Wage Theft": Evasion and Avoidance Responses to Minimum Wage Increases. NBER Working Paper No. 26969.

Cottini, E. and C. Lucifora (2013). Mental Health and Working Conditions in Europe. *ILR Review 66*(4), 958–988.

Courtenay, W. H. (2000). Constructions of masculinity and their influence on men's well-being: a theory of gender and health. *Social science & Medicine 50*, 1385–1401.

Cox, N. J. and G. M. Longton (2008). Speaking Stata: Distinct observations. *Stata Journal 8*(4), 557–568.

Danna, K. and R. W. Griffin (1999). Health and well-being in the workplace: A review and synthesis of the literature. *Journal of Management 25*(3), 357–384.

Dauth, W., S. Findeisen, J. Suedekum, and N. Woessner (2021). The Adjustment of Labor Markets to Robots. *Journal of the European Economic Association 19*(6), 3104–3153.

De Nardi, M., G. Fella, M. Knoef, G. Paz-Pardo, and R. Van Ooijen (2021). Family and government insurance: Wage, earnings, and income risks in the Netherlands and the U.S. *Journal of Public Economics 193*, 104327.

Deardorff, A. V. and F. P. Stafford (1976). Compensation of Cooperating Factors. *Econometrica 44*(4), 671–684.

Department for Business, Innovation and Skills (2010). National Minimum Wage Compliance Strategy. Policy Paper, BIS/10/678.

Deutscher Bundestag (2016a). Drucksache 18/7525: Finanzkontrolle Schwarzarbeit - Kontrolle von Mindestlöhnen 2015.

Deutscher Bundestag (2016b). Drucksache 18/8513: Kontrolle der Einhaltung des gesetzlichen Mindestlohns.

Dickens, W. T. and S. J. Lundberg (1993). Hours Restrictions and Labor Supply. *International Economic Review 34*(1), 169–192.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica 64*(5), 1001–1044.

Dobkin, C., A. Finkelstein, R. Kluender, and M. J. Notowidigdo (2018). The Economic Consequences of Hospital Admissions. *American Economic Review 108*(2), 308–352.

Doll, H. A., S. E. Petersen, and S. L. Stewart-Brown (2000). Obesity and physical and emotional well-being: Associations between body mass index, chronic illness, and the physical and mental components of the SF-36 questionnaire. *Obesity Research 8*(2), 160–170.

Dustmann, C., A. Lindner, U. Schönberg, M. Umkehrer, and P. Vom Berge (2022). Reallocation Effects of the Minimum Wage. *Quarterly Journal of Economics 137*(1), 267–328.

Dustmann, C., J. Ludsteck, and U. Schönberg (2009). Revisiting the German Wage Structure. *Quarterly Journal of Economics 124*(2), 843–881.

Ehrlich, I. (1996). Crime, punishment, and the market for offenses. *Journal of Economic Perspectives 10*(1), 43–67.

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

European Foundation for the Improvement of Living and Working Conditions (2011). European Company Survey 2009: Part-time work in Europe. Luxembourg: European Union, 2011.

Euwals, R. (2001). Female labour supply, flexibility of working hours, and job mobility. *The Economic Journal 111*(471), 120–134.

Euwals, R. and A. Van Soest (1999). Desired and actual labour supply of unmarried men and women in the Netherlands. *Labour Economics 6*(1), 95–118.

Faberman, R. J., A. I. Mueller, and A. Şahin (2022). Has the Willingness to Work Fallen during the Covid Pandemic? NBER Working Paper No. 29784.

Faberman, R. J., A. I. Mueller, A. Şahin, and G. Topa (2020). The shadow margins of labor market slack. *Journal of Money, Credit and Banking 52*(S2), 355–391.

Fadlon, I. and T. H. Nielsen (2021). Family Labor Supply Responses to Severe Health Shocks: Evidence from Danish Administrative Records. *American Economic Journal: Applied Economics 13*(3), 1–30.

