Essays in Empirical Economics

Britta Rude

A thesis presented for the degree of
doctor oeconomiae publicae

München 2022

Essays in Empirical Economics

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

2022

vorgelegt von
Britta Rude

Referent: Prof. Panu Poutvaara, Ph.D.
Korreferent: Prof. Helmut Rainer, Ph.D.
Promotionsabschlussberatung: 1. Februar 2023

Acknowledgements

I thank Panu Poutvaara, my supervisor, for making me part of the ifo team and allowing me to explore the topics I am most passionate about. I am grateful for many invaluable comments, helpful feedback, and the support Panu provided me on my thesis. I am most grateful to Helmut Rainer, my second supervisor, for his invaluable feedback on this thesis. I thank my co-authors Michele Battisti and Ilpo Kauppinen for the joint work on the second chapter of this dissertation. Thank you for always listening to me and for the unwavering support throughout my doctoral programme. I am grateful to Claudia Steinwender for being an inspiring professor and serving as a third advisor on this thesis.

Acknowledgement goes to my colleagues, Sarah Reiter, Christa Hainz, Tanja Stitteneder, Carla Rhode, Mirely Kollmannsberger and Clara Albrecht, who have all contributed to supporting me at ifo over the years and providing feedback. I would like to express my gratitude to Yvonne Giesing for inviting me to the Pillars Research Consortium and feedback on the third essay of this dissertation. I thank the administrative and printing staff at ifo for their support. I thank the following friends and colleagues for valuable feedback and proofreading: Cristina Rujan, Madhinee Valeyatheepillay, Regina Mestre, Marcela Monsalve Rincon. I am grateful to Andreas Peichl, Oliver Falck, and Carsten Eckel for being inspiring professors and providing me with guidance. I am grateful for the feedback provided by participants in conferences and seminars.

The first chapter of my dissertation uses raw data provided by the Chilean Agency for Quality Assurance in Education. I thank the Chilean Agency for Quality Assurance in Education for allowing me access to the data. The results shown here are the author's and in no way those of the Chilean Agency for Quality Assurance in Education.

The second chapter of my dissertation uses data provided by the Institute for Social Research at the University of Michigan (ICPSR) from the National Incident-Based Reporting System. I thank the ICPSR for making this data publicly available to the broader research community. The research project in the second chapter also uses data retrieved from the Twitter API for Academic Research, which was accessible to the research community at no costs during the time I conducted this research. I thank Twitter for granting me access to their data. Moreover, I use the Tweepy Python Library among several other python packages and would like to thank their developers for making these packages accessible.

Regarding the third chapter of my dissertation, I thank the LMU-ifo Economics and Business Data Center for providing access to the IFR data. Research conducted as part of the third chapter received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement number 101004703. The data access to the SIAB was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently remote data access.

I express my gratitude to my colleagues at the World Bank, who have formed me as a young economist, especially Monica Robayo-Abril, Julian Lee, Juan Diego Alonso, Francis V. Fragano, Yoanna Kraus, Robert Davis, Ricardo Larrobla, and Ruth González Llamas. Thank you for your guidance and support, and everything you taught me along the way. I would like to thank the Carlo Schmid Programme for financing my stay with the World Bank in Paraguay. Thanks go to Adrianna Mancuello, Juliana Chen, and Flavia Sacco who all supported me during my time in Paraguay.

I would like to thank the Musi family for receiving me in their home back in 2008 and many inspiring conversations that significantly influenced my journey. I am grateful to all the people I met through Aguablanca e.V. and my time in Cali, Colombia, which inspired great parts of my dissertation. I am thankful for being a scholar of the Regional Academy on the United Nations during my doctoral programme, and my colleagues Chelsea Couture and Anna Arias Duart.

My dissertation would not have been possible without the support I received from my friends around the world, and Sergio who has been my rock during this time. Special thanks go to my brother, my cousins, and Maria Teresa who are always there for me. Lastly and most importantly, I thank my mum and my dad, who always support me and who opened all possible doors for me.

Contents

List of Figures

List of Tables

Introduction

"I believe that virtually all the problems in the world come from inequality of one kind or another." - Amartya Sen

Motivation

In 2018, more than 4 out of 10 people lived in poverty[1] and nearly every tenth person lived in extreme poverty[2] (The World Bank 2021). While poverty rates have fallen during the last decades, many are still trapped in poverty, with limited possibilities for better living conditions. In addition, inequality around the globe remains high. In 2021, the richest 10 percent held more than half (52.3 percent) of global pre-tax national income, a value similar to the one observed in 1985 (World Inequality Database 2022). For the richest one percentile, this share has even increased over time (from 16.6 percent back in 1985 to 19.2 percent in 2021).

Understanding and approaching poverty and inequality is a complex task given that there are many drivers behind them. Traditionally, researchers took a pure income perspective on poverty and inequality, which mainly targeted the poor via the social safety net and temporary alleviation strategies, but then increasingly moved towards more comprehensive approaches (Attanasio and Székely 1999). Starting with Attanasio and Székely (1999) the focus shifted towards an asset-based understanding of poverty and inequality based on three different channels: human capital, physical capital, and social capital. In line with this work, Carter and Barrett (2006) and McKague, Wheeler, and Karnani (2015) apply similar models to gain a deeper understanding of poverty and emphasize that there are many structural drivers involved.

Approaching poverty and inequality is complex, not only because of their multidimensional nature, but also given that poverty alleviation strategies involve a variety of different

[1]The poverty line is the international poverty line of 5.5 US-Dollars per day (at 2011 international prices).

[2]The extreme poverty line is the international poverty line of 1.90 US-Dollars a day (at 2011 international prices).

stakeholders. McKague, Wheeler, and Karnani (2015) define three different stakeholders relevant to the poor: the private sector, governments, and civil societies. According to their model, each of these players can assume targeted but interconnected roles and responsibilities when addressing poverty and inequality.

My dissertation connects to this overarching literature on poverty alleviation and inequality by taking a deep dive into all three channels of the asset-based framework by Attanasio and Székely (1999) and connecting them to each of the three stakeholders identified by McKague, Wheeler, and Karnani (2015). By the work conducted in my dissertation, I hope to shed light on how to potentially address some of the multidimensional aspects of inequality.

To start with, my work presented in Chapter 1 is located within the human capital channel of the asset-based framework and explores the potential role that governments can play in addressing inequalities related to human capital. In the first chapter, I analyze the impact of a pioneer comprehensive early childhood development program implemented by the government of Chile. My motivation for the work in this chapter is based on substantial evidence showing that the first years of a child's life lay the basis for sustainable development (Daelmans et al. 2017). Early childhood development is therefore a crucial pillar to address poverty and inequality.

Next, my work in Chapter 2 connects to the social capital channel of the asset-based framework. I investigate the effect of social movements against gender-based violence (GBV) that take place on Twitter on GBV-related crime rates. GBV has economic relevance as it generates large costs for both individuals and for the society as a whole. The World Bank estimates that costs related to GBV amount to one to two percent of gross domestic product (GDP) (Duvvury et al. 2013). Nevertheless, to date, GBV is a largely understudied topic. The research question in my second dissertation chapter explores one of the potential roles that civil societies can assume to address structural inequality, in this case GBV, and consequently contributes to closing this knowledge gap.

Lastly, Chapter 3 speaks to the physical capital channel of the asset-based model and to the potential responsibilities of the private sector in addressing poverty and inequality. Chapter 3 is about the impact of automation on migration flows and on labor market outcomes of foreigners and natives. This work is motivated by evidence showing that most OECD countries exhibit a wage gap between migrants and natives due to differences in labor productivity as well as outside options of natives (Battisti et al. 2018). In addition, technological change might affect certain skill groups differently than others (Acemoglu and Restrepo 2018b). Consequently, it is relevant from an equality perspective to study how automation affects labor market outcomes and migration dynamics of foreigners and natives.

Outlook of Thesis

In Chapter 1, I investigate whether comprehensive and universal early childhood development programs are effective tools to address human capital gaps. I exploit the birth eligibility cutoff of a pioneer early childhood intervention in Chile, called *Chile Crece Contigo* (hereinafter ChCC), and assess its impact on educational outcomes in middle childhood. I use administrative data on grade point averages, standardized test scores of all 4th graders, and an extensive early childhood development survey. Program exposure raises standardized math scores by 0.3 percent, standardized reading scores by 0.8 percent, and grade point averages by 0.5 percent. However, the effect is less pronounced for girls and socioeconomically vulnerable children. Several of the intended channels of the program seem to be ineffective. A cost-benefit analysis suggests that targeted programs might be more cost-efficient in addressing human capital gaps.

This chapter makes a contribution to the literature that studies the impact of social safety nets. A large body of literature investigates the impact of having access to social safety nets on infant health (Almond, Hoynes, and Schanzenbach (2011); Hoynes, Page, and Stevens (2011); Amarante et al. (2016); Goodman-Bacon (2018); Ko, Howland, and Glied (2020)). Related work focuses on alternative child development outcomes (Milligan and Stabile (2011); Akee et al. (2018)) and others analyze long-term effects of social safety nets (Chetty et al. (2011); Hoynes, Schanzenbach, and Almond (2016a); Deming (2009); Bailey, Sun, and Timpe (2021); Akee, Jones, Simeonova, et al. (2020)). My work also connects to the literature studying the effect of policy interventions for children, such as the Perry Highschool Project (Heckman et al. (2010); Heckman, Pinto, and Savelyev (2013)), or the ABC/CARE program (García, Heckman, and Ziff 2018). Moreover, it is related to a number of papers asking how to improve schooling outcomes (see for example Duflo (2001) or Black et al. (2014)).

My work speaks to this literature by analyzing the impact of a comprehensive and universal early childhood development program. To date, there is limited evidence on the effect of these types of programs. To the best of my knowledge, there is only one paper so far that studies the impact of ChCC. Work by Clarke, Méndez, and Sepúlveda (2020) demonstrates positive effects of the program on birth weight and other early human capital outcomes.

Given that ChCC was one of the first comprehensive and universal early childhood developing (ECD) programs designed, the first chapter of my dissertation generates valuable insights on how to best close human capital gaps early in life. In this chapter, I investigate if predictions made by experts on the effectiveness of multi-sectoral programs (Richter et al. 2017) hold. The chapter also addresses a so-called "missing middle" within the literature on early childhood development, referring to a lack of knowledge about how early and middle childhood interact (Almond, Currie, and Duque 2018). Lastly, my work generates empirical evidence on theoretical predictions made by Heckman (2006) and Cunha and Heckman (2007).

The evidence raised in the first chapter of my dissertation has important implications for policymakers. First, it demonstrates that, while comprehensive and universal early childhood interventions lead to improved schooling outcomes, these types of programs have several flaws. In comparison to other types of programs, the Marginal Value of Public Funds (MVPF) is low, indicating that targeted programs might be more cost-efficient to address human capital gaps early in life. Chapter 1 also provides evidence stressing the importance of piloting early childhood development programs. Lastly, policymakers need to improve the design of such programs so that girls and socioeconomically vulnerable children benefit equally from them.

In the second chapter of my dissertation, I ask whether social movements against GBV that take place on Twitter affect GBV-related crime rates. Using Twitter data and machine learning methods, me and my co-authors construct a novel data set on the prevalence of Twitter conversations about GBV. We then link this data to information on crime reports and arrests at the federal state by week level in the United States provided by the FBI. We take advantage of the high-frequency nature of our data and an event study design to establish a causal impact of Twitter social movements on GBV. Our results point out that Twitter discussions on GBV lead to a decrease in reported crime rates. The evidence shows that perpetrators commit these crimes less due to increased social pressure and perceived social costs. We also find a significant increase in the arrest per crime rate. The results indicate that social media could significantly decrease reported GBV.

The second chapter of my dissertation makes a valuable contribution to two overarching strands of the economic literature. First, it contributes to the economic literature studying GBV. One branch of literature within the economics of GBV studies potential strategies to reduce GBV. Previous research analyzes the potential of public transfer programs (Bobonis, González-Brenes, and Castro 2013), anti-poverty programs (Amaral, Bandyopadhyay, and Sensarma 2015), the employment of female police officer (Miller and Segal (2019); Amaral, Bhalotra, and Prakash (2021)), or the exposure to video dramas (Cooper, Green, and Wilke 2020). Morrison, Ellsberg, and Bott (2007) give an early literature review on potential interventions.

Another branch of literature within the economics of GBV analyzes potential drivers of GBV. To name a few examples, Aizer (2010) shows that a decreasing wage gap comes along with a decrease in domestic violence at the household level in the US. Related work by

Bhalotra et al. (2021a) illustrates that an increase in male unemployment or a decrease in female unemployment increases intimate partner violence (IPV). Similarly, Brassiolo (2016) demonstrate that a decrease in divorce costs leads to a decrease in IPV. Closely related work by González and Rodríguez-Planas (2020) finds that gender norms are important drivers of IPV. Related work analyzes the association between polygony and IPV (Cools and Kotsadam 2017), family structures and IPV (Tur-Prats 2019), and colonialism and IPV (Guarnieri and Rainer 2021). There is also an increasing literature studying the impact of COVID-19 related lock-downs on IPV (Agüero (2021), Berniell and Facchini (2021); Bullinger, Carr, and Packham (2021)).

A related subfield of papers investigates the impact of GBV (Welsh (1999); Chakraborty et al. (2018); Fitzgerald and Cortina (2018); Folke et al. (2020); Mishra, Mishra, and Parasnis (2021); Siddique (2022)).

Our work brings a new aspect to this literature by analyzing the impact of social media movements. We argue that Twitter might act as a facilitator for the signaling of shifting social norms. This channel would be in line with previous evidence on the erosion of existing social norms, such as work by Bursztyn, Egorov, and Fiorin (2020) showing significant effects of Donald Trump's rise in polarity on publicly expressed xenophobic views.

This rationale is related to the literature on social movements (Besley and Ghatak (2018); Francois and Vlassopoulos (2008); Bénabou and Tirole (2006)) and social norms (Agranov, Elliott, and Ortoleva (2021); Viscusi, Huber, and Bell (2011)). One particular type of social movements studied more extensively in the political economy literature are political protests (Bremer, Hutter, and Kriesi (2020); Bursztyn et al. (2021); Matta, Bleaney, and Appleton (2021)). We contribute to this literature by showing that social movements in online spaces translate into offline behavioral changes, indicating that Twitter is a facilitator of the signaling of social norms. These findings are in line with a limited number of studies illustrating a significant association between social media usage and hate crimes (Müller and Schwarz 2020) as well as political outcomes (Levy (2021); Zhuravskaya, Petrova, and Enikolopov (2020)). Also, recently, several papers have analyzed the impact of the *#blacklivesmatter* movement, another movement connected to social media usage (Dave et al. (2020); Agarwal and Sen (2022)).

The findings of my second dissertation chapter generate relevant policy insights. First, policymakers should use social media platforms as a tool to signal social norms in favor of gender equality. They can also purposefully use these platforms to create informal support networks and connect advocates of gender equality to each other. Second, our work presents novel evidence on the importance of stigmatization and tabooing for the reporting of GBV. We draw this insight from analyzing the impact of social movements against GBV on Twitter on different types of GBV, such as sexual, physical and emotional violence. Our findings from this analysis suggest that policymakers should design strategies to address harmful gender norms, especially within the authorities. When training police forces, it is crucial to raise awareness about the barriers to the reporting of GBV. Lastly, although we do not find evidence in favor of backlash behavior, institutions should still secure environments safe of backlash cultures, especially within the police, to impede GBV.

In Chapter 3, I address the question if automation affects migration patterns. To answer this question we analyze the impact of artificial intelligence (AI) on internal migration dynamics as well as immigration inflows from abroad. For this purpose, we harness online job vacancy data as well as matched employer-employee data. We apply a novel shift-share instrument and determine an increase in net internal migration inflows for German citizens as a response to high-skilled task automation. The effect is less pronounced for foreign citizens. Surprisingly, there is no education gradient in geographic responsiveness to automation of high-skilled tasks and no significant increase in immigration inflows from abroad. A comparison to automation of low-skilled tasks, which we approximate by the adoption of industrial robots, shows that the impact of automation differs by the skill types that technologies likely replace.