Fahr, R. and U. Sunde (2009). Did the Hartz reforms speed-up the matching process? A macro-evaluation using empirical matching functions. *German Economic Review 10*(3), 284–316.

Fairris, D. and L. F. Bujanda (2008). The dissipation of minimum wage gains for workers through labor-labor substitution: evidence from the Los Angeles living wage ordinance. *Southern Economic Journal*, 473–496.

Federal Institute for Research on Building, Urban Affairs and Spatial Development (2015). INKAR: Indikatoren und Karten zur Raum und Stadtenwicklung.

Federal Ministry of Finance (2014). Erlass vom 7. Mai 2014: Neuausrichtung der Aufgabenwahrnehmung im Bereich der Finanzkontrolle Schwarzarbeit (FKS). Decree by the Federal Ministry of Finance. Berlin, 2014.

Federal Ministry of Finance (2017a). Arbeitgeberprüfungen der Finanzkontrolle Schwarzarbeit in 2014. Response to an inquiry based on the Freedom of Information Act. Retrievable at `https://fragdenstaat.de/files/foi/63099/IFG-Schreiben_2017_0191978.pdf`.

Federal Ministry of Finance (2017b). Arbeitgeberprüfungen der Finanzkontrolle Schwarzarbeit in 2016. Response to an inquiry based on the Freedom of Information Act. Retrievable at `https://fragdenstaat.de/files/foi/75745/IFG-Schreiben_2017_0771286.pdf`.

Federal Office for Migration and Refugees (2016). Das Bundesamt in Zahlen 2015. Nürnberg, 2016.

Federal Statistical Office (2017a). Bevölkerung und Erwerbstätigkeit – Schutzsuchende, Ergebnisse des Ausländerzentralregisters. Fachserie 1, Reihe 2.4.

Federal Statistical Office (2017b). Diagnosedaten der Patientinnen und Patienten in Krankenhäusern. Fachserie 12 Reihe 6.2.1.

Federal Statistical Office (2021a). Genesis-online, 23131-0001: Krankenhauspatienten: Deutschland, Jahre, Hauptdiagnose ICD-10 (1-3-Steller Hierarchie).

Federal Statistical Office (2021b). Genesis-online, 23611-0005: Gesundheitsausgaben: Deutschland, Jahre, Ausgabenträger, Erweiterter Leistungsbereich des Gesundheitswesens.

Federal Statistical Office (2021c). Genesis-online, 81000-0133: VGR des Bundes - Einnahmen und Ausgaben der Sozialversicherung: Deutschland, Jahre, Sozialversicherungszweige, Einnahme- und Ausgabearten.

Fedorets, A. and C. Schröder (2019). Economic Aspects of Subjective Attitudes towards the German Minimum-Wage Reform. *FinanzArchiv 75*(4), 357–379.

Fitzenberger, B., K. Kohn, and Q. Wang (2011). The erosion of union membership in Germany: determinants, densities, decompositions. *Journal of Population Economics 24*(1), 141–165.

Fletcher, J. M., J. L. Sindelar, and S. Yamaguchi (2011). Cumulative Effects of Job Characteristics on Health. *Health Economics 20*, 553–570.

Fournier, J.-M. and I. Koske (2013). The determinants of earnings inequality. *OECD Journal: Economic Studies 2012*(1), 7–36.

Friedman, J., T. Hastie, and R. Tibshirani (2017). *The Elements of Statistical Learning*, Volume 1. Springer series in statistics, New York.

Galama, T. and A. Kapteyn (2011). Grossman's Missing Health Threshold. *Journal of Health Economics 30*(5), 1044–1056.

Galvin, D. J. (2016). Deterring Wage Theft: Alt-Labor, State Politics, and the Policy Determinants of Minimum Wage Compliance. *Perspectives on Politics 14*(2), 324–350.

García-Gómez, P. (2011). Institutions, health shocks and labour market outcomes across Europe. *Journal of Health Economics 30*(1), 200–213.

García-Gómez, P. and A. Nicolás (2006). Health shocks, employment and income in the Spanish labour market. *Health Economics 15*, 997–1009.

García-Gómez, P., H. van Kippersluis, O. O'Donnell, and E. van Doorslaer (2013). Long- term and spillover effects of health shocks on employment and income. *Journal of Human Resources 48*(4), 873–909.

Garloff, A. (2019). Did the German minimum wage reform influence (un)employment growth in 2015? Evidence from regional data. *German Economic Review 20*(3), 356–381.

Gerold, S. and U. Stein (2020). Are working hours a new driver of inequality? On the relation between hours inequality, earnings inequality, and collective bargaining in Germany. IMK Discussion Paper No. 203.

Gewerkschaft der Polizei – Zoll (2014). Termin 01.10.2014: Startet die Neuausrichtung der FKS. Newsletter, 1 October, 2014.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik 239*(2), 345–360.

Golden, L. (2016). Still falling short on hours and pay: Part-time work becoming new normal. Report, Economic Policy Institute.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Goldstein, J. R., C. Koulovatianos, J. Li, and C. Schröder (2017). Evaluating how child allowances and daycare subsidies affect fertility. CFS Working Paper Series No. 568.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277.

Goos, M., A. Manning, and A. Salomons (2009). Job Polarization in Europe. *American Economic Review 99*(2), 58–63.

Goos, M., A. Manning, and A. Salomons (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review 104*(8), 2509–2526.

Grabka, M. M. (2020). SOEP-core v35-codebook for the PEQUIV file 1984-2018: CNEF variables with extended income information for the SOEP.

Grabka, M. M. and C. Schröder (2018). Inequality in Germany: decrease in gap for gross hourly wages since 2014, but monthly and annual wages remain on plateau. *DIW Weekly Report 8*(9), 83–92.

Grenier, G. (1982). On compliance with the minimum wage law. *Journal of Political Economy 90*(1), 184–187.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy 80*(2), 223–255.

Haan, P. and M. Myck (2009, 12). Dynamics of health and labor market risks. *Journal of Health Economics 28*(6), 1116–1125.

Haan, P. and K. Wrohlich (2011). Can child care policy encourage employment and fertility? Evidence from a structural model. *Labour Economics 18*(4), 498–512.

Ham, J. C. (1982). Estimation of a labour supply model with censoring due to unemployment and underemployment. *The Review of Economic Studies 49*(3), 335–354.

Hardy, W., R. Keister, and P. Lewandowski (2018). Educational upgrading, structural change and the task composition of jobs in Europe. *Economics of Transition 26*(2), 201–231.

Heathcote, J., F. Perri, and G. L. Violante (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 1967-2006. *Review of Economic Dynamics 13*(1), 15–51.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Henze, P. (2014). Structural change and wage inequality: Evidence from German micro data. Center for European, Governance and Economic Development Research Working Paper 204.

Hosseini, R., K. Kopecky, and K. Zhao (2021a). How Important Is Health Inequality for Lifetime Earnings Inequality? FRB Atlanta Working Paper 2021-1.

Hosseini, R., K. Kopecky, and K. Zhao (2021b). The evolution of health over the life cycle. *Review of Economic Dynamics*.

Jacobi, L. and J. Kluve (2007). Before and after the Hartz reforms: The performance of active labour market policy in Germany. *Journal for Labour Market Research 40*(1), 45–64.

Jann, B. (2019). ISCOGEN: Stata module to translate ISCO codes. Statistical Software Components, Boston College Department of Economics.

Johnson, R. W. and A. T. Lo Sasso (2006). The impact of elder care on women's labor supply. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing 43*(3), 195–210.

Johnson, W. R. and J. Skinner (1986). Labor supply and marital separation. *American Economic Review 76*(3), 455–469.

Kahn, S. and K. Lang (1991). The effect of hours constraints on labor supply estimates. *The Review of Economics and Statistics 73*(4), 605–611.

Kahn, S. B. and K. Lang (1995). The causes of hours constraints: Evidence from Canada. *Canadian Journal of Economics 28*(4a), 914–928.