The last chapter of my dissertation makes an important contribution to the literature studying labor market effects of technological change. On the effect of AI, Acemoglu et al. (2020) find that AI has not yet had any significant aggregate labor market effects, while Webb (2019) predicts a decrease in inequality through replacement effects on the high-skilled. In contrast to that, Felten, Raj, and Seamans (2019) show that AI could exacerbate current levels of inequality as it leads to an increase in wages of high-skilled occupations. Finally, Alekseeva et al. (2021) document an increased skill demand of AI in the US and a wage premium for these jobs. We contribute to this literature by studying if AI could potentially result in skill shortages, which companies address by recruiting workers from outside their local labor market, either from abroad or from within Germany. In addition, our study provides insights regarding the extent to which emerging technologies trigger internal migration dynamics. Additionally, we explore the heterogeneity of labor market effects with respect to the origin of employees by differentiating between natives and foreigners. This contribution sheds further light on underlying drivers behind labor market effects of technological change.

The last chapter of my dissertation is closely connected to several papers tying the topic of technological change to migration economics. Hanson (2021), for example, illustrates that foreign-born workers have accounted for more than half of the job growth in AI-related occupations since 2000. Our paper sets apart by studying the reversed causality of this question. Additionally, several other papers study the interaction between different forms of technologies and migration. Basso, Peri, and Rahman (2020) study the effect of computerization on immigration and show that newly arrived immigrants specialize in manual service occupations. Their findings indicate that immigrants attenuate job and wage polarization faced by natives owing to computerization. Hanson (2021) shows that an increase in the supply of high-skilled immigrants leads to an increase of AI in local labor markets in the United States. Moreover, Beerli, Indergand, and Kunz (2021) demonstrate that a higher exposure to Information and Communication Technology (ICT) leads to a significant inflow of high-skilled immigrants. For Germany, Danzer, Feuerbaum, and Gaessler (2020) analyze the effect of immigrant inflows on innovation, and find that they reduce innovation.

In addition, our work speaks to the literature that studies the effect of migration on innovation. Hunt and Gauthier-Loiselle (2010) find that immigrants patent at double the rate of natives. Related work by Peri and Sparber (2011) shows that immigration influences the specialization of the native population. Research by Lewis (2011) suggests that firms could see low skilled foreigners and automation machinery as substitutes.

The last chapter of my dissertation is also connected to the literature studying drivers and consequences of internal migration. Work by Piyapromdee (2021), for example, demonstrates that internal migration is a mitigation mechanism for immigration inflows. Similarly, internal migration potentially mitigates the impact of economic downturns (Cadena and Kovak 2016).

The main contribution of Chapter 3 is threefold. First, it generates additional evidence on Germany, a highly automated economy. Second, it brings a new aspect, namely the migration aspect, to the literature studying labor market effects of technological change. Lastly, it combines a rich set of novel data sets to study the underlying research question, and thus enhances empirical research strategies applied in previous studies.

Conclusion

My dissertation shows that all three stakeholders identified by McKague, Wheeler, and Karnani (2015) can be powerful actors in addressing inequalities and structural drivers of poverty. Chapter 1 makes the case for governments' investments in early childhood development. Chapter 2 shows that the civil society can significantly impact the structural drivers of inequality, in this case GBV. Lastly, Chapter 3 illustrates that the adoption of technologies in the private sector can have important equity implications for foreigners and natives.

Although the findings of my dissertation showcase possible channels to address the structural drivers of poverty and inequality, my results also illustrate that there is a need for differentiated approaches. The fact that ChCC led to lower effects for girls and the socioeconomically vulnerable population is troublesome. Similarly, my findings on social movements that take place on Twitter show that effects differ by types of GBV. Lastly, the last chapter of my dissertation illustrates that the impact of technologies differs by citizenship and skill types. These results bolster the case for differentiated evaluations and careful, evidence-based assessments of structural drivers behind poverty and inequality.

Chapter 1

Middle-run Impacts of Comprehensive Early Childhood Interventions: Evidence from a Pioneer Program in Chile

"The only international language in the world is a child's cry." - Eglantyne Jebb

1.1 Introduction

There is substantial evidence showing that the first years of a child's life lay the basis for sustainable development (Daelmans et al. 2017). As documented by Currie and Almond (2011) it is these early years during which children lay the foundation for their human capital accumulation and labor market outcomes during the rest of their lives. At the same time, estimates by the World Health Organization (2020) find that 250 million children, or 43 percent of all children from low- and middle-income countries, were unable to fulfill their full development potential in 2016. It is crucial to study how we can most effectively close these development gaps.

While there is an increasing literature studying the potential of early childhood interventions to close these development gaps, the focus is mainly on targeted programs.[1] However, several scholars have recently pointed out the need for universal and comprehensive early childhood development (ECD) programs (see for example Richter et al. (2017), Black et al. (2017), and Daelmans et al. (2017)). Comprehensive ECD programs are interventions that incorporate multi-sectoral entry points, which include a variety of coexisting components, such as health, nutrition, security and safety, responsible care, and early learning.[2] To date, there is limited evidence on the effectiveness of these types of programs.

This paper tries to close this evidence gap and asks if universal and comprehensive early childhood interventions are effective tools to decrease human capital gaps in developing countries.[3] To study this question, I exploit the roll-out of a pioneer early childhood intervention of this nature in Chile - called *Chile Crece Contigo*[4] (hereinafter ChCC) - which was introduced in 2007. I study the program's impact on educational outcomes in middle childhood. ChCC is a pioneer program based on two characteristics. First, the program is universal, meaning that it targets the full universe of children in Chile. Children are integrated into the program from the first prenatal checkup until they enter the school system. Second, ChCC takes a comprehensive approach. This means that ChCC offers a variety of services to children and their families, such as access to the public health system, technical help, nurseries, kindergartens, and the *Chile Solidarity* program, which supports vulnerable families. ChCC gives families preferential access to the whole network of social services of the State.

[1] Examples include the Perry Preschool program (Heckman et al. 2010), the Jamaica study (Gertler et al. 2014), the Abecedarian experiment (Campbell et al. 2014), or targeted programs in Colombia (Attanasio et al. 2020).

[2] The researchers highlight that families who cannot provide their children with the necessary input to reach their developmental potential need support. This support should consist of materials, financial resources, knowledge, time and professional assistance, as well as protection, prevention and education.

[3] For previous versions of this chapter and the underlying working papers see Rude (2022a) and Rude (2022b).

[4] In English: Chile Grows With You.

Thus, ChCC implemented an approach which was recently identified as a best practice for child development in the literature already back in 2007. I can therefore examine whether these best practices do indeed work as predicted by experts. To the best of my knowledge, this is the first paper to study the overall impact of ChCC on child development in middle childhood.

To analyze the underlying research question, I exploit the birth eligibility cutoff of the program and assess ChCC's effects on educational outcomes 12 years after its introduction. I apply a regression discontinuity approach by matching the date of ChCC's introduction by municipality with children's dates of birth and places of residence. I create a staggered eligibility threshold based on treatment variation by county and birth cohort. This staggered eligibility threshold allows me to compare children born before and after the program's roll-out. I use a variety of rich datasets. To measure educational achievement, I rely on administrative data on grade point averages for the entire student population in Chile. I additionally look at standardized test scores in reading and math of all 4th graders in Chile.

I find that the program has a positive effect on educational outcomes. Program exposure increases standardized math scores by 0.3 percent, standardized reading scores by 0.8 percent and grade point average by 0.5 percent. However, effects are more marked for boys than for girls. The impact is also smaller for socioeconomically vulnerable children. To assess the cost-efficiency of the program, I conduct a cost-benefit-analysis. I find that ChCC's marginal value of public funds (MVPF) is lower than the MVPF for early childhood interventions in the US.

My paper contributes to the literature studying effects of having access to social safety nets during early childhood. While one literature stream analyzes the impact on health outcomes early in life[5], a large number of papers assess long-term effects[6]. My paper talks to this literature by exploring the potential of multi-sectoral early childhood interventions. Additionally, most of the work in this field focuses on the developed world.[7] I generate evidence on impacts of ECD programs in developing countries. Another contribution of this paper to the existing literature on ECD is that it analyzes outcomes in middle childhood. Almond, Currie, and Duque (2018) identify a lack of research focusing on middle childhood within this field. My work helps to close this gap by showing that there is a significant interaction between early childhood development and development outcomes in middle childhood.

[5]See for example work by Ludwig and Miller (2007), Almond, Hoynes, and Schanzenbach (2011), Hoynes, Page, and Stevens (2011), Amarante et al. (2016), Goodman-Bacon (2018), Ko, Howland, and Glied (2020).

[6]Example studies conducted by Chetty et al. (2011), Campbell et al. (2014), Hoynes, Schanzenbach, and Almond (2016b), Akee, Jones, Simeonova, et al. (2020), Bailey et al. (2020) and Bailey, Sun, and Timpe (2021) give a great entry to the topic.

[7]One paper by Amarante et al. (2016) analyzes the effect of the PANES program on birth outcomes in Uruguay.

Studying ChCC is of high policy-relevance since it has been the basis for the design of several similar programs in numerous countries.[8] It is also one of the showcase models used by international organizations (Richter et al. 2017). Chile offers a relevant context for the underlying research question as it is a benchmarking country for Latin American countries, but also other developing countries.[9] To date, there is only one other paper studying the effect of ChCC, that of Clarke, Méndez, and Sepúlveda (2020). Unlike my work, they study the effect of one specific component of *Chile Crece Contigo* on different health variables at birth. They find that the health components of ChCC have a positive effect on birth weight and the rate of fetal deaths. In fact, these improvements in early health outcomes could partly drive the positive impact on educational outcomes in middle childhood observed in this paper.

Following the current state of the literature and the multi-sectoral setup of the program, I explore three key channels, which could drive my results. First, based on work by Cunha, Heckman, and Schennach (2010) showing the importance of cognitive and non-cognitive skills for school performance, I investigate ChCC's impact on these skills. Second, following the literature on the significant influence of the home environment on learning outcomes (Currie and Almond (2011); Almond, Currie, and Duque (2018)), I analyze its effects on parent-child relationships and the home environment. Third, given the work on the impact of early childhood education (Temple and Reynolds (2007); Carneiro and Ginja (2014); Williams (2019)), I analyze if ChCC results in increased attendance rates in early childhood education facilities. To study these three channels, I leverage data from the Longitudinal Survey of Early Childhood (ELPI), which contains rich and standardized information on these outcomes.

My evidence shows that the positive effect on school performance seems to be driven by important improvements in intra-household relations and raising attendance rates in early childhood facilities. I do not find evidence on improved cognitive and non-cognitive skills for the overall sample. Analyzing ChCC's impact by socioeconomic group reveals that the program's impact on these intermediate outcomes differs by gender and socioeconomic status. For girls, I find significant improvements in several of the parental outcomes investigated. The program also raises girls' executive functioning.[10] For boys, I find significant improvements in the availability of learning material at home. Moreover, there is some evidence

[8]ChCC has inspired similar programs in Colombia, Peru, Uruguay, El Salvador and South Africa (Ministry of Health 2017).

[9]Chile is a high-income OECD member country in Latin-America. Although it joined the OECD in 2010, it is still considered a developing country (United Nations 2022).

[10]Executive functioning refers to mental skills, which allow us to remember instructions, to mentally play with ideas and to plan.

pointing towards increased attendance rates in early childhood education for boys and socioeconomically privileged children. Overall, the pattern of results suggests that several of the program's multi-sectoral entry points might not work as intended.

My findings are robust to varying the size of bandwidth. They also hold when estimating a linear local randomization approach. My results are not driven by treatment manipulation around the cutoff or observable student characteristics. Furthermore, I employ an alternative empirical strategy, namely an event study design, showing consistent results. I verify that my findings are not confounded by the financial crisis in 2008, the development of copper prices, migration patterns, school entry dates, nor children's maturity. In addition, I generate evidence illustrating larger effects for children forming part of later roll-out groups. This evidence speaks for piloting early childhood interventions.

My work shows that pre-existing gaps between different socioeconomic groups could increase under comprehensive, universal early childhood development programs. Policymakers should revise how they can design these type of programs more effectively and also reach those most in need. The relatively low MVPF for ChCC indicates that targeted early childhood development programs might be more cost-efficient in creating long-lasting changes in the human capital accumulation of developing countries.

This paper is structured as follows. Section 1.2 describes the underlying early childhood development program *Chile Crece Contigo*. Section 1.3 gives an overview of the current state of literature relevant to the paper at hand. Section 1.4 describes the datasets at use and Section 1.5 the empirical methodology. Section 1.6 presents my main results and Section 1.7 several robustness checks. I apply a heterogeneity analysis in Section 1.8. In Section 1.9 I study possible underlying mechanisms, followed by a cost-benefit analysis in Section 1.10. Section 1.11 concludes.

1.2 Institutional Background and Program Description

1.2.1 Inequality and Human Capital Gaps in International Context

Chile is a high-income OECD member country in Latin-America. Although it joined the OECD in 2010, it is still considered a developing country (United Nations 2022). This is mainly because of high levels of inequality. While its GDP per capita stood at approximately 25,000 US-Dollar (after adjusting for purchasing power parity in constant 2017 US-Dollar) in 2019, a value close to Bulgaria's GDP per capita, its Gini coefficient is close to the one of other very unequal countries like Mexico or Bolivia (World Bank Group 2022). The persistently high levels of inequality are reflected in low social mobility. Chile only ranks 47th out of 82 countries on the *Global Social Mobility Index* by the World Economic Forum (2020).[11]

Human capital gaps are persistent in Chile. According to the *Human Capital Index* by the World Bank Group (2022) a child is only 65 percent as productive in adulthood as it could be if it enjoyed full access to education and health. This positions the country in the upper second quintile when compared globally. It performs similarly when restricting the index to harmonized test scores or learning-adjusted years of schooling. Educational attainment rates in primary and secondary schooling are close to the Latin-American and OECD average (World Bank Group 2022). Gross enrollment rates in preprimary education stood at 85 percent in 2019, a value slightly above the OECD average and 8 percentage points higher than the Latin-American average (World Bank Group 2022).

To summarize, Chile presents a highly relevant setting for the underlying research question. Not only is it characterized by persistent inequality and human capital gaps, which is the main motivation of this study. Chile is also a widely used bench-marking country for developing countries, especially in the Latin-American region. Comparisons to OECD countries are also reasonable.

1.2.2 Institutional Background of Early Childhood Development

Early childhood development has been one of the priorities of Chilean politics since the 20th century. As a result, child mortality decreased from 370 per 1,000 births in 1900 to 7.6 per 1,000 births in 2006 (Villalobos 2011). In 2001, Chile introduced its *Integrated Action Plan for Early Childhood and Adolescence.* The plan involved the creation of a public institution with the task of informing the presidency about the progress in the implementation of children's rights. The institution was established in 2003, at the same time as *Chile Solidario*[12].

In 2006, there were still persistent gaps in early childhood development.[13] This led to

[11] Chile assumes a score of 60.3 on the index. In comparison, the US ranks 27th with a score of 70.4 (World Economic Forum 2020).

[12] *Chile Solidario* is the social protection system for the poor population in Chile. It offers several programs and services aimed at improving the living conditions of these people.

[13] The 2006 socioeconomic household survey (CASEN) showed that 21.9 percent of children under the age of four lived in poverty, a higher share than in the overall population (13.7 percent) (Villalobos 2011). Moreover, the National Survey on Life Quality and Health revealed some troubling results. The study found that 30 percent of children below five years old did not meet internationally established development goals, and it revealed that significant developmental gaps existed between income quintiles with respect to child development (Villalobos 2011). Another gap was observed in early education. Coverage of early education in general was low. Only 26.5 percent of children between two and three years old attended kindergarten, while only 6 percent of children under two attended pre-kindergarten (Villalobos 2011). The gaps between income quintiles were marked, with four times more children from the top quintile attending early education facilities than children from the bottom quintile (Villalobos 2011).

the establishment of the *National Advisory Council for the Reform of Policies for Children* in 2006. The mission of the council was to develop a social protection system for early childhood development, laying the foundation for *Chile Crece Contigo*, which President Bachelet announced in October 2006 (Villalobos 2011).

1.2.3 Program Description

Chile Grows with You (ChCC – *Chile Crece Contigo*) is a comprehensive early childhood protection system which, alongside the social sub-programs *Chile Cuida* and *Chile Seguridad y Oportunidad*, is part of the overall social protection system of the Chilean government. The aim of the program is to accompany, protect and support all children and their families in an integrated manner. The program operates via an integrated network, combining services of several public sector institutions. It was introduced in 2007 with the goal of reducing inequalities during the first years of a child's life in Chile.