Karpoff, J. M. and J. R. Lott Jr. (1993). The reputational penalty firms bear from committing criminal fraud. *The Journal of Law and Economics 36*(2), 757–802.

Kemptner, D. (2019). Health-related life cycle risks and public insurance. *Journal of Health Economics 65*, 227–245.

Keynes, J. M. (1930). Economic possibilities for our grandchildren. In *Essays in persuasion*. New York: Norton.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 133*(1), 97–126.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2019). Child penalties across countries: Evidence and explanations. *AEA Papers and Proceedings 109*, 122–26.

Knabe, A., R. Schöb, and M. Thum (2014). Der flächendeckende Mindestlohn. *Perspektiven der Wirtschaftspolitik 15*(2), 133–157.

Kosfeld, R. and A. Werner (2012). Deutsche Arbeitsmarktregionen – Neuabgrenzung nach den Kreisgebietsreformen 2007–2011. *Raumforschung und Raumordnung 70*(1), 49–64.

Lemos, S. (2008). A Survey of the Effects of the Minimum Wage on Prices. *Journal of Economic Surveys 22*(1), 187–212.

Lindeboom, M. and M. Kerkhofs (2009). Health and Work of the Elderly: Subjective Health Measures, Reporting Errors and Endogeneity in the Relationship between Health and Work. *Journal of Applied Econometrics 24*, 1024–1046.

Longmuir, M., C. Schröder, and M. Targa (2020). De-routinization of jobs and polarization of earnings: Evidence from 35 countries. LIS Working Paper Series.

Low Pay Commission (2015). National minimum wage: Low pay commission report 2015. The Stationery Office, Cm 9017.

Low Pay Commission (2017). Non-compliance and enforcement of the National Minimum Wage: September 2017. Report.

MaCurdy, T. (2015). How effective is the minimum wage at supporting the poor? *Journal of Political Economy 123*(2), 497–545.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. In *Proceedings of the National Institute of Science of India*, Volume 12, pp. 49–55.

Mahutga, M. C., M. Curran, and A. Roberts (2018). Job tasks and the comparative structure of income and employment: Routine task intensity and offshorability for the LIS. *International Journal of Comparative Sociology 59*(2), 81–109.

Marcus, J. (2013). The effect of unemployment on the mental health of spouses – Evidence from plant closures in Germany. *Journal of Health Economics 32*(3), 546–558.

Martinez-Granado, M. (2005). Testing labour supply and hours constraints. *Labour Economics 12*(3), 321–343.

McGuinness, S., P. Redmond, and J. Delaney (2020). Minimum wage non-compliance. *Applied Economics Letters 27*(20), 1663–1666.

Mehnert, A., A. d. Boer, and M. Feuerstein (2013, 6). Employment challenges for cancer survivors. *Cancer 119*, 2151–2159.

Metcalf, D. (2008). Why has the British national minimum wage had little or no impact on employment? *Journal of Industrial Relations 50*(3), 489–512.

Meyer, B. D. and W. K. Mok (2019). Disability, earnings, income and consumption. *Journal of Public Economics 171*, 51–69.

MiLoG (2014). Gesetz zur Regelung eines allgmeinen Mindestlohns. Bundesgesetzblatt I, S. 1348.

Minimum Wage Commission (2016). Erster Bericht zu den Auswirkungen des gesetzlichen Mindestlohns.

Minimum Wage Commission (2018). Zweiter Bericht zu den Auswirkungen des gesetzlichen Mindestlohns.

Minimum Wage Commission (2020). Dritter Bericht zu den Auswirkungen des gesetzlichen Mindestlohns.

Mookherjee, D. and A. Shorrocks (1982). A decomposition analysis of the trend in UK income inequality. *The Economic Journal 92*(368), 886–902.

Müller, K.-U. and V. Steiner (2013). Distributional effects of a minimum wage in a welfare state–The case of Germany. SOEPpapers, No. 617.

Nazareno, L. and D. S. Schiff (2021). The impact of automation and artificial intelligence on worker well-being. *Technology in Society 67*, 101679.

Neumark, D., M. Schweitzer, and W. Wascher (2004). Minimum wage effects throughout the wage distribution. *Journal of Human Resources 39*(2), 425–450.

Neumark, D. and W. Wascher (2003). Minimum wages and skill acquisition: Another look at schooling effects. *Economics of Education Review 22*(1), 1–10.

OECD (2016). Income Inequality Update: Income Inequality Remains High in the Face of Weak Recovery. Paris: 2016.

OECD (2019a). *Health at a Glance 2019*. Paris: OECD Publishing.

OECD (2019b). *OECD Employment Outlook: The Future of Work*. Paris: OECD Publishing.

OECD and European Union (2016). The labour market impacts of ill-health. In *Health at a Glance: Europe 2016: State of Health in the EU Cycle*. Paris: OECD Publishing.

Patel, P. C., S. Devaraj, M. J. Hicks, and E. J. Wornell (2018). County-level job automation risk and health: Evidence from the United States. *Social Science and Medicine 202*, 54–60.

Paul, M. (2016). Is there a causal effect of working part-time on current and future wages? *The Scandinavian Journal of Economics 118*(3), 494–523.

Pencavel, J. (2016). Whose preferences are revealed in hours of work? *Economic Inquiry 54*(1), 9–24.

Pezzin, L. E. and B. S. Schone (1999). Intergenerational household formation, female labor supply and informal caregiving: A bargaining approach. *Journal of Human Resources 34*(3), 475–503.

Piketty, T. (2015). Putting distribution back at the center of economics: Reflections on capital in the twenty-first century. *Journal of Economic Perspectives 29*(1), 67–88.

Prowse, V. (2012). Modeling employment dynamics with state dependence and unobserved heterogeneity. *Journal of Business & Economic Statistics 30*(3), 411–431.

Pusch, T. and H. Seifert (2017). Unzureichende Umsetzung des Mindestlohns bei Minijobbern. *Wirtschaftsdienst 97*(3), 187–191.

Rebitzer, J. B. and L. J. Taylor (1995). Do labor markets provide enough short-hour jobs? An analysis of work hours and work incentives. *Economic Inquiry 33*(2), 257–273.

Rentenverischerung, D. (2021). Erwerbsminderungsrenten im Zeitablauf 2021.

Richter, D. and J. Schupp (2015). The SOEP Innovation Sample (SOEP IS). *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts- und Sozialwissenschaften 135*(3), 389–400.

Rick, O., E. M. Kalusche, T. Dauelsberg, V. König, C. Korsukéwitz, and U. Seifart (2012). Reintegrating Cancer Patients Into the Workplace. *Deutsches Arzteblatt 109*(42), 702–708.

Ronconi, L. (2010). Enforcement and Compliance with Labor Regulations in Argentina. *ILR Review 63*(4), 719–736.

Rürup, B. (2002). 5. The German Pension System: Status Quo and Reform Options. In *Social security pension reform in Europe*, pp. 137–170. University of Chicago Press.

Salyers, M. P., H. B. Bosworth, J. W. Swanson, J. Lamb-Pagone, and F. C. Osher (2000). Reliability and validity of the SF-12 health survey among people with severe mental illness. *Medical Care 38*(11), 1141–1150.

Schiele, V. and H. Schmitz (2016). Quantile treatment effects of job loss on health. *Journal of Health Economics 49*, 59–69.

Schmitz, S. (2019). The Effects of Germany's Statutory Minimum Wage on Employment and Welfare Dependency. *German Economic Review 20*(3), 330–355.

Schröder, C., J. König, A. Fedorets, J. Goebel, M. M. Grabka, H. Lüthen, M. Metzing, F. Schikora, and S. Liebig (2020). The economic research potentials of the German Socio-Economic Panel study. *German Economic Review 21*(3), 335–371.

Schurer, S. (2017). Bouncing back from health shocks: Locus of control and labor supply. *Journal of Economic Behavior and Organization 133*, 1–20.