ChCC offers a variety of social services for children in their early life stages. The services offered through the program are adapted to the different needs that develop at each stage of life. It also addresses the needs of families, pregnant women, primary caregivers, and the family as a whole. The program is a universal program offered to all children and is part of the public health system (Asesorías para el Desarrollo 2012). Originally, children entered the program at their mother's first prenatal examination and left when they entered kindergarten or preschool. The program was expanded to include children from five to nine years of age in 2017.

ChCC is a program with a strong socioeconomic development focus, trying to address cognitive, emotional as well as behavioral lags in children's development through the program's comprehensive health, education and parental approach. Parental investment, cognitive as well as emotional stimulation, and a comprehensive health program are the entry points of ChCC to foster children's development. It therefore diverges from programs trying to lift people out of poverty through cash transfers.

The implementation approach of ChCC is an integrated one, recognizing that the municipality is the environment which forms and fosters the development of its children. The entry point and first contact point with the target population is the health sector, mainly through the Biopsychosocial Development Programme (PADB). The services offered through the program fall into three categories: An educational program for the Chilean citizenship and children's caregivers with the goal of raising awareness of the importance of early childhood development; services for children under the Biopsychosocial Development Programme PADB (Programa de Apoyo al Desarrollo Biopsicosocial), benefiting children from the womb to age four; special services for children belonging to the lowest 40th percentile in terms of income or non-income vulnerabilities.

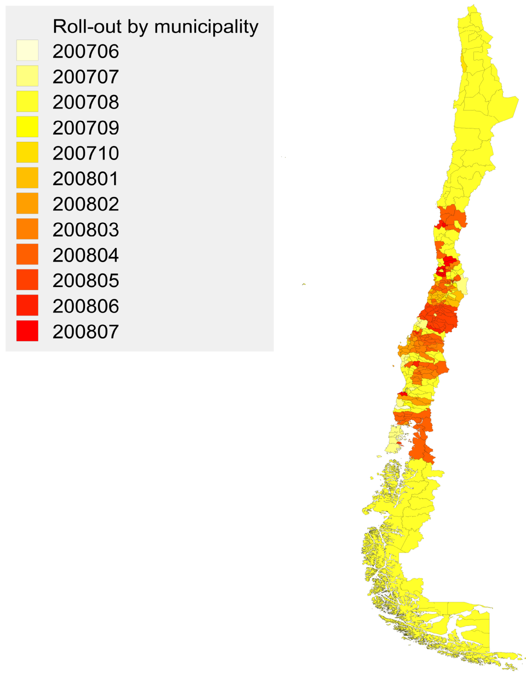
Appendix A.2 presents a detailed list of the services provided through ChCC.[14] The focus program of ChCC is the PADB, through which all children enter the program. The main changes that ChCC implemented in early childhood services are the following: an increase in the time for prenatal screenings from 10-20 minutes to 40 minutes and the inclusion of psychosocial factors in risk assessments, additional to biomedical factors; a comprehensive home visit program for at-risk patients; educational workshops on pregnancy and parenting as well as the distribution of educational materials; a guarantee of personalized services during childbirth; the availability of local facilities to care for at-risk children or children with developmental delays; the development of a local network to address all children's needs (The World Bank 2018).

ChCC's roll-out was gradual at the municipality by month level. The system was first implemented in 159 municipalities which were best prepared for its implementation. Experiences gained in the first round of roll-outs were then used for program implementation in the remaining municipalities during the second phase of ChCC's roll-out. Geographically, the roll-out of ChCC was dispersed as shown in Figures 1.1 and 1.2.

The inclusion of beneficiaries was also gradual (see Appendix A.1). First, the first generation of women was included in the system. In the next year, the second generation of women and all newborns were included in the system. By 2011, the system included all pregnant women and children under four years of age. The system also introduced different services gradually, to reflect the aging of beneficiaries. The roll-out of activities took place gradually targeting children according to their age. The system immediately offered these services to the whole target population. In the first year, the program mainly provided services for pregnant women and newborns. Importantly, there was strong compliance in ChCC's roll-out. Children born slightly before and after program implementation did not form part of ChCC.

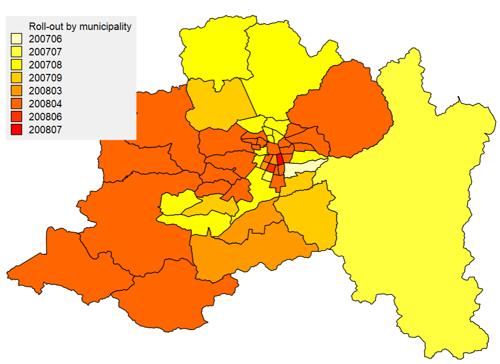
[14]Importantly, ChCC did not introduce all services listed in the Appendix, but enhanced them, developed them further, increased their scope and coverage, and improved their coordination and linkage with each other.

Figure 1.1: Geographic roll-out of ChC



Notes: The figure plots ChCC's roll-out over time by municipalities in Chile. ChCC's roll-out took place at the month by municipality level. Importantly, there was strong compliance in the monthly roll-out and the exclusion of children born slightly before the program. Source: Clarke, Méndez, and Sepúlveda (2020).

Figure 1.2: Geographic roll-out of ChCC (metropolitan area of Santiago de Chile)



Notes: The figure plots ChCC's roll-out over time by municipalities in the metropolitan area of Santiago de Chile. ChCC's roll-out took place at the monthly basis at the municipality level. Source: Clarke, Méndez, and Sepúlveda (2020)

1.3 The Economics of Early Childhood Development

This paper contributes to three different strands of the existing literature.

First, my results empirically manifest predictions made by theoretical models of early childhood development, such as those developed by Heckman (2006) and then later Almond, Currie, and Duque (2018). Heckman (2006) states that early investments strongly affect the productivity of later inputs and that they are dynamic complementarities[15] rather than perfect substitutes (Cunha and Heckman 2007). Consequently, investments in early childhood are especially important. The framework developed by Almond, Currie, and Duque (2018) confirms this. The authors highlight that a reallocation of resources from later to earlier in life creates pareto improvements. My results show that investments made in early childhood translate into positive human capital outcomes in middle childhood. Thus, my findings confirm the theory of dynamic complementarities.

Moreover, my paper builds on previous research showing positive effects of having access to social safety nets. The program design of ChCC links the paper at hand to this literature by comparing these findings on the impact of specialized social safety nets to a comprehensive ECD program. To the best of my knowledge there is only one paper so far analyzing the causal effect of ChCC. Clarke, Méndez, and Sepúlveda (2020) study the neonatal health component of *Chile Crece Contigo* and show that it has significant positive effects on birth weight and other early human capital outcomes.

Several studies within the literature on social safety nets focus on infant health. Almond, Hoynes, and Schanzenbach (2011), for example, show that participation in the Food Stamp Program three months prior to pregnancy leads to increased birth weight, with the largest gains at the lowest birth weights. Similarly, Hoynes, Page, and Stevens (2011) show that the implementation of WIC results in an increase in average birth weight. Related to that, Amarante et al. (2016) study the effects of cash transfers to poor pregnant women in Uruguay, that are part of the PANES program. They find that the incidence of low birth weights decreases by 20 to 19 percent. A paper by Goodman-Bacon (2018) analyzes the effect of Medicaid on infant and child mortality. The paper shows that infant and child mortality decline due to the program. Ko, Howland, and Glied (2020) examine the Supplemental Security Income (SSI) program, which includes cash transfers for poor children with disabilities. They find positive effects on a variety of health outcomes for children in the first eight years of life.

A related literature studies the impact of social safety nets on other dimensions besides

[15] Dynamic complementarities refer to the fact that early inputs in human capital affect the productivity of later inputs, a phenomenon that Cunha and Heckman (2007) call self-productivity.

infant health. Milligan and Stabile (2011) study the effect of an increase in child benefits, that translates into higher family income. They find significant positive effects on test scores, maternal health, and mental health, among other measures, with significant differences by gender.[16] Similarly, Akee et al. (2018) evaluate the impact of quasi-experimental unconditional household income transfers on children's emotional and behavioral health and personality traits, as well as on parental relationships. They find large positive effects.

Similarly, a connected stream of literature looks at the long-term impact of access to social safety nets during early childhood. Chetty et al. (2011) investigate the effects of the project STAR during kindergarten on earnings and find positive effects. Hoynes, Schanzenbach, and Almond (2016b) illustrate that participation in the Food Stamp Program leads to a reduction in the incidence of metabolic syndrome and an increase in economic self-sufficiency. Bailey, Sun, and Timpe (2021) study the long-term effects of the Head Start program[17]. Deming (2009) follows up and finds positive effects on adult human capital, adult economic self-sufficiency, the quality of adult neighborhoods and an increase in life-expectancy. The program leads to a large increase in adult human capital and economic self-sufficiency. The author demonstrates a positive effect of 0.23 standard deviations on a summary index of young adults' outcomes. Moreover, Akee, Jones, Simeonova, et al. (2020) study how the EITC program affects the next generation. They find significant and mostly positive effects, varying by household type and gender. Related work by Bharadwaj, Eberhard, and Neilson (2018a) documents positive effects of parental investment on academic outcomes.

Most of the interventions outlined in the literature above focus on the effect of income or in-kind transfers on children's short or long-term outcomes. Additionally, most of them are located in developed countries like the US or Canada. My paper contributes in that it goes beyond looking at the income component of child development by studying a comprehensive, integrated early childhood intervention. It also examines whether the positive effects found in the literature to date also apply to developing countries, where human capital needs are greatest.

Next, my work builds on previous research studying the effect of policy interventions for children. Two examples are the well-known Perry Highschool Project[18] (Heckman et al.

[16] While benefits have stronger effects on educational outcomes and physical health for boys, in the case of mental health outcomes effects are larger for girls.

[17] Head Start is a nationwide preschool program for poor children in the US, established in 1965 as part of the federal government's "War on Poverty".

[18] The Perry Highschool Project is a pre-school intervention targeting socioeconomically disadvantaged children. The High/Scope Perry Preschool Project started in 1962 to analyze the influence of pre-school education on children's learning outcomes (Weikart et al. 1970). The project was created when David Weikert noticed that poor children were doing much worse in school and formed a committee to address these gaps. As part of the project, a randomly selected group of vulnerable, ultra-poor children ages three to four were given access to pre-school as well as a weekly 90-minute home visits by a social worker, while a second

(2010) or Heckman, Pinto, and Savelyev (2013)) and the ABC/CARE program (García, Heckman, and Ziff 2018). In the same way, Attanasio et al. (2020) study the impact of a targeted early childhood intervention in Colombia and find significant gains in cognitive and socio-emotional skills among disadvantaged children. Felfe and Lalive (2018) analyze the expansion of early child care in Germany, showing strong but diverging effects on children's motor and socio-emotional skills. Most of these interventions are programs that target vulnerable children. My work contributes by asking whether universal programs can have similar effects and how they differ.

While there is a number of papers that investigate the effects of universal childhood interventions, none of these interventions follows the comprehensive approach of ChCC. Moreover, most of these studies analyze ECD programs in developed countries. One example is work by Baker, Gruber, and Milligan (2008) who analyze the introduction of universal child care in Quebec. According to their study, the provision of universal child care leads to an increase in maternal labor supply, but leaves children worse off. On the contrary, Cascio (2017) finds that attending a state-funded universal preschool in the US leads to increased test scores, particularly for the poor. Similarly, in the case of Germany, universal child care has larger treatment effects for disadvantaged children (Cornelissen et al. 2018). Furthermore, a universal ECD program in Norway is associated with long-term improvements in educational outcomes, as well as labor market outcomes (Havnes and Mogstad 2011). Similarly, Havnes and Mogstad (2015) show that the childcare expansion in Norway results in income gains during adulthood for children from the lower and middle parts of the income distribution, but income losses for those in the upper part.

My paper also talks to the literature which analyzes how to improve children's school performance. One example is the influential paper by Duflo (2001) that studies the impact of education supply on schooling outcomes. Black et al. (2014) ask how childcare subsidies impact student performance and several papers investigate the interaction between initial endowments and educational outcomes[19]. A stream within this literature analyzes the effect of income increases.[20] Similar to my contribution to the social safety net literature, my work expands this literature by looking beyond a pure income channel and analyzing the effects of a more comprehensive approach, bringing together several income and non-income channels.

Finally, Almond, Currie, and Duque (2018) single-out the necessity to further study the

group of vulnerable, ultra-poor children with similar characteristics served as a control group. Decades later, researchers compared several socioeconomic outcomes of both groups, such as criminal activities, income, and educational outcomes (Manning and Patterson 2006).

[19]See, for example, the work by Bharadwaj, Løken, and Neilson (2013), Bharadwaj, Eberhard, and Neilson (2018b) or Almond, Mazumder, and Van Ewijk (2015).

[20]For a good introduction to the topic, see studies by Dahl and Lochner (2012), Aizer et al. (2016), Muralidharan and Prakash (2017), Barrera-Osorio, Linden, and Saavedra (2019) or Millán et al. (2020).

effect of the "missing middle" years, meaning trajectory effects of early childhood and middle childhood. They identify a lack of knowledge about how early childhood, middle childhood and adulthood interact. My paper contributes to this identified gap in the literature through connecting early and middle childhood.

1.4 Data

In this section, I document the data I use to analyze ChCC's program impact on child outcomes in middle childhood. I mainly rely on administrative datasets provided by different entities of the government of Chile. I additionally use a rich survey on early childhood development as well as data on standardized test scores.

Standardized test scores. The first dataset used in this paper is from the national student achievement testing system (SIMCE). The data is provided by the National Agency of Educational Quality in Chile (Agencia de Calidad de la Educación 2021) and measures educational achievements along several dimensions, such as math or reading skills. The evaluation takes place every year and evaluates all second, fourth, sixth and eighth graders in elementary school, as well as the second and third graders in secondary school. I focus here on standardized test scores in reading and math of fourth graders tested between 2015 and 2018. To enter the schooling system in Chile a child must be at least six years old on March 31st of the respective school year (Ministerio de Educación 2021b). The treated children in 2007 would therefore enter primary education in 2013 at the earliest and be in fourth grade in 2016 at the earliest. The treated children in 2008 would be in fourth grade in 2017. Including the 2015 and 2018 evaluation years allows me to include children born one year before to one year after the introduction of ChCC.

Student register. The second dataset is provided by the Ministry of Education of Chile (Ministerio de Educación 2021a).It is the Student Register, containing information on the entire student body based on administrative school registry data. The data contains information about students' municipality of residence, date of birth, grade point average, school assistance rate, whether they passed the school year, the school and class they attend, and more. It also contains information on the socioeconomic status of students, divided into priority and preferential students. Priority students are those who belong to households with a socioeconomic background that make it more difficult for them to manage the educational process. These are students who belong to *Chile Solidario*, students who belong to the most vulnerable 30-percentile as defined by the Social Protection Scorecard (in Spanish: *Ficha de Protección Social* - FPS); students belonging to group A of the National Health Fund (in Spanish: *Fondo Nacional de Salud* - FONASA) who do not have a FPS (families in poverty, and receiving a family subsidy); students whose household income is below the poverty line; students whose mothers have less than four years of education; and students living in rural or poor communities. Preferential students are students who belong to the 80-percentile of the population, as defined by the social characterization score (in Spanish: *Instrumento de caracterización social vigente del Registro Social de Hogares*). The key outcome variable of interest is school performance, that is, the grade point average achieved by a student in a respective year. I use data on grade point averages from 2015-2018 and merge the data with SIMCE data based on the electronic student ID (MRUN).

Roll-out data. I use data on the monthly roll-out of ChCC data at the municipality level provided by Clarke, Méndez, and Sepúlveda (2020). The main explanatory variable is an indicator equal to one if a student was born after the implementation of ChCC in her respective community of residence. Table 1.1 gives an overview of the underlying student population by treatment group. The largest difference between both groups is the share of vulnerable students.