Sellekaerts, B. and S. Welch (1984). An econometric analysis of minimum wage noncompliance. *Industrial Relations: A Journal of Economy and Society 23*(2), 244–259.

Semykina, A. and J. M. Wooldridge (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics 157*(2), 375–380.

Siegers, R., V. Belcheva, and T. Silbermann (2019). Data documentation: Documentation of sample sizes and panel attrition in the German Socio-Economic Panel (SOEP)(1984 until 2017). SOEP Survey Papers 606: Series C. Berlin: DIW/SOEP.

Smith, J. P. (2004). Unraveling the SES: health connection. *Population and Development Review 30*, 108–132.

Smith, T. J. and T. Nichols (2015). Understanding the millennial generation. *The Journal of Business Diversity 15*(1), 39.

Snower, D. J. and C. Merkl (2006). The caring hand that cripples: The East German labor market after reunification. *American Economic Review 96*(2), 375–382.

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics 24*(2), 235–270.

Stechert, M. (2018). Eine kritische Analyse ausgewählter Effekte unter der Einführung des gesetzlichen Mindestlohns in Deutschland. *Wirtschaft und Statistik 3*(2018), 40–53.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics 225*(2), 175–199.

Van Soest, A. (1995). Structural models of family labor supply: a discrete choice approach. *Journal of Human Resources 30*(1), 63–88.

Van Soest, A., I. Woittiez, and A. Kapteyn (1990). Labor supply, income taxes, and hours restrictions in the Netherlands. *Journal of Human Resources 25*(3), 517–558.

Von Gleichen, R. D. and M. Seeleib-Kaiser (2018). Family policies and the weakening of the male-breadwinner model. In *Handbook on Gender and Social Policy*, pp. 153–178. Edward Elgar Publishing.

Wanger, S., T. Hartl, and F. Zimmert (2019). Revision der IAB-Arbeitszeitrechnung. IAB Forschungsbericht 7, 2019.

Ware, J. E., M. Kosinski, and S. D. Keller (1994). *SF-36 Physical and Mental Health Summary Scales: A User's Manual*. Boston, MA: Health Assessment Lab.

Ware, J. E., M. Kosinski, and S. D. Keller (1996). A 12-Item Short-Form Health Survey : Construction of Scales and Preliminary Tests of Reliability and Validity. *Medical Care 34*(3), 220–233.

Weil, D. (2005). Public enforcement/private monitoring: Evaluating a new approach to regulating the minimum wage. *ILR Review 58*(2), 238–257.

Wooden, M., D. Warren, and R. Drago (2009). Working time mismatch and subjective well-being. *British Journal of Industrial Relations 47*(1), 147–179.

Yaniv, G. (1988). Enforcement and monopsonistic compliance with the minimum wage law. *Southern Economic Journal 55*(2), 505–509.

Yaniv, G. (2006). On the employment effect of noncompliance with the minimum wage law. *International Review of Law and Economics 26*(4), 557–564.

English Summary

This dissertation comprises four empirical chapters which contribute to the fields of labor economics, inequality research, and health economics.

The first chapter studies the relationship between the spatial distribution of labor market inspections and non-compliance with Germany's Minimum Wage Law. By combining novel administrative data on labor market inspections with the German Socio-economic Panel (SOEP), we document that the inspection probability is higher in regions with higher non-compliance. This implies a risk-based allocation of inspection efforts and, hence, their endogeneity. Using fixed-effects and an instrumental variable approach, we show that higher inspection efforts have a limited effect on compliance. Based on a theoretical framework and international evidence, we discuss challenges for law enforcement, the political importance of compliance, and potential improvement measures.

The second chapter focuses on inequality in monthly earnings in Germany and the role of desired and actual working hours. We document a significant rise in monthly earnings inequality between 1993 and 2018. The main contributors are inter-temporal increases in working hours inequality and increases in the covariance between working hours and hourly wages, while changes in the distribution of hourly wages play a minor role. We develop a novel double decomposition technique which reveals that these results are particularly pronounced in the growing groups of female employees and service sector employees. If employees had been able to realize their desired optimal working hours, the increase in inequality would have been more moderate. This is mainly because employees with low hourly wages work less than desired, a finding that is reinforced over time—even after controlling for various covariates.