Table 1.1: Summary statistics of 4th-graders (2015-2018)

	Control group		Treatment group	
VARIABLES	N	Mean	N	Mean
Standardized math score	565,928	261.6	269,114	261.3
Standardized reading score	565,928	267.2	269,114	271.7
Grade point averages	565,928	5.808	269,114	5.886
Rural (share)	565,928	0.0977	269,114	0.101
Age (years)	565,907	9.501	269,093	9.107
Assistance (%)	565,928	91.06	269,114	91.49
Female share	565,928	0.490	269,114	0.514
Retention rate	565,928	0.0108	269,114	0.00660
Share of vulnerable students	565,928	0.824	269,114	0.737

Notes: The information above is based on SIMCE data and the National School Register from 2015-2018. Treated children are all children born after the implementation of ChCC in the respective municipality of residence. Standardized math scores range from 93 to 395 points. Standardized reading scores range from 115 to 406 points. Grade point averages represent the grade point average achieved by the respective child in a given school year and range from 1.0 (lowest) to 7.0 (highest). The retention rate is a dummy variable that takes the value of one once a child has not successfully completed the school year. Vulnerability refers to socioeconomic vulnerability based on a variety of characteristics defined by the Ministry of Education in Chile.

Survey data. To analyze potential channels through which ChCC affects school outcomes, I look at intermediate factors that could impact a child's performance in school. More specifically, I investigate the influence of ChCC on parental attitudes towards child-care, as well as developmental indicators such as psychomotor development, executive functioning, socio-emotional development, and anthropometric measures. I also analyze the program's impact on attendance rates in early childhood education facilities. To this end, I use data from the Longitudinal Survey of Early Childhood (ELPI) published by the Ministry of Social Development (Ministerio de Desarrollo Social y Familia 2021). The ELPI Survey consists of several questionnaires, addressed both to the children themselves and to their families. It also includes the use of child evaluation instruments that measure child development, as well as caretaker development and the interaction between the two. Survey questions and evaluation tools differ by age group (UNICEF 2018). Consequently, the sample size varies by variable. The survey consists of three survey waves from 2010, 2012, and 2017. I include those children into my sample who are in the 2010, 2012, and 2017 survey waves. I exclude children who are only reported in 2010 or 2012 as no information is available for them on their place or date of birth. I weight the observations using the sample weights provided by the ELPI evaluation team. Table A.2 in Appendix A.4 shows some basic characteristics of the underlying sample by treatment group.

The package of evaluation instruments for children consists of a set of tests measuring the following areas of child development: psychomotor development, executive functioning, socio-emotional functioning, as well as anthropometric measures. For the purpose of my analysis, I focus on instruments that were used with children who were part of the treatment as well as control group.[21]

To measure children's general and cognitive development I consider three different measures: TEPSI (the Psychomotor Development Test), TVIP (the Peabody Picture Vocabulary Test) and TADI (test of general infant learning). The TEPSI measures the psychomotor development of children and was part of the survey in 2010. A higher score indicates a better psychomotor development. The TVIP Score consists of 145 questions, and the ELPI gives an overall score generated from these questions. It is a norm-referenced measure of Spanish hearing vocabulary analyzing verbal reasoning, as well as language skills. A score below 70 is considered extremely low, and a score of more than 145 is considered extremely high. Instructors used this instrument with children in all three survey waves. The TADI score evaluates children ages three months to six years and measures four dimensions of child development: cognition, motor skills, language and socio-emotional development. This evaluation instrument was part of the 2012 survey. It consists of a task given to the child, a set of questions for the primary caregiver and a professional observation of the child. The TADI score is standardized for the Chilean population.

I measure the effect of ChCC on children's executive functioning using the BDST (Back-

[21]For an overview of all instruments see the ELPI User Manual (UNICEF 2018). For a detailed explanation and description of all instruments see a report published by the Universidad de Chile (2015).

ward Digit Span Task) as well as the HTKS (Head Toes Knees Shoulders Task). The BDST consists of 16 questions and measures the working memory. The ELPI reports an overall score based on these questions. The HTKS is a game for children, in which they are asked to do the opposite of what an instructor says.

I then analyze ChCC's impact on children's socio-emotional development via the CBCL1 (Child Behavior Checklist 1). The CBCL1 is a caregiver report identifying behavioral problems in children, based on the following symptoms: aggressive behavior, anxiety, attention problems, rule-breaking behavior, somatic complaints, social problems, thinking problems, and depression. The CBCL1 consists of 99 questions. The ELPI, in turn, generates an overall test score from these questions. A percentile score of less than 93 is considered normal, and a score greater than 98 is considered clinical range. A total scale score of less than 60 is considered normal, while a total scale score of greater than 83 is considered clinical range. This evaluation instrument was part of all three rounds in the ELPI survey.

For the anthropometric measures, I create a dummy variable that equals one if the interviewer observes some kind of abnormality in a child's height.

To measure the impact of ChCC on a children's immediate environment and on caregiver parenting, I use the PSI (Parental Stress Index), PSCS (perceived self-confidence scale), CESD-10 (Center for Epidemiologic Studies Depression Scale 10) and HOME Index (Home Observation Measurement of the Environment Index). The PSI consists of 36 questions answered directly by the principal caregiver. Each question relates to a subdomain of parental stress and is scored on a five-point scale. A score of less than 80 is considered normal, while a score greater than 90 is within the clinical range. The PSCS consists of 17 items measuring the self-assessment of parenting skills. Higher scores represent greater parent confidence in their parenting skills. The CESD-10 is based on 10 items. People with higher scores are more prone to depression. I also use the HOME (Home Observation for the Measurement of the Environment) Index to measure household quality. The HOME Index consists of 13-43 questions. It measures several dimensions of household quality, such as the emotional interaction between principal caregivers and children, the presence of learning material, as well as maternal commitment. The interviewer assigns points for each dimension, with eight points being the maximum score. Lastly, I retrieve information on parenting practices (such as inadequate dental care) from the survey.

Table A.3 in Appendix A.4 gives an overview of the evaluation instruments under consideration.

1.5 Identification Strategy

1.5.1 Regression Discontinuity Design

In this section, I describe the identification strategy I use to empirically assess the intention-to-treat (ITT) effect of ChCC on human capital accumulation in middle childhood. Simply regressing an indicator variable, which is equal to one if a child was born after ChCC's roll-out, and zero otherwise, on child outcomes in middle childhood might lead to biased estimators. Children from earlier birth-cohorts might significantly differ from children in later birth-cohorts. This is problematic especially if they differ on unobservable dimensions, which also affect the outcome variables of interest. To give an example, children of the pre-treatment group might be subject to different education policies than children of the treatment group. These policies might significantly affect schooling outcomes, but are unobservable in the data at hand. Therefore, a simple ordinary least square regression might mistake the effects generated from changes in education policies for changes generated through the implementation of ChCC.

To address these endogeneity concerns, I exploit the fact that there is a random cutoff for potential exposure to ChCC, that is, the date of birth of a child, and apply a regression discontinuity design (RDD). The intuition behind RDDs is that students are very similar around the cutoff. Therefore, the potential existence of unobservable confounding factors is less likely. Comparing outcomes of students located closely below the cutoff to outcomes of students located closely above the cutoff delivers the local treatment effect of ChCC. Hence, I estimate the following equation:

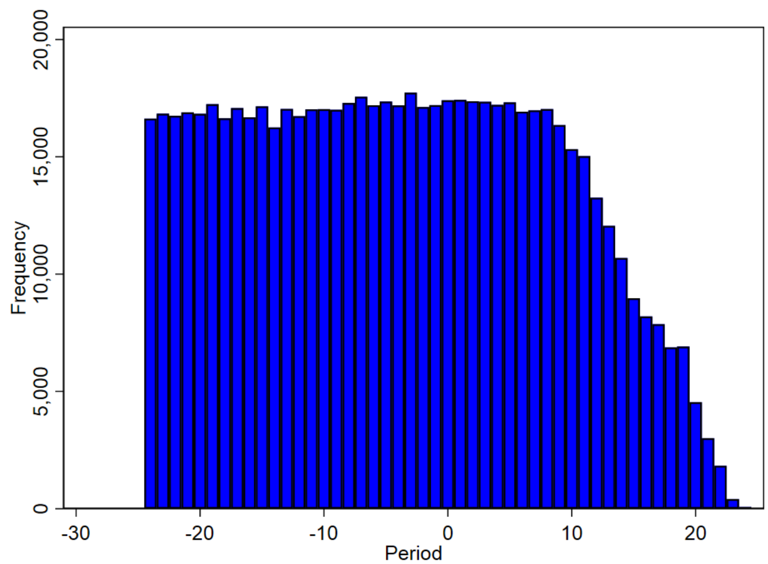
$$Y_i = \alpha + \beta ChCC_i + \gamma_1(X_i - c) + ChCC_i\gamma_2(X_i - c) + \varepsilon_i \tag{1.1}$$

, where:

$$ChCC_i = \begin{cases} 1, & X_i \geq c \\ 0, & X_i < c \end{cases} . \tag{1.2}$$

Being exposed to ChCC is determined by the threshold c (being born after ChCC) of the discrete variable X_i, the date of birth of the respective child. As ChCC's roll-out was staggered, I first normalize the threshold. I do so through setting the roll-out date to zero and then calculating the difference between each child's date of birth and the roll-out date. The running variable is equal to the number of months ChCC was in place in a respective municipality. X_i depends on the bandwidth b of the data used. The bandwidth is equal to the number of periods under consideration before and after ChCC's implementation took

Figure 1.3: Histogram of the running variable (months since roll-out)



Notes: The figure shows a histogram of the running variable. The x-axis presents the running variable. The running variable is the number of months relative to ChCC's roll-out. 0 is the staggered roll-out period. The y-axis indicates the frequency in absolute numbers. Source: SIMCE (2014-2018).

place. X_i is therefore as follows:

$$c - b \leq X_i \leq c + b \tag{1.3}$$

RDDs were originally designed for settings with continuous running variables. Importantly, it is only possible to apply RDDs to settings with discrete running variables when the number of mass points is large (Cattaneo, Idrobo, and Titiunik 2019). If the number of mass points is small, a local randomization approach might be more appropriate than a RDD. Figure 1.3 shows a histogram of the running variable. I restrict the sample to all children born 18 months before to 18 months after ChCC's roll-out, a setup under which there are sufficient observations in each of the cells of the underlying dataset. This results in 37 mass points, a relatively small number.

Consequently, I decide to implement a local randomization approach instead of a continuity-based method. I employ finite-sample methods to determine the cutoff window under which the assumption of randomized treatment assignment is most plausible. I follow Cattaneo, Titiunik, and Vazquez-Bare (2016) and implement a window-selection procedure based on balance tests. I find that the optimal window is equal to four periods around the cutoff.[22]

[22] This applies to a functional form of polynomial order zero. The optimal cutoff window of a local randomization approach with higher polynomial orders is the minimum cutoff window. As it is challenging to estimate the slope with only two data points, I consider a bandwidth of ten as the main empirical

This means that the optimal cutoff window consists of the two birth cohorts previous to ChCC's implementation and the first three birth cohorts exposed to ChCC. For the underlying figures and details behind the optimal window length selection see Appendix A.5.

The baseline estimation strategy relies on the assumption that the relationship between exposure to ChCC and schooling outcomes in middle childhood has a regression slope equal to zero. But if the true relationship is linear or even non-linear, the local randomization approach could mistake linear and non-linear relationships for discontinuities. To account for this caveat, I employ a parametric regression specification of the local randomization approach. In detail, I follow Angrist and Pischke (2009) and estimate the following regression:

$$Y_i = \alpha + \sum_{p=1}^{n} \gamma_p c_i^p + \beta_0 ChCC_i + \sum_{p=1}^{n} \beta_p c_i^p ChCC_i + \varepsilon_i \tag{1.4}$$

, where:

$$ChCC_i = \begin{cases} 1, & c_i \geq c \\ 0, & c_i < c \end{cases} . \tag{1.5}$$

I follow Pei et al. (2022) and select the optimal polynomial order p based on a mean squared error estimation. I find that an increase in the polynomial order does not lead to a lower mean squared error. I therefore rely on a local randomization approach, which does not include slope terms, as my baseline estimation. I later account for higher polynomial orders to test the validity of my findings.

1.5.2 Testing the Validity of Identification Assumptions

The underlying assumption of the local randomization approach is that the assignment of each child to the treatment was random and that there was no manipulation into treatment. To test this assumption I conduct a falsification test. I find that there are 17,203 observations in the month before ChCC's roll-out, and 17,410 observations during the roll-out. In the month after ChCC's roll-out there are 17,426 observations. This suggests that there is no manipulation or non-random selection into the treatment. The ratio of observations close to the cutoff is nearly 1. This is consistent with the assumption that treatment assignment is random and close to a probability of 0.5. Additionally, in this specific setup manipulation might be of low concern, as it might be difficult to time child birth to the monthly level. A pregnancy takes 10 months and it is unlikely that the roll-out date of ChCC played a role

specification. In these cases, I show that findings are robust to specifications, which only consider the minimal cutoff window.

in the monthly timing of pregnancies. I conduct a binomial test to confirm this empirically, and the p-value of a binomial test is close to 1. This confirms that treatment manipulation is of low concern in this setup.

In addition, I investigate whether the chosen window around the cutoff potentially drives my empirical results. When considering different nested windows, namely up to 20 months around the cutoff, the ratio of observations around the cutoff remains balanced (see Figure 1.3). Consequently, the probability of treatment assignment remains around 0.5 and it is unlikely that the window size drives my results.

Additionally, the local randomization approach relies on the assumption that individuals close to the cutoff are similar on observable and unobservable characteristics. While I cannot analyze the similarity of unobservable student characteristics around the cutoff, this is possible for observable covariates. I test if treated and control groups at the cutoff are on average similar in terms of observable characteristics. I can observe three covariates in the data at use, namely students' gender, socioeconomic vulnerability and the degree of urbanization of the school they attend.

Table 1.2 shows the mean values of the three observable student characteristics in the optimal window around the cutoff. It also shows the resulting p-value of a t-test, which investigates the equality of means in the optimal cutoff window in Column 3. I cannot reject the null hypothesis of no significant difference in the means in the minimum cutoff window. This applies to all of the three observable covariates.

The evidence presented speaks for the identification assumptions of the local randomization approach to be likely fulfilled in this setting. Most importantly, there seems to be no manipulation into treatment around the cutoff. Moreover, randomness of the program's roll-out is plausible. Lastly, there is no discontinuity of observable student characteristics around the threshold (see Figure 1.4 to Figure 1.6).

Table 1.2: Baseline municipality characteristics (4 periods around cutoff)

	Control mean	Treatment mean	T-test p
Female	0.50	0.50	0.50
Vulnerable student	0.74	0.74	0.20
Rural	0.10	0.10	0.37
Standardized math score	262.36	263.24	0.36
Standardized reading score	270.35	272.41	0.00
Grade point averages	5.86	5.89	0.05
Observations	34,324	52,206	86,530

Notes: The table shows the baseline characteristics of children in the treatment and control group. I restrict the window to the optimal bandwidth (4 periods around the cutoff). Column 1 presents the mean value for children in the control group, Column 2 for children in the treated group and Column 3 the p-values for a two-sample t-test that tests for systematic differences between both groups in each variable in the table. Standardized math scores range from 93 to 395 points. Standardized reading scores range from 115 to 406 points. Grade point averages represent the grade point average achieved by the respective child in a given school year and range from 1.0 (lowest) to 7.0 (highest). Source: SIMCE (2015-2018), MINEDUC (2015-2018).

Figure 1.4: Regression Discontinuity Plot (female)

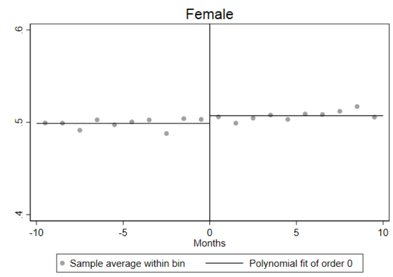


Figure 1.5: Regression Discontinuity Plot (vulnerable)

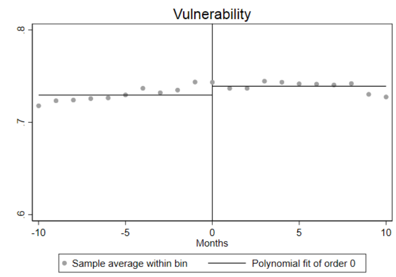
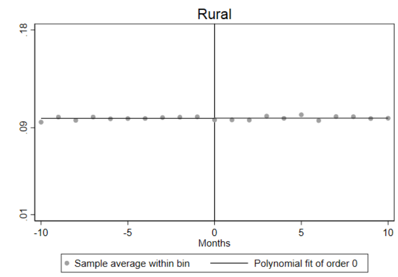


Figure 1.6: Regression Discontinuity Plot (rural school)



Notes: The figures above show the local randomization design plots for observable control variables. I do not include a slope in this case. The left panel shows the plot for gender, the middle panel the one for socioeconomic vulnerability, and the right panel the one for urbanity. *Female* is a dummy variable equal to 1 for girls and 0 otherwise. *Vulnerable* is a dummy variable equal to 1 for socioeconomically vulnerable by definition of the Ministry of Education and 0 otherwise. *Rural school* is a dummy variable equal to 1 for students who attend a school in a rural area and 0 otherwise. I restrict the periods shown to a bandwidth of 20. This means that the figures show the average values of the socioeconomic controls for all children born ten months previous to ChCC's roll-out to ten months after its roll-out. The black horizontal line features the threshold of the local RD approach, namely zero. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

1.6 Impacts on Schooling Outcomes

After establishing the plausibility of the underlying identification assumptions, I analyze the local randomization discontinuity (RD) effect of exposure to ChCC on the main outcome variables of interest, namely the standardized math and reading score as well as grade point averages. If the program successfully increases the accumulation of human capital in middle childhood, I expect to see positive and significant effects of program exposure on schooling outcomes.