The third chapter investigates the labor market effects of transitory and persistent health shocks. Using machine learning based on sick days and hospitalizations, we derive two novel health shock indicators: one for transitory and one for persistent shocks. By using these new indicators, we overcome issues in the measurement of health, such as heterogeneity and measurement error. In an event study framework, we analyze the respective effects of either shock type on employment, yearly working hours, and labor earnings, but also partner earnings and household net income. Persistent shocks induce large negative employment effects that end up impacting household net incomes. In contrast, transitory shocks induce only minor employment effects that leave household net incomes unaffected. We also investigate effect heterogeneity and find that individuals over 50 years of age are particularly affected by health shocks. Accordingly, persistent health shocks reduce employment of individuals above 50 by 25 percentage points.

The fourth chapter analyzes the effect of occupational routine task intensity on workers' mental and physical health in the context of technological progress and automation driving the deroutinization of job tasks. By combining individual-level health information of German employees with data on occupational task profiles and applying an instrumental variable strategy, I find that male and female workers are oppositely affected by occupational routine task intensity. For women, routine tasks are more likely cognitive routine tasks that negatively affect mental health. For men, routine tasks are more likely manual routine tasks that negatively affect physical health, but have a positive effect on mental health. When considering the overall workforce, the effects on mental health balance out, but a significant negative effect of routine task intensity on physical health remains.

Deutsche Zusammenfassung

Die vorliegende Dissertation besteht aus vier Kapiteln, die Beiträge zu den Gebieten der Arbeitsmarktökonomie, der Ungleichheitsforschung und der Gesundheitsökonomie liefern.

Das erste Kapitel untersucht die den Zusammenhang zwischen der räumlichen Verteilung von Arbeitsmarktkontrollen und Verstößen gegen den gesetzlichen Mindestlohn in Deutschland. Dazu kombiniert es neue administrative Daten zu Arbeitsmarktkontrollen mit Daten des Sozioökonomischen Panels. Es wird gezeigt, dass die Wahrscheinlichkeit einer Arbeitsmarktkontrolle in Regionen mit mehr Mindestlohnverstößen höher ist. Dies lässt auf eine risikobasierte Allokation der Arbeitsmarktkontrollen schließen und impliziert, dass der Zusammenhang zwischen Mindestlohnverstößen und Arbeitsmarktkontrollen endogen ist. Basierend auf einer Fixed-Effects Schätzung sowie einer Schätzung mit Instrumentenvariable zeigen wir, dass eine höhere Kontrolldichte nur wenig Einfluss auf die Einhaltung des Mindestlohns hat. Darauf aufbauend diskutieren wir die politische Bedeutung sowie die Herausforderungen der Durchsetzung des Mindestlohngesetzes unter Einbezug theoretischer Modelle sowie internationaler empirischer Evidenz.

Das zweite Kapitel befasst sich mit der Ungleichheit in Monatseinkommen in Deutschland und beleuchtet dabei die Rolle von gewünschten und tatsächlichen Arbeitszeiten. Wir zeigen, dass sich die Ungleichheit in Monatseinkommen zwischen 1993 und 2018 signifikant erhöht hat. Die Hauptfaktoren hierfür waren ein Anstieg in der Ungleichheit von Arbeitszeiten sowie ein Anstieg in der Kovarianz zwischen Arbeitszeiten und Stundenlöhnen. Dagegen haben Veränderungen in der Verteilung von Stundenlöhnen nur eine geringfügige Rolle gespielt. Wir entwickeln eine neue Doppel-Dekompositionsmethode, mittels derer wir zeigen können, dass diese Ungleichheitstrends besonders ausgeprägt bei weiblichen Beschäftigten sowie Beschäftigten im Dienstleistungssektor sind. Wären die Beschäftigten dagegen in der Lage gewesen ihre gewünschten Arbeitszeiten zu realisieren, wäre der Anstieg der Ungleichheit nur etwa halb so groß gewesen. Dies liegt daran, dass Beschäftigte mit geringen Stundenlöhnen häufig weniger arbeiten als sie gerne würden. Dieser Trend nimmt im Verlauf der Zeit zu – auch nach Einbezug diverser Kontrollvariablen.