Before analyzing the program's impact on schooling outcomes, I further investigate the role of potential confounding factors around the cutoff. To do this, I estimate the local RD effect on the three observable student characteristics. The last three columns in Table 1.3 confirm that there is indeed no discontinuity in any of the observable covariates in the optimal cutoff window. The resulting p-values from a local RD estimation on the three observable covariates in the optimal cutoff window are larger than the most commonly used significance levels. This shows that the assumption of similarity between observed covariates is plausible in this cutoff window.

Table 1.3 implies that exposure to ChCC leads to improved schooling outcomes in the optimal window length. Column 2 shows that the program increases standardized math scores by 0.883 points, standardized reading scores by 2.059 points and grade point averages by 0.03 points. The p-values in Column 3 are zero or very close to zero. Consequently, the reported point estimates are significant at the 1 percent significance level. Compared to the mean values in the optimal window, this corresponds to an increase of 0.337 percent in standardized math scores, 0.762 percent in standardized reading scores, and 0.512 percent in grade point averages. I show that my findings are robust to restricting the window to the minimum bandwidth as well as a bandwidth of 20. See Appendix A.4 for details.

Figure 1.7 to 1.9 show the related local randomization design plots for schooling outcomes. The solid lines present the predicted values from a regression of schooling outcomes on a zero-degree polynomial in months to the birth eligibility cutoff. Negative values of months indicate children born before program implementation. The figures show that there is a clear increase in standardized math scores and grade point averages for children born after ChCC's roll-out. The jump in standardized math scores, on the other hand, seems to be negligible. Figure A.3 to A.5 in Appendix A.6 present the same figures restricting the window length to the optimal bandwidth of four. These figures confirm my insights from the RD plots in the window length of 20.

My results illustrate that the program successfully improves schooling outcomes in middle childhood. The results are positive and significant across the three educational variables investigated in this paper. This shows that universal, comprehensive ECD programs like ChCC can indeed successfully foster children's human capital accumulation. Still, it is important to assess if these improvements outnumber the program's costs. To evaluate the cost-efficiency of the program, I conduct a cost-benefit analysis later in this paper.

Table 1.3: Local RD effect of ChCC on schooling outcomes and observable control variables in the optimal window around cutoff

	Variable	RD Estimate	P-Value	N (left)	N (right)
1	Standardized math score	0.883	0.008	34,324	52,206
2	Standardized reading score	2.059	0.000	34,324	52,206
3	Grade point averages	0.030	0.000	34,324	52,206
4	Gender	0.000	0.983	34,324	52,206
5	Vulnerability	-0.000	0.921	34,324	52,206
6	Rural	-0.003	0.185	34,324	52,206

Notes: The table shows the local RD effect of ChCC in the four months around the cutoff. This means that the estimation considers all students born two months before and after ChCC's roll-out as well as those born during its roll-out. The first column shows the intention-to-treat (ITT) effect of ChCC on the three schooling outcomes and observable covariates. Column 2 presents the related coefficient p-values. Column 3 and 4 show the number of observations N on each side of the cutoff. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

Figure 1.7: Regression Discontinuity Plot (standardized math scores)

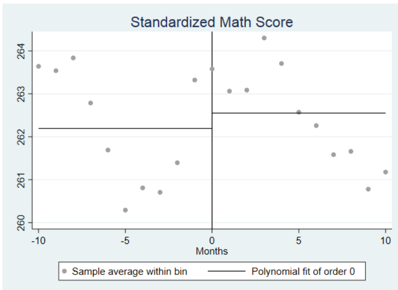


Figure 1.8: Regression Discontinuity Plot (standardized reading scores)

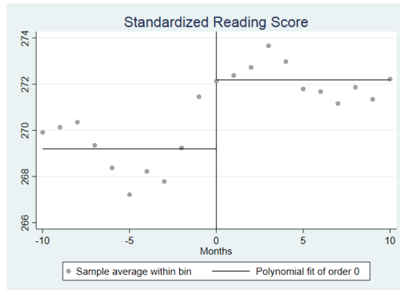
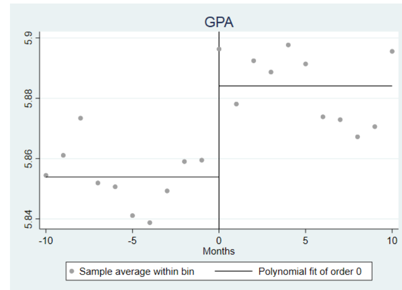


Figure 1.9: Regression Discontinuity Plot (grade point averages)



Notes: The figures above show the local randomization design plots for schooling outcomes. The left panel shows the plot for standardized math scores, the middle panel the one for standardized reading scores, and the right panel the one for grade point averages. I restrict the periods shown to a window length of 20. This means that the figures show the average values of schooling outcomes for all children born 10 months previous to the roll-out of ChCC to 10 months after its roll-out. The black horizontal line features the threshold of the local RD approach, namely zero. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

1.7 Robustness Checks

The underlying functional form or unobservable factors taking place at the same time as ChCC's roll-out could confound my main findings. Hence, I conduct several robustness checks to validate my results. First, I employ a parametric estimation of the local RD approach. Next, I employ an event study to show that my main results hold under an alternative empirical strategy. Lastly, I validate that my findings are not confounded by the financial crisis, the development of copper prices, migration patterns, school entry dates nor children's maturity.

1.7.1 Parametric Estimation

To investigate if assumptions about the underlying functional form drive my results, I include slope terms. Table 1.4 presents the results of a local linear randomization approach. Based on the p-values in Column 2, all three coefficients are significant and positive at the 1 percent significance level. When comparing the point estimates in Column 1 to the point estimates from the baseline specification, the local RD estimator is larger. This could mean that the non-parametric estimation underestimates the true impact of the program. The other way around, it could also indicate that accounting for a polynomial order of one overestimates the intention-to-treat (ITT) effect of exposure to ChCC.

Figure 1.10 and 1.11 plot the discontinuity for standardized test scores and Figure 1.12

the one for grade point averages. The solid lines give the predicted values from a regression of schooling outcomes on a one-degree polynomial in months to the birth eligibility cutoff. The figures confirm that there is a clear jump in standardized reading scores and grade point averages for children born after program implementation. The graph on standardized math scores is less convincing. Note that all graphs have a downward slope. Schooling outcomes decrease as one moves away from the birth eligibility cutoff for children exposed to the program, and similarly increase for children born before program implementation. These slopes are driven by window lengths and disappear when increasing the window size to 11 or 12 periods. Due to the staggered nature of the program I am confident that these downward trends do not represent birth cohort or school-starting ages. They might be related to the slight periodicity visible in the RD plots.

Table 1.4: Local RD effect of ChCC on schooling outcomes in the 20 periods around the cutoff with a polynomial fit of order 1

		RD Estimate	P-Value	N (left)	N (right)
1	Standardized math score	3.157	0.000	172,695	186,707
2	Standardized reading score	3.887	0.000	172,695	186,707
3	Grade point averages	0.043	0.000	172,695	186,707

Notes: The table shows local RD effects of ChCC in the 20 months around the cutoff window. This means that the estimation considers all students born 10 months before and after the roll-out of ChCC as well as those born during the roll-out. The results are based on a parametric regression specification of the local randomization approach of degree one. Column 1 reports the point estimates, Column 2 the coefficient p-values, and Column 3 and 4 the number of observations N on each side of the threshold. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

Overall, the results from parametric estimations confirm my main findings from the baseline local randomization approach. Exposure to ChCC leads to significant improvements in schooling outcomes in middle childhood. Still, the exact magnitude of the program's impact on standardized test scores is sensitive to parametric assumptions behind the empirical estimation.

Figure 1.10: Regression Discontinuity Plot - p=1 (stand. math scores)

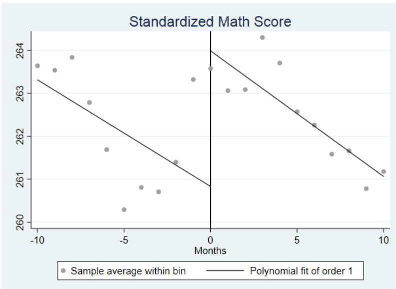


Figure 1.11: Regression Discontinuity Plot - p=1 (stand. reading scores)

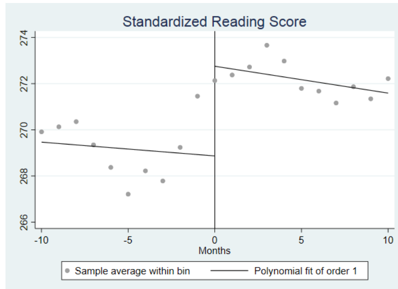
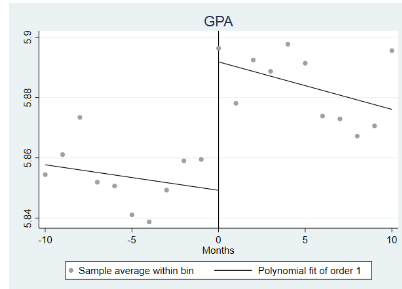


Figure 1.12: Regression Discontinuity Plot - p=1 (grade point averages)



Notes: The figures above show the local randomization design plots for schooling outcomes assuming a polynomial fit of order one. I consider a cutoff window of ten periods before and ten periods after the actual cutoff. The cutoff is equal to zero and represented by the black horizontal line. The left panel shows the plot for standardized math scores, the middle panel the one for standardized reading scores, and the right panel the one for grade point averages. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

1.7.2 Event Study

To further validate the local randomization approach, I employ an event-study design. My main explanatory variable is the number of months that ChCC had been in place in a certain municipality when a child was born. I explain this further using the example of a child born in August 2008. If ChCC was introduced in her respective municipality in August 2007, the main explanatory variable has a value of 12. If the child was born in August 2006, the main explanatory variable has a value of -12. The regression estimation for the event study is as follows:

$$y_{mb} = \alpha + \sum_{p=-13}^{13} \beta_p \times I_p + \eta_m + \lambda_b + \theta_s \times b + \delta M_{pre} \times b + \gamma X_i + \varepsilon_{imb} \tag{1.6}$$

,where m stands for the municipality, and b for the birth-cohort. One cell in the sample represents a combination of a specific municipality and birth-cohort. y_{mb} is the outcome of interest (as, for example, the average municipality-level standardized test score for a certain birth-cohort) and β_p the effect of leads and lags (denoted as I_p) included in the event study design. η_m is a municipality fixed effect, λ_b a birth-cohort fixed effect, and $\theta_s \times b$ a state-specific linear time of birth trend. Standard errors are clustered at the regional level. I omit period -1. Additionally, I interact some pre-treatment municipality characteristics with a time of birth trend ($M_{pre} \times b$) and control for individual time-varying controls (X_i).

Schmidheiny and Siegloch (2019) recommend a binning approach in which the number of pre-periods included in the event study is equal to the first year of data for the dependent variable (in my case, July of 2007) minus the effect window (which in my case is 13 periods). Figure 1.13 shows the results for standardized math scores, Figure 1.14 for standardized reading scores and Figure 1.15 for grade point averages.[23]

Figure 1.13: Event Study Graph (standardized math score)

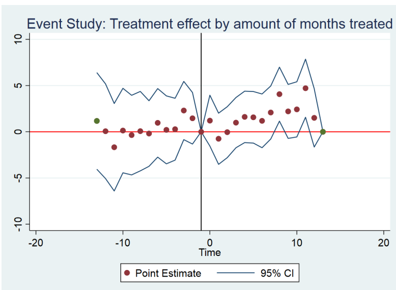


Figure 1.14: Event Study Graph (standardized reading score)

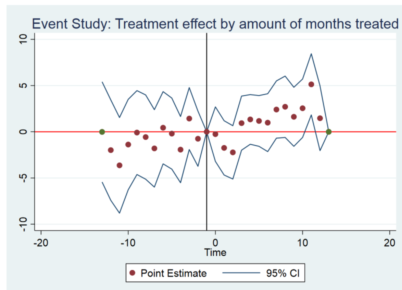
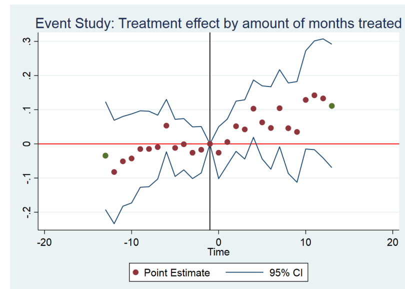


Figure 1.15: Event Study Graph (grade point averages)



Notes: The plots above show results for an event study on standardized math and reading scores as well as grade point averages. I aggregate data at the date of birth by municipality level. The main explanatory variable is the number of treatment periods relative to ChCC's roll-out. I control for municipality fixed effects as well as date of birth fixed effects. I include an interaction term between birth cohorts and regions of residence, as well as pre-treatment controls. I include the following pre-treatment municipality characteristics: poverty rates at the municipality level, the number of families receiving subsidies, the available budget per municipality, the share of education spending from the Ministry of Education, the type of administrative cooperation in education, the student-teacher ratio, having or not having a primary health unit, the health transfer per capita from the Ministry of Health, and the share of votes in the 2004 mayoral elections. I control for the following individual characteristics: the share of female students, the share of vulnerable students, and the share of students from rural areas. The omitted event time is period -1, represented by the vertical black line. Standard errors are clustered at the regional level. Source: SIMCE and MINEDUC (2014-2018).

Figure 1.13 shows that there is no pre-trend for standardized math scores and that math scores increase consistently after the introduction of ChCC. The same is true for standardized reading scores (Figure 1.14), although there might be a small periodic trend in the pre-treatment period. Still, standardized reading scores increase after the introduction of ChCC. Figure 1.15 illustrates that there is evidence of a slight but negligible pre-trend in the case of grade point averages. Grade point averages increase steadily in the post-treatment period. Overall, my event study confirms the results from the local randomization approach. I conduct a joint significance test of the 13 lags included in the event study. I can reject the null hypothesis that all coefficients are zero at the commonly used levels of significance.

My results confirm that exposure to ChCC has a positive effect on human capital accumulation in middle childhood.

[23]I take advantage of the command *eventdd* provided by Clarke and Tapia-Schythe (2021).

1.7.3 Additional Robustness Checks

Migration as a Confounding Factor

I do not have information on the children's municipality of birth, only on the children's municipality of residence. Therefore, my treatment and control group could be confounded by internal migration patterns. To test this, I analyze these patterns using data from the latest 2017 micro-census. The data shows that 15.8 percent of the population are internal migrants (Instituto Nacional de Estadísticas 2020). Internal migrants are defined as all people who changed their residence between 2012 and 2017 by moving between regions or within one region but between municipalities. Importantly, households with children are less likely to migrate internally, and the share is lowest among the youngest and oldest population (less than 2 percent). Internal migration patterns only begin to take hold for children over age 15. As these groups are not included in my sample, I conclude that internal migration patterns of less than 2 percent for my target group should not be a significant confounding factor in the definition of my treatment and control group.

The Influence of the Financial Crisis and Copper Prices

Since the implementation of ChCC took place just before the onset of the financial crisis in 2008, I analyze the potential effect of the financial crisis as a confounding factor. The financial crisis would be a problem for my estimation strategy if it had a systematically different effect on schooling and developmental outcomes of children in municipalities introducing ChCC earlier than those municipalities introducing ChCC at a later stage. This channel could arise from a transitory effect of the financial crisis on income and poverty and then on schooling outcomes.

The implementation of ChCC was completed in August 2008. Like most emerging economies, the financial crisis hit Chile later than developed countries. This is why Chile did not enter a severe recession until late 2008 (Cortés 2016). Real GDP growth (year-to-year) began to fall in the third and fourth quarter of 2008 (3.5 percent and 0.9 percent respectively), and quarterly growth was negative by the first quarter of 2009 (-2.6 percent) (OECD 2021). Based on this data I conclude that the financial crisis is no threat to my identification strategy.

With respect to copper prices, the same reasoning applies. If municipalities introducing ChCC earlier are municipalities which depend heavily on the Chilean copper industry, and if these industries are then hit harder by a negative development of copper prices, the development of copper prices might be a confounding factor. As the copper price did not start to fall sharply until September 2008, I can rule it out as a confounding factor.[24]

School Entry Date and Maturity

One important requirement for RDDs is that there exists a clear cutoff which divides the population of interest into treated and control units. One concern could be that there is an alternative unobservable cutoff which coincides with ChCC's birth threshold but refers to another program. If this alternative program affects educational outcomes I might confound its effects with the impact of ChCC. Because of the staggered rollout of ChCC I am less concerned about potential unobservable programs that were rolled out on one specific date. School entry dates, for example, are not a concern.[25]

Another possible identification concern is the maturity of children. Children born after the birth cutoff in a respective municipality are automatically older than those born before the cutoff. If maturity plays a role in educational outcomes the increase in schooling outcomes could be due to the underlying age structure. Again, because of the staggered nature of ChCC's roll-out, I am less concerned about this potential confounding factor. I verify that maturity does not play a role in my findings by controlling for the age of children. I find that my estimates more than double in size when including age as a control variable.[26]

1.8 Impact on Schooling Outcomes by Subgroups

Next, I analyze the program's effects on educational outcomes by gender and socioeconomic vulnerability. I do so to investigate if comprehensive early childhood development programs have the potential to reduce inequalities between these groups. I divide the treatment group by gender and socioeconomic vulnerability. I then estimate the local RD effect of the program's impact in the optimal window, namely in the four periods around the cutoff.

Table 1.5 shows that the program's impact significantly differs by gender. Comparing the point coefficients in Column 1 and 2 of Panel 1 reveals that the program has larger effects on boys' standardized math scores than on girls'. The local RD estimate is barely significant at the 10 percent significance level for girls. According to Column 2, exposure to ChCC increases standardized math scores by 0.996 points for boys but only by 0.771 points

[24]For the detailed development of copper prices see `https://tradingeconomics.com/commodity/copper`, and for a graphical overview see Figure A.1 in the Appendix.

[25]School entries are defined via the 31st of March of each year. To enter school, children must be at least six years old on this day. I do not account for cohort-fixed effects in my regression, as more than 99 percent of all children belong to 4th grade.

[26]The RD coefficient is 7.836 for standardized math scores, 7.342 for standardized reading scores, and 0.106 for grade point averages.

for girls. When compared to the average in the optimal window, this is a relative increase of 0.376 percent for boys compared to 0.295 percent for girls. Turning attention to Panel 2 and 3, similar patterns become apparent for the other two schooling outcomes investigated. While in the case of standardized reading scores and grade point averages all RD estimators are significant at the 1 percent significance level, exposure to ChCC results in smaller effects for girls than for boys. Column 2 shows that the program increases standardized reading scores by 2.315 points for boys, but only by 1.805 points for girls. The impact on grade point averages is also lower for girls (see Column 2 and 3 in Panel 3).

Differences are even larger when making a distinction by socioeconomic vulnerability. Table 1.5 shows that the program's impact on standardized math scores is insignificant for socioeconomically vulnerable children. The p-value reported in Column 3 is above 0.1. Column 3 and 4 in Panel 1 present the respective point estimates. While the program raises standardized math scores by 1.639 points for non-vulnerable children, this same effect is only 0.615 points for vulnerable children. The effect is therefore nearly three times larger for the socioeconomically privileged group. Similarly, the program's impact on standardized reading scores is 1.8 times larger for the socioeconomically privileged group (3.107 points versus 1.689 points). It is 2.8 times larger in the case of grade point averages (0.056 points versus 0.020 points).

These findings hold when accounting for a larger cutoff window, as well as for polynomial orders of degree one by subgroups. Appendix A.4.3 presents the details.

The heterogeneous impact of ChCC by subgroups could mean that the program fails to address important human capital gaps between different groups and benefits those who are already more privileged most. This hints towards important shortcomings in the inclusiveness of universal, comprehensive early childhood interventions. Less privileged groups might be less likely to comply with the program, or the inclusion error might be larger for these groups. Another possible explanation is that the quality and quantity of the services offered to these groups might be worse. The heterogeneous impact of exposure to ChCC could also be evidence of the program falling short on addressing important drivers behind human capital gaps, such as gender stereotypes. The results presented cast doubt on the effectiveness of universal, comprehensive programs like ChCC in closing human capital gaps between socioeconomic groups. While the program's impact is overall positive for all groups investigated, the heterogeneous impact by vulnerability and gender is worrisome. Targeted programs might be more suited to address the needs of vulnerable children.

Table 1.5: Local RD effect of ChCC in the optimal window around the cutoff on schooling outcomes by groups

	Subgroup	Boys	Girls	Vulnerability	Non-vulnerability
	Panel 1: Standardized math scores				
1	RD Estimate	0.996	0.771	0.615	1.639
2	P-Value	0.037	0.094	0.116	0.008
3	Panel 2: Standardized reading scores				
4	RD Estimate	2.315	1.805	1.689	3.107
5	P-Value	0.000	0.000	0.000	0.000
6	Panel 3: Grade point averages				
7	RD Estimate	0.034	0.025	0.020	0.059
8	P-Value	0.000	0.000	0.000	0.000
9	N (left)	17,039	17,285	25,375	8949
10	N (right)	25,912	26,294	38,579	13,627

Notes: The table shows the local RD effects of ChCC in the optimal window by subgroups. The optimal window consists of four months around the threshold. This means that the estimation considers all students born two months before and after the roll-out of ChCC as well as those born during the roll-out. Panel 1 shows the results for standardized math scores, Panel 2 for standardized reading scores, and Panel 3 for grade point averages. I first report the local RD estimator, and then the coefficient p-value. Column 1 reports the results for boys, Column 2 the ones for girls, Column 3 the ones for socioeconomic vulnerable children, and Column 4 for non-vulnerable children. For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016). Source: SIMCE (2015-2018) and MINEDUC (2015-2018).

1.8.1 Early versus Late Roll-out Group

As detailed earlier in this paper, ChCC was rolled out in two phases. The early roll-out group consisted of all municipalities best prepared for its implementation. The experiences gained in the first round of roll-outs was then used for the implementation in the remaining municipalities. To investigate if the two-phased roll-out confounds my findings, I analyze the program's impact for the early and late roll-out group separately.

Table 1.6 suggests that the program's impact is driven by the late roll-out group. Based on the p-values reported in Column 1, the coefficients on all three schooling outcomes are insignificant at the 1 percent significance level for the early roll-out group. This is not the case for the late roll-out group. The p-values reported in Column 2 are zero in all three cases. Moreover, the point coefficients reported in Column 2 are larger than the ones reported in Column 1. These results imply that policymakers and implementing partners in the second phase of the program's roll-out benefited from important insights and experiences gained in the first phase of the roll-out. This, on the other hand, speaks for the importance of pilot projects when implementing early childhood interventions.

Table 1.6: Local RD effect of ChCC on schooling outcomes in the optimal window - early versus late roll-out group

		Early	Late
0	Panel 1: Standardized math scores		
1	RD Estimate	-1.012	4.355
2	P-Value	0.035	0.000
3	Panel 2: Standardized reading scores		
4	RD Estimate	-0.527	4.355
5	P-Value	0.326	0.000
6	Panel 3: Grade point averages		
7	RD Estimate	-0.003	0.059
8	P-Value	0.736	0.000
9	No. of observations (left)	16,116	18,208
10	No. of observations (right)	24,685	27,521

Notes: The table shows local RD effects and coefficient p-values of ChCC's effect in the optimal window around the cutoff by roll-out group. The estimation considers all students born two months before and after the roll-out of ChCC as well as those born during the roll-out. The early roll-out group consists of all municipalities, which implemented ChCC before 2008. The late roll-out group consists of those municipalities, which implemented ChCC during the second phase of its roll-out. Panel 1 shows the results for standardized math scores, Panel 2 the ones for standardized reading scores, and Panel 3 the ones for grade point averages. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

1.9 Potential Drivers of Improved Schooling Outcomes

ChCC is a comprehensive program addressing several different aspects of early childhood development. To shed some light on the mechanisms that could drive the positive outcomes on education in middle childhood, I analyze the program's impact on intermediate outcomes. Following the current state of the literature and the multi-sectoral setup of the program, I explore three key channels, which could drive my results.

First, Cunha, Heckman, and Schennach (2010) show that cognitive and non-cognitive skills are important drivers behind school performance.[27] Based on these results, I investigate ChCC's impact on these skills. Second, work by Currie and Almond (2011) and Almond, Currie, and Duque (2018) stresses the importance of the home environment for learning outcomes. Therefore, I analyze ChCC's effects on parent-child relationships and the home environment. Third, the literature shows that early childhood education significantly influences child development outcomes and educational outcomes later in life (Temple and Reynolds (2007); Carneiro and Ginja (2014); Williams (2019)). Due to this, I analyze if ChCC results in increased attendance rates in early childhood education facilities.

I apply the same local randomization approach described earlier in this paper. I report local RD estimators in the optimal cutoff window of four periods when abstracting from polynomial orders. I also present results on the linear local randomization approach using a bandwidth of 20 periods. Table 1.7 presents summary statistics for the two scenarios. In what follows, I report local RD estimators and coefficient p-values for these model specifications. Table 1.8 presents the results. For a detailed description of the variables at use see the data section.

The evidence supporting significant effects of ChCC on cognitive and non-cognitive child development is limited. Table 1.8 shows that exposure to ChCC decreases TEPSI Scores. The coefficient reported in Column 1 and 3 is -1.89 and the p-values reported in Column 2 and 4 are close to 0.01. This result is surprising as it implies a decline in children's cognitive development. These negative effects could be related to a more rigorous evaluation of this dimension after the roll-out of ChCC or to unintended adverse effects. The program does not alter children's TEVI or TADI score.

To shed further light on potential drivers behind the program's diverging impact on educational outcomes by gender, I investigate its effects for boys and girls separately. I find that the negative impact on TEPSI Scores is only significant for boys (see Table A.10

[27] Cunha, Heckman, and Schennach (2010) analyze the interaction between cognitive and non-cognitive skills and their importance for learning outcomes. They find that students with higher cognitive and non-cognitive skills in the early years of life are more successful in learning these skills later in life. These skills then affect a variety of outcomes, as, for example, test scores, schooling, and wages.

Table 1.7: Summary Statistics of intermediate outcomes in optimal windows

	Bandwidth=4		Bandwidth=20	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Psychomotor dev. (TEPSI)	52.639	11.479	52.960	11.853
Hearing vocabulary (TVIP)	102.222	17.124	102.265	17.185
TADI	50.091	9.009	50.529	9.011
Executive functiong (BDS)	44.630	7.506	44.931	7.360
Executive functiong (HTKS)	49.039	10.841	49.380	10.846
Abnormal behavior (CBCL1)	57.581	10.969	57.392	11.208
Abnormal height (ECD)	0.196	0.397	0.207	0.405
Parental Stress Index (PSI)	40.223	34.892	41.346	35.332
Self-confidene (PSCS)	67.307	10.428	66.801	10.428
Depression Scale (CESD)	7.203	5.271	7.205	5.417
Gender-neutral parenting	0.831	0.375	0.834	0.372
Space for toys	0.889	0.314	0.894	0.308
Learning equipment	0.715	0.452	0.708	0.454
Books	0.878	0.327	0.872	0.334
Reading (Mom)	0.453	0.498	0.451	0.498
Sharing meals	0.855	0.352	0.860	0.347
Lacking dental care	0.367	0.482	0.363	0.481
Early childhood educ. (ECE)	0.503	0.500	0.499	0.500

Notes: The table shows summary statistics of the intermediate outcomes investigated in this paper. I restrict the sample to the optimal window. The first two columns report summary statistics in a window length of four while the last two columns report them in a window length of 20. I leverage data from the ELPI survey. For this purpose, I pool survey waves from 2010, 2012, and 2017. *TEPSI* is a score that measures children's psycho-motor development. *TVIP* is a norm-referenced measure of Spanish hearing vocabulary analyzing verbal reasoning, as well as language skills. *TADI* scores evaluate children ages three months to six years and measures four dimensions of child development: cognition, motor skills, language and socio-emotional development. *BDS* and *HTKS* measure children's executive functioning and *CBCL1* behavioral abnormalities. *PSI* is an index measuring parental stress, *PSCS* is a perceived self-confidence scale, *CESD* is the Center for Epidemiologic Studies Depression Scale, and *HOME* is the Home Observation Measurement of the Environment Index. All other variables are dummy variables. *ECE* stands for early childhood education. I weight each variable by the survey weights. Source: ELPI (2010-2017).

Table 1.8: Local RD effect of ChCC on intermediate outcomes

Specification	Bandwidth=4	P=0	Bandwidth=20	P=1
Variable	RD Estimate	P-Value	RD Estimate	P-Value
Panel 1: Cognitive child development outcomes				
Psychomotor dev. (TEPSI)	-1.894	0.012	-1.891	0.013
Hearing vocabulary (TVIP)	-0.139	0.875	0.971	0.251
TADI	0.631	0.304	0.818	0.165
Panel 2: Non-cognitive child development outcomes				
Executive functioning (BDS)	0.446	0.353	0.428	0.362
Executive functioning (HTKS)	1.168	0.102	1.325	0.055
Abnormal behavior (CBCL1)	-0.704	0.157	-0.102	0.834
Abnormal height (ECD)	-0.025	0.212	-0.015	0.431
Panel 3: Home environment and parent-child interactions				
Parental Stress Index (PSI)	-2.278	0.3	-1.986	0.352
Self-confidence (PSCS)	0.656	0.317	0.852	0.176
Depression Scale (CESD)	-0.596	0.077	-0.583	0.078
Home environment (HOME)	0.083	0.612	0.028	0.859
Gender-neutral parenting	0.006	0.473	-0.003	0.744
Space for toys	0.006	0.006	-0.010	0.235
Learning Equipment	0.005	0.005	0.027	0.001
Books	0.016	0.07	0.028	0.001
Reading (mother)	-0.023	0.336	0.037	0.106
Sharing meals	-0.008	0.637	-0.005	0.768
Lacking dental care	0.011	0.365	0.036	0.003
Panel 4: Early childhood education				
Attends ECE	0.005	0.665	0.027	0.021

Notes: The table shows two different local RD estimators and the respective coefficient p-values on intermediate outcomes of exposure to ChCC. The first two columns refer to a local randomization approach considering the optimal window around the cutoff and abstracts from polynomial orders (P=0). The next two columns consider 20 windows around the cutoff and a polynomial order of one (P=1). Panel 1 presents the results on cognitive child development outcomes. *TEPSI* is a score that measures children's psycho-motor development. *TVIP* is a norm-referenced measure of Spanish hearing vocabulary analyzing verbal reasoning, as well as language skills. *TADI* scores evaluate children ages three months to six years and measures four dimensions of child development: cognition, motor skills, language and socio-emotional development. Panel 2 presents the results on non-cognitive child development outcomes. *BDS* and *HTKS* measure children's executive functioning and *CBCL1* behavioral abnormalities. Panel 3 reports estimators and coefficient p-values on parent-child interactions and the home environment. *PSI* is an index measuring parental stress, *PSCS* is a perceived self-confidence scale, *CESD* is the Center for Epidemiologic Studies Depression Scale, and *HOME* is the Home Observation Measurement of the Environment Index. All other variables are dummy variables. *ECE* stands for early childhood education. For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016). Source: ELPI (2010-2017) and Clarke and Tapia-Schythe (2021).

and A.11). One possible interpretation is that the program achieved a greater inclusion of girls, with negative effects on boys. Similar patterns arise when dividing the sample by socioeconomic vulnerability (see Table A.12 and A.13). Privileged children experience declines in TEPSI Scores, while there is no significant impact of ChCC on socioeconomically vulnerable children.