Das dritte Kapitel untersucht die Arbeitsmarkteffekte von transitorischen und persistenten Gesundheitsschocks. Mittels eines auf maschinellem Lernen basierenden Algorithmus leiten wir neue Indikatoren für Gesundheitsschocks her: einen für transitorische Schocks, und einem für persistente Schocks. Diese Indikatoren helfen dabei, bei der Definition von Gesundheitsschocks übliche Probleme wie Messfehler und unbeobachtete Heterogenität zu überwinden. Im Rahmen einer Event-Study Methode analysieren wir die Effekte beider Schocktypen auf die Beschäftigung, die geleisteten Arbeitsstunden pro Jahr, auf Bruttoeinkommen der erkrankten Person sowie des Partners oder der Partnerin und auf Haushaltsnettoeinkommen. Persistente Gesundheitsschocks haben große negative Effekte auf die Beschäftigung der betroffenen Personen, die sich auch in reduzierten Haushaltsnettoeinkommen widerspiegeln. Transitorische Schocks haben dagegen nur kleine Beschäftigungseffekte und wirken sich nicht signifikant auf Haushaltsnettoeinkommen aus. Heterogenitätsanalysen zeigen, dass über 50-jährige Beschäftigte besonders stark betroffen sind. Bei dieser Gruppe führt ein persistenter Gesundheitsschock zu einer Reduzierung der Beschäftigungsquote um 25 Prozentpunkte.

Das vierte Kapitel analysiert die Effekte von Routinetätigkeiten im Beruf auf die mentale und physische Gesundheit von Beschäftigten vor dem Hintergrund, dass durch technologischen Fortschritt und immer mehr Routinetätigkeiten automatisiert und somit ersetzt werden. Dazu kombiniere ich Daten zum Gesundheitszustand von Beschäftigten mit Daten zu Tätigkeitsprofilen von Berufen. Unter Anwendung einer Instrumentenvariablenschätzung, zeigt sich, dass die Routineintensität ihrer Berufe unterschiedliche Auswirkung auf die Gesundheit von Männern und Frauen hat. Bei Frauen sind Routinetätigkeiten häufiger kognitive Routinetätigkeiten, die sich negativ auf die mentale Gesundheit auswirken. Dagegen sind Routinetätigkeiten bei Männern häufig manuelle Routinetätigkeiten, die sich negativ auf die physische, aber positiv auf die mentale Gesundheit auswirken. Betracht man alle Erwerbstätigen gemeinsam, gleichen sich die Effekte auf die mentale Gesundheit aus, während ein signifikanter negativer Effekt der Routineintensität von Berufen auf die physische Gesundheit bestehen bleibt.

Erklärung

Erklärung gemäß §4 Abs. 2

Hiermit erkläre ich, dass ich mich noch keinem Promotionsverfahren unterzogen oder um Zulassung zu einem solchen beworben habe, und die Dissertation in der gleichen oder einer anderen Fassung bzw. Überarbeitung einer anderen Fakultät, einem Prüfungsausschuss oder einem Fachvertreter an einer anderen Hochschule nicht bereits zur Überprüfung vorgelegen hat.

(Ort, Datum, Unterschrift)

Erklärung gemäß §10 Abs. 3

Ich habe meine Dissertation soweit im Folgenden nicht anders vermerkt selbständig verfasst.

Folgende Hilfsmittel wurde benutzt

- Statistik und Mathematik: Stata, Excel
- Schriftsatz und Formatierung: LaTeX

(Ort, Datum, Unterschrift)