In addition, the program seems to increase children's executive functioning when measured via the HTKS. The coefficient is around 1.2 and significant at the 10 and 5 percent significance level. My analysis by gender illustrates that this is driven by positive effects on girls' executive functioning (see Table A.11). The impact is significant at the 5 percent significance level. Boys do not seem to experience positive effects on their executive functioning (see Table A.10). Similarly, executive functioning only improves for socioeconomically vulnerable children (see Table A.13).

To measure the program's impact on the home environment, I rely on the official evaluation instruments measuring intra-household relations included in the ELPI survey. These are the parental stress index (PSI), the self-assessment of parenting skills (PSCS), the presence of depressive symptoms (CESD), and the home environment index (HOME). For a detailed explanation of these evaluation instruments see the data section. I also consider a number of additional variables measuring the home environment: the degree of gender-neutral parenting, the space at home available for children's toys, the amount of learning equipment and books at home, if the mother reads out to her child, if parents share a meal together with their child, and evidence of inappropriate dental care.

Table 1.8 shows that there is some evidence of the program altering parental outcomes. The p-value of the CESD reported in Column 2 and 4 is below 0.1. This indicates that program exposure significantly decreases the incidence of parental depression at the 10 percent significance level. The point coefficient in Column 1 and 3 indicates that the program decreases the index by approximately -0.595 index points. When restricting the sample by gender, it becomes evident that the program improves parental outcomes for girls but not for boys. Girls' parents report a lower parental stress index, higher self-confidence, and less depression as a result of ChCC (see Table A.11). These effects are significant at the 1 percent significance level for most coefficients. A different picture emerges for boys. There is no significant improvement on these dimensions for boys' parents (see Table A.10). Similarly, these indicators only report significant improvements for socioeconomic non-vulnerable children (see Table A.12). There are no significant changes for parents of less privileged children on these dimensions (see Table A.13).

In addition, the program seems to increase the availability of learning equipment and books at home. The p-value on these two coefficients is below 0.1 in the case of learning equipment for both model specifications. Exposure to ChCC increases the probability that a household disposes of learning equipment for children by 0.005 (0.027) percentage points. Compared to the mean, this is an increase of 0.711 (3.762) percent. Similarly, ChCC alters the probability to have children books at home by 0.016 (0.028) percentage points. This is equivalent to a relative increase of 1.839 (3.248) percent. Lastly, while the p-value on available space for toys in Column 2 is below 0.01 it increases to 0.235 in Column 4. Consequently, while there is some evidence on the program increasing the share of households with available space for toys, this evidence does not hold under the local linear randomization approach.

My heterogeneity analysis shows that these improvements in parental outcomes are solely significant for boys but not for girls (see Table A.10 and A.11). Analyses that make a distinction by socioeconomic status confirm that there are significant and positive effects for the more privileged children (see Table A.12). Socioeconomically vulnerable children experience a significant increase in the number of children books at home, but also some adverse effects (see Table A.13). There is a significant decrease in the space available for toys. In addition, dental care deteriorates for these children.

To sum up, ChCC has significant effects on mechanisms within households. These mechanisms differ significantly by gender and socioeconomic status. While there are significant improvements in the executive functioning and several parental outcomes for girls, the program's impact is limited to material channels in the case of boys. Analyses by socioeconomic vulnerability also show important differences. Parenting outcomes significantly increase for the more privileged children. Socioeconomically vulnerable children experience improvements in their executive functioning, but also adverse effects.

ChCC also involved an increase in the supply of early childhood education provided by the public sector. To measure the effectiveness of this channel, I analyze the program's impact on the attendance rate of children in early childhood education facilities. Panel 4 in Table 1.8 shows that there is some evidence supporting increased attendance rates as a result of the program. The p-value in Column 4 is below 0.05. Therefore, the point estimate of the local linear randomization approach is significant at the 5 percent significance level. ChCC increases the probability to attend an early childhood education facility by 0.027 percentage points. This is a relative increase of 5.294 percent when compared to the average. Still, the significance of this result does not hold when abstracting from polynomial orders. The p-value in Column 2 is 0.665.

The evidence suggesting raising attendance rates only holds for boys not for girls (see Table A.10 and A.11). Socioeconomically vulnerable children do not seem to benefit from the increased supply of early childhood education (see Table A.13). Raising attendance rates are limited to the more privileged children (see Table A.12). These findings could explain why the program's impact on educational outcomes in middle childhood are larger for boys and the more privileged children.

1.10 Cost-Benefit Analysis

To compare the costs and benefits of ChCC to alternative programs, I use a framework developed by Hendren (2016) and Hendren and Sprung-Keyser (2020) to calculate the marginal value of public funds (MVPF). Benefits are captured by beneficiaries' willingness to pay and costs capture initial program spending and fiscal externalities. The MVPF is the ratio of both.

To calculate beneficiaries' willingness to pay, I estimate the average lifetime earnings in Chile based on the 2017 socioeconomic household survey (CASEN). I first calculate the mean income of all individuals included in the survey by age. I restrict the working population to all individuals between the ages of 18 and 64. I then assume a discount rate of 3 percent and calculate the average present value of lifetime earnings (PVLE) in Chile. I estimate that the PVLE in Chile is 220,312.4 US-Dollar.[28]

Based on data provided by the government of Chile, the estimated average unit cost of ChCC is 23,647.2 US-Dollar. Distributing these 23,647.2 US-Dollar over the life of an average person in Chile yields a present value of the annual average unit costs of 12,864.5 US-Dollar. Work by French et al. (2015) shows that a 1 percent increase in GPAs leads to an average increase in income of around 12 to 14 percent. The average GPA in my sample is 5.8 (see Table 1.1). Based on the different model specifications investigated in this paper, the average impact of ChCC on GPAs is approximately 0.031 points. This corresponds to an increase of 0.53 percent. The equivalent increase in income would therefore be approximately 7.5 percent.

A 7.5 percent increase in lifetime earning leads to a difference in the present value of lifetime earnings between the pre- and post-program world of 16,523 US-Dollar per participant and an additional present value of tax revenues of 1,156.6 US-Dollar per participant. The MVPF is then 1.41 per participant.

The MVPF is lower than the MVPF of the Food Stamp Program in the US of 56.25 (Bailey et al. 2020) or the Perry Preschool Project's MVPF of 43.61 (Hendren and Sprung-Keyser 2020). This could be due to the fact that girls and the socioeconomically vulnerable do not benefit equally from the program and that several of the intended channels seem to be unaffected by the program.

[28]Appendix A.6 presents a detailed overview of the methodology.

1.11 Discussion and Conclusion

In this paper I investigate the effect of comprehensive and universal early childhood development programs on educational outcomes in developing countries. In particular, I study the effect of the pioneer program *Chile Crece Contigo* on schooling outcomes 12 years after its introduction. The program is universal and combines components of health, education and parental care. I find that exposure to ChCC has positive effects on grade point averages as well as standardized math and reading test scores. Effects are more marked for boys than for girls. The program's impact is smaller for socioeconomically vulnerable children. My findings pass several robustness tests including alternative polynomial orders and variations in the underlying bandwidth. Furthermore, my results hold when employing an alternative empirical estimation strategy, namely an event study design.

The positive impact on school performance seems to be driven by improvements in intra-household relations and increased attendance rates. Still, these effects differ by gender and socioeconomic status. For girls, I find significant improvements in several parental outcomes as well as in their executive functioning. Boys seem to benefit by an increase in the availability of material goods. Raising attendance rates are limited to the more privileged children and boys. The latter could explain why the program's impact on educational outcomes is larger for boys and socioeconomically non-vulnerable children.

My work contributes to a large literature analyzing early childhood development. I show that a comprehensive approach to early childhood development can lead to improvements in child development across several dimensions. My findings illustrate that, in addition to targeted programs for poor children, universal interventions also have positive effects on child development. Still, the relatively low MVPF indicates that targeted programs with a higher MVPF might be more effective in reducing human capital gaps. This paper also fills the gap of the "missing middle years", showing positive effects of investments in early childhood on outcomes observed during middle childhood.

Policymakers should use the insights of this paper to design more integrated and comprehensive approaches to early childhood development and to decrease disparities in human capital. My work also illustrates potential limitations of these types of programs. I show that several of the multi-sectoral entry points might not work as intended. In addition, policymakers should pay special attention to the gender dimension of such programs so that boys and girls benefit equally. Additionally, there is a need for the development of mechanisms, which maximize the effect of such programs on the most vulnerable children. Lastly, I illustrate that the program's impact is larger for municipalities implementing the program during the second phase of its roll-out. This makes a case for piloting early childhood interventions.

Further research should analyze if trajectory effects carry over from early to middle childhood into adulthood, and study the impact of ChCC on long-term outcomes, like tertiary education, wages and health in the long run.

Chapter 2

Twitter and Crime: the Effect of Social Movements on Gender-Based Violence

"If people were silent, nothing would change." - Malala Yousafza

This chapter is based on joint work with Michele Battisti and Ilpo Kauppinen.

2.1 Introduction

Gender-based violence (GBV), which refers to violence against individuals based on their gender, remains a problem in contemporary societies. According to UN Women (2021) every third woman falls victim to some sort of GBV at least once in their life. In addition to personal costs, such as physical, mental, justice system costs, lost economic output, social welfare expenses and the need for specialized support services (Walby and Olive 2014). It is therefore crucial to study GBV, its drivers and potential methods to reduce it. However, GBV is surrounded by social stigma, shame, discriminatory and stereotypical attitudes and other factors that lead to many victims remaining silent.[1] The emergence of social movements such as the *#metoo* movement has, in contrast, led many to openly share their experiences of GBV on social media.

This paper asks whether social movements on Twitter affect GBV-related crime rates and arrest per crime rates.[2]) We focus on Twitter as one of the most well known social movements, the *#metoo* movement, took place on Twitter. Our key hypothesis is that these movements increase the social costs and social pressure of committing GBV and hence reduce the prevalence of GBV. In addition, we hypothesize that victims feel empowered and increasingly report GBV and given this, perpetrators may perceive increased costs also via this channel. This paper is the first to conduct an in-depth analysis of how social movements on Twitter (hereinafter: Twitter social movements) impact crime rates and arrest per crime rates related to GBV. Our paper also makes an important contribution to the economic crime literature as it explores innovative ways on how to distinguish between crime perpetration and reporting of GBV-related crime rates. Moreover, we shed light on the role of social stigma and tabooing in the alteration of GBV, a highly understudied question to date. Lastly, we investigate to which extent the polarity of written text plays a role in our results via a sentiment analysis.

In this paper, we first construct a novel dataset on conversations around GBV using data from Twitter. We use a set of machine learning techniques applied to hashtags used on Twitter to construct a weekly measure of the prevalence of conversations about GBV on Twitter across federal states in the United States and over time. Our Twitter-based dataset consists of around 11.4 million tweets for the period 2014-2016. We focus on this period as we match this data to crime-incidence level data and arrests gathered by the FBI,

[1]For instance, a study by Palermo, Bleck, and Peterman (2014) on GBV in 24 developing countries finds that only seven percent of female victims of GBV had reported to the authorities or other formal institutions. The problem of under-reporting in GBV is well established in the literature (see for example Joseph et al. (2017) or Fernández-Fontelo et al. (2019)).

[2]For the working paper version of this chapter see Battisti, Kauppinen, and Rude (2022).

which is only available up to 2016 at the time of this study. We take advantage of the high frequency of our datasets, and conduct regressions at the federal state by week level. Our main outcome variables are the crime rate as well as the arrest per crime rate. The main explanatory variable is the number of tweets per 100 cellphone internet subscriptions at the federal state by week level. We control for month of the year and federal state fixed effects to account for potential confounding factors. We introduce lagged coefficients as people may take some time to react to what they observe online. The introduction of lagged coefficients also establishes the causal direction of our estimators. It is unlikely that future crime rates would affect the current number of tweets.

We confirm the causality of our findings by employing an event study design. For this purpose, we exploit the sudden emergence of a Twitter social movement in 2014 using the hashtag *#yesallwomen*. *#yesallwomen* is a hashtag under which Twitter users share their experiences of GBV and misogyny. It was first used after the *Isla Vista Killings* in 2014. In terms of magnitude, it outnumbered the *#metoo* hashtag in the first month after its emergence. We exploit the fact that the movement spread to different states at different points in time.

Our results show that the number of tweets per 100 cellphone internet subscription decreases GBV-related reported crime rates per 100,000 inhabitants to the authorities by about 1 percent. If Twitter social movements also empower victims and lead to more reporting, our coefficients are lower bound estimates of the true underlying effect of Twitter social movements on GBV. Moreover, our evidence points to positive and statistically significant effects of social movements on GBV-related arrest per crime rates.

To establish if the decrease in crime rates as a response to Twitter social movements is due to victims reporting these crimes less or perpetrators committing these crimes less, we employ three empirical strategies. First, we analyze the impact of Twitter social movements on Google search activities on informal support networks. The underlying rationale is that crime reporting could decrease as victims substitute formal for informal support networks. These regressions reveal no clear pattern of results.

Second, we investigate if tweets in favor of conservative gender norms, such as tweets using the hashtag *#alphamale*, impact GBV-related crime rates differently than tweets about GBV. If the number of tweets using *#alphamale* has a positive effect on GBV-related crime rates, this might point towards perpetrators changing their behavior and committing these crimes more. Therefore, analyzing the impact of Twitter tweets in favor of conservative gender norms can help to shed light on the reporting or perpetrator channel. The analysis demonstrates that there is limited evidence in favor of a positive effect.

Lastly, we restrict our outcome variable to GBV-related violent crimes, namely homicides or aggravated assault. The idea is that victims of violent crimes are less likely to report these crimes. Hence, crime rates of violent crimes are more likely to mirror crime perpetration instead of reporting behavior. The pattern of results suggests a negative impact of Twitter social movements on GBV-related violent crime rates.

These analyses illustrate that decreasing crime rates are most likely caused by changes in the behavior of perpetrators, instead of a decrease in reporting behavior. This finding is in line with the main economic perspective on crime according to which criminal activity varies with the price of conducting a crime (Becker et al. 1995). This price increases with peer pressure and neighborhood effects, which emerge from social interactions (Falk and Fischbacher 2002). The number of Twitter tweets would, in this case, be a signal of this peer pressure. Moreover, perpetrators might observe the social prosecution of those alike online, and interpret this as neighborhood effects. If Twitter social movements also empower victims and lead to more reporting, our coefficients are lower bound estimates of the true underlying effect of Twitter social movements on GBV. Twitter might act as a facilitator for the signaling of shifting social norms. This channel would be in line with previous evidence on the erosion of existing social norms, such as work by Bursztyn, Egorov, and Fiorin (2020) showing significant effects of Donald Trump's rise in polarity on publicly expressed xenophobic views.

The second stage of the analysis differentiates between different types of GBV. We believe that we can shed light on the role of social stigma and tabooing by making this distinction. We find that the impact of Twitter use on crime rates and arrest per crime rates is strongest in the case of sexual violence. We interpret these findings as stigmatization, tabooing and silencing being especially persistent in the case of sexual violence.

Lastly, this paper asks if the polarity of the tweets' text plays a role in affecting human behavior offline. For this purpose, we apply text analysis methods to the tweets in our Twitter sample to study sentiments involved in the overall conversation. Our analysis of the tweets' written content shows that the polarity of tweets does not play a significant role in the change of crime rates or arrest per crime rates. Consequently, what matters is the pure magnitude of Twitter social movements.

We establish the validity of our findings by employing four robustness tests. First, to ensure that our findings are not driven by simultaneous unobserved shocks, we conduct placebo regressions. For this purpose, we use non-GBV related crime rates and arrests per crime as an outcome variable. These placebo regressions validate our findings. Second, we show that our main findings align with the evidence generated from the event study design. Third, we investigate if Twitter users engaging in the GBV-related debate and victims of GBV systematically differ from each other. To do so, we employ a face recognition technique to Twitter users' profile pictures and deduce their age and ethnicity. In addition, we deduce the authors' gender from their first name. Our evidence shows that, while victims of GBV and those tweeting about it are on average similar in their age and ethnicity, there are some systematic differences in their gender. A larger share of victims of GBV than Twitter users are female. Lastly, accounting for spatial spillovers between neighboring states confirms our main results.

This paper makes a significant contribution to the economic literature studying GBV. To the best of our knowledge, this is the first paper using Twitter data to study the effect of online social movements on GBV-related crime rates and arrest per crime rates. While there has been a recent interest in this research question, no paper so far has used social media data. Closest to our paper is Levy and Mattsson (2021), who analyze Google Trends data. Their analysis focuses on the extent to which individuals try to get informed on this topic, rather than on online conversations. Moreover, we take advantage of the tweets' text and study sentiments around Twitter social movements related to GBV. We therefore believe that our analysis goes one step further in analyzing the conversation taking place on social media.[3]

Furthermore, our work contributes to economic papers studying potential drivers of GBV. To name a few examples, Aizer (2010) shows that a decreasing wage gap comes along with a decrease in domestic violence at the household level in the US. Related work by Bhalotra et al. (2021a) illustrates that an increase in male unemployment or a decrease in female unemployment increases intimate partner violence (IPV). Similarly, Brassiolo (2016) demonstrate that a decrease in divorce costs leads a decrease in IPV. Closely related work by González and Rodríguez-Planas (2020) finds that gender norms are important drivers of IPV. Related work analyzes the association between polygony and IPV (Cools and Kotsadam 2017), family structures and IPV (Tur-Prats 2019), and colonialism and IPV (Guarnieri and Rainer 2021). There is also an increasing literature studying the impact of COVID-19 related lock-downs on IPV (Agüero (2021); Berniell and Facchini (2021); Bullinger, Carr, and Packham (2021)).

Our paper advances the current understanding of potential strategies to reduce GBV. Previous research studies focus on public transfer programs (Bobonis, González-Brenes, and Castro 2013), anti-poverty programs (Amaral, Bandyopadhyay, and Sensarma 2015), employment of female police officer (Miller and Segal (2019); Amaral, Bhalotra, and Prakash

[3]The paper by Levy and Mattsson (2021) focuses on an international setup while our paper takes place in the US and considers lower geographic variation. Lastly, differently from their paper, which solely relies on the *#metoo* movement, we focus on Twitter social movements taking place in earlier periods. We believe that focusing on the years prior to the *#metoo* movement is a more appropriate setup for our underlying research question based on lower awareness on GBV, lower digitalization, and less exposure to confounding factors such as the election of President Trump.

(2021)), or the exposure to video dramas (Cooper, Green, and Wilke 2020). Morrison, Ellsberg, and Bott (2007) give an early literature review on potential interventions.

Our work also talks to the literature on the impact of experiencing GBV. Example studies within this stream of literature are studies by Welsh (1999), Fitzgerald and Cortina (2018), or Folke et al. (2020) finding negative effects for victims of sexual harassment. A number of related papers shows negative effects of GBV on economic activities (Duvvury et al. (2013); Ouedraogo and Stenzel (2021)). Similarly, GBV leads to work deterioration (Chakraborty et al. (2018); Siddique (2022)).

Lastly, this paper talks to the emerging literature on social movements (Besley and Ghatak (2018); Francois and Vlassopoulos (2008); Bénabou and Tirole (2006)) and social norms (Agranov, Elliott, and Ortoleva (2021); Viscusi, Huber, and Bell (2011)). One particular type of social movements studied more extensively in the political economy literature are political protests (Bremer, Hutter, and Kriesi (2020); Bursztyn et al. (2021); Matta, Bleaney, and Appleton (2021)). Recently, several paper have analyzed the impact of the *#blacklivesmatter* movement (Dave et al. (2020); Agarwal and Sen (2022)). We contribute to this literature by showing that social movements in online spaces translate into offline behavioral changes, indicating that Twitter is a facilitator of the signaling of social norms. These findings are in line with a limited number of studies illustrating a significant association between social media usage and hate crimes (Müller and Schwarz 2020) as well as political outcomes (Levy (2021); Zhuravskaya, Petrova, and Enikolopov (2020)).

Our findings have several important policy implications. Those who are interested in decreasing the prevalence of GBV should explore the potential of social media platforms to do so. Our findings also point to the important role of stigmatization, tabooing and silencing. Policymakers should design strategies to facilitate the reporting of and conversation on GBV. Moreover, they should design interventions addressing harmful gender norms and leading to long-term changes in the beliefs and attitudes concerning GBV.

The rest of the paper proceeds as follows. Section 2.2 presents our economic rationale, the definition of GBV used in this paper, and a review of the related literature. Section 2.3 outlines the creation of our Twitter dataset and describes additional datasets used in this paper. Section 2.4 explains our methodology and section 2.5 presents our main results. Section 2.6 analyzes some of the mechanisms behind our main findings, while section 2.7 presents our robustness checks. Section 2.8 concludes.

2.2 Economic Relevance and Motivation

2.2.1 The Economics of Gender-Based Violence

The United Nations defines GBV as violence against individuals based on their gender (UNHCR 2022). It can be of physical, sexual, mental or economic nature. GBV can take place in private, for instance in the case of child abuse or intimate partner violence, or in the public space, as for example in the form of rape or street harassment. Victims of GBV can be both male and female, or non-binary, as long as the root of it lies with gender inequality, the abuse of power and harmful social norms. Child marriage, female genital mutilation, and honour crimes also form part of GBV. Globally, one third of every woman falls victim of GBV at least once in their lifetime.

GBV has economic relevance as it generates large costs for both individuals and for the society as a whole. The World Bank estimates that costs related to GBV amount to one to two percent of GDP (Duvvury et al. 2013). Nevertheless, it is still a largely understudied topic.

GBV is related to the economic field through three channels. First, one form of GBV is economic violence. One example of economic violence is the control of female-owned property or financial resources, as well as their exclusion from those resources, as is the case when women are excluded from inheritance or property rights. This, on the other hand, affects female empowerment in general and creates serious barriers for countries to secure their full economic potential.[4]

Second, GBV leads to serious economic harm for those who fall victim to it. Ouedraogo and Stenzel (2021), for example, explore the economic consequences of violence against women in Sub-Saharan Africa. They show that an increase in GBV decreases economic activity. In fact, an increase in the share of women experiencing GBV of one percentage point, leads to a downturn of economic activity by up to eight percentage points. A related study by Chakraborty et al. (2018) studies the effect of crime against women on women's work force participation in India. The authors show that it leads to serious work deterioration. They also demonstrate that the impact is larger for women from more conservative families and the lower end of the wage distribution. Similarly, Siddique (2022) shows that an increase in media reports about sexual assault in India decreases female labor force participation. On the contrary, an increase in crime rates results in raising male labor force participation (Mishra, Mishra, and Parasnis 2021). Further studies on the negative effects of sexual harassment have been conducted by Welsh (1999), Fitzgerald and Cortina (2018) and Folke et al. (2020).

[4]See for example Swamy (2014) on the positive effect of female financial inclusion on poverty reduction.

Third, several studies show that economic circumstances affect the occurrence of GBV. An example study by Li, Pandya, and Sekhri (2019) show that global economic integration, measured through foreign direct investment, leads to an increase in female economic empowerment and a decrease in rape incidences in India. Household bargaining power also plays an important role. Related work by Aizer (2010) shows that a decreasing wage gap comes along with a decrease in domestic violence at the household level in the US. Similarly, an increase in male unemployment or a decrease in female unemployment increases IPV (Bhalotra et al. 2021a). Cools and Kotsadam (2017) illustrate that resource inequality is associated with higher abuse and a decrease in divorce costs leads to a decrease in IPV (Brassiolo 2016). Moreover, recent work by Bullinger, Carr, and Packham (2021) shows that COVID-19 lock-downs led to an increase in domestic violence-related calls for police service.

In addition, a recent body of literature analyzes how to decrease GBV. As an example, Amaral, Bhalotra, and Prakash (2021) show that opening of women police stations leads to an increase in police reports of crimes against women. Miller and Segal (2019) show similar evidence on an increased share of female police officers in the US. In addition, Cooper, Green, and Wilke (2020) come to the conclusion that the exposure to videos that dramatize violence against women and girls (VAWG) increases its reporting. Amaral, Bandyopadhyay, and Sensarma (2015) study the effect of an anti-poverty programme in India on GBV and find that an increase in female labor force participation leads to an increase in GBV. On the contrary, Bobonis, González-Brenes, and Castro (2013) show that beneficiary women of a public transfer program in Mexico are less likely to experience physical abuse. Others have shown a significant effect of municipal female political leaders on GBV in the US (Wen 2021), Brazil (Delaporte and Pino 2022) and India (Iyer et al. 2012).

Recently, more studies explore new data sources to analyze the topic of GBV. One example is work by ElSherief, Belding, and Nguyen (2017) from computational science. They create a one percent sample of Twitter's public stream and then filter on certain keywords or hashtags related to GBV. They then measure certain attributes of these tweets, such as user engagement, linguistic properties and sentiments, among others. Similarly, Khatua, Cambria, and Khatua (2018) derive insights about the occurrence of different forms of sexual violence by analyzing 700,000 tweets from the *#metoo* movement.

To the best of our knowledge, there is very limited application of social media data to study GBV in the economic literature to date. While Levy and Mattsson (2021) analyze a similar research question to ours, they rely on google trends data. The authors use this data as their main measure of the *#metoo* movement's intensity by country. While they use a pre-existing Twitter dataset to validate this approach, our paper uses Twitter data per se. Their work also refrains from using social media to analyze emotions or sentiments around this topic. Additionally, they restrict their Twitter dataset to the tweets from October 2017 which were geotagged. Consequently, our paper is an important contribution to their work, as it goes one step further in the depth of the underlying research question. We leverage Twitter data for a broader time period, make use of the tweets' text, and investigate the dynamics at a more dis-aggregated level, namely at the week by state level in the United States.

2.2.2 The Economics of Crime

Economists mostly base their understanding of criminal behavior on a cost-benefit-approach. The first one to apply theoretical rationales to criminology was Becker et al. (1995). The author based his understanding of criminal behavior on an individual choice model. In this model, individuals commit crimes as soon as the benefits of the potential criminal act outrun the costs. Benefits can be of monetary or non-monetary nature, such as feeling a thrill of danger, excitement, entitlement, or satisfaction. The latter two might be especially important when applying the theory of crime to GBV. Costs from committing a crime, on the other hand, can take different forms as well: material costs, psychic costs, such as fear, guilt, anxiety, social punishment, as well as opportunity costs. Opportunity costs could be lost income due to the time spent in prison. Lastly, there are direct punishment costs, meaning legal fees and formal and informal sanctions.

The perceived costs and benefits of committing a crime depend on individual characteristics (Becker et al. 1995). A gang member's peer recognition from committing a crime will be higher than the one experienced by a non-gang member. To give another example, opportunity costs vary with the potential income level of individuals. Poorer people might experience lower opportunity costs of committing a crime than richer people.

Our paper tests the model's empirical implications, applying it to the setting of Twitter social movements and GBV. We ask if Twitter social movements create potential costs to perpetrators due to a perceived increase in peer pressure and informal sanctions. Moreover, legal and police authorities might also feel pressured by society and increase the degree to which they control and punish these type of crimes which could increase the potential punishment costs. Based on this rationale, we would expect to see a negative effect of social movements on GBV-related crimes committed.

On the contrary, the perceived benefits could also increase under the model when applied to our underlying research question. First of all, the degree of thrill and excitement might increase when the perceived costs increase. Next, perpetrators might increase their engagement in GBV, as they might see a long-term benefit from protecting the status-quo. This type of backlash behavior has been outlined in previous studies (see for example Amaral, Bandyopadhyay, and Sensarma (2015), or Bandyopadhyay, Jones, and Sundaram (2020)). This would lead to a positive impact of Twitter social movements on GBV.

Another possibility for a positive effect of Twitter tweets on GBV is a scenario, in which victims of GBV feel empowered and report these types of crimes more often. To understand this potential channel in more detail, we next look at the economic literature on social movements.

2.2.3 The Economics of Social Movements and Networks

Our work talks to the economic literature on social movements, although to a lesser extent. Breton and Breton (1969) develop a supply- and demand framework of social movements. Since then, several scholars have stressed the importance of social movements for the provision of public goods (Besley and Ghatak 2018), the delivery of social services (Francois and Vlassopoulos 2008) and the design of incentives (Bénabou and Tirole 2006). Others have analyzed its impact on the provision of social policies (Amenta and Halfmann 2000) or wages (Rudé 1954). While the economic literature on social movements is scarce, one specific form of social movements has been studied more in depth: the impact of political protests. To name some examples, Bremer, Hutter, and Kriesi (2020) study the impact of social movement on electoral outcomes. Similarly, Bursztyn et al. (2021) investigate how political protests and political engagement interact with each other. Related work by Matta, Bleaney, and Appleton (2021) estimate the overall economic impact of mass protests. Recently, several paper have analyzed the impact of the *#blacklivesmatter* movement (Dave et al. (2020); Agarwal and Sen (2022)).

Our paper brings a new aspect to this literature through analyzing the role of social movements in the online space. We ask whether these online movements translate into real behavioral changes. This can shed further light on the role of social movements in political and economic change. Recently, the importance of social norms and movements as drivers of lasting change have gained more attention (see for example Agranov, Elliott, and Ortoleva (2021) or Viscusi, Huber, and Bell (2011)), and the work at hand provides additional evidence to this stream of literature. Moreover, our paper provides answers to the question on the role of social media platforms for public policies (for a detailed review on this literature see Zhuravskaya, Petrova, and Enikolopov (2020)). A limited number of studies illustrate a significant association between social media usage and hate crimes (Müller and Schwarz 2020) as well as political outcomes (Levy 2021).

2.3 Data

The paper at hand generates a novel dataset on GBV-related Twitter tweets. In addition, it leverages a variety of different datasets to answer the underlying research question. In the following, we describe these in more detail and outline the construction of the Twitter dataset used in this paper.

2.3.1 Tweets related to Gender-Based Violence

In the following we outline the construction of the Twitter dataset used in this paper, which approximates GBV-related social movements taking place on Twitter. We restrict our dataset to the years 2014-2016, as this is the time period, for which we have data available on crime incidences in the US. Our goal is to construct a dataset of English Twitter posts that is as representative as possible of GBV-related social movements taking place on Twitter during this period. To generate our dataset, we take advantage of the Twitter API for Academic Research. This API gives academic researchers access to the full universe of Twitter tweets since the first Twitter tweet in 2006. The academic access is subject to a monthly cap and a rate limit. This is why we need to filter for a subset of tweets when retrieving data from the API. To do so, we apply a hashtag based approach when designing our search query with which we retrieve raw data from the Twitter API. This means that we use hashtags to filter on the full universe of Twitter tweets accessible via the Twitter API for Academic Research. For this purpose, we define a list of hashtags, which approximates GBV-related Twitter movements as good as possible.

To generate this list of hashtags, we start by looking at ten of the most well-known Twitter movements on this topic (see Annex B.1 for the concrete list of these movements). We then gather a list of all hashtags mentioned in relation with these ten movements. To this end, we first retrieve all tweets from the first month of each of the ten movements, a total of 818,003 tweets. We then list all hashtags mentioned in these 818,003 tweets, a list of 73,430 different hashtags in total. Many of these hashtags do not uniquely identify a GBV-related topic when standing on their own, which is why we classify our list of hashtags into GBV-related hashtags and unrelated hashtags.

To do so, we apply supervised learning methods and several machine learning classifiers. These methods help us to classify our sample of 818,003 different tweets and 73,430 different hashtags into the ones clearly related to the topic of GBV and those relating to broader or different topics as well as other areas of gender equality. To achieve the highest possible precision of our underlying classification exercise, we spot-check ten different machine learning algorithms.[5] We train a Multinomial Naive Bayes classifier, K-Nearest Neighbor Classifier, a Regularized Logistic Model and Support Vector Machine, a Stochastic Gradient Descent, a Decision Tree Classifier as well as several Ensemble Classifiers. Our training process involves 32,487 hashtags coded by hand. We split this hand-coded data into a training and testing dataset. We choose the best performing classifier based on a variety of different performing measures: The Accuracy Score, Precision and Recall, as well as Confusion Matrices. As our hand-coded data is characterized by a highly imbalanced classification (90 percent of the hand-coded hashtags belong to our "Garbage" category[6]), we address the imbalanced classification problem by oversampling our training data first.

Based on our different performance measures, the Linear Support Vector Machine (SVM) is the best performing classifier for our underlying classification problem. We then run the linear SVM Algorithm on the remaining 40,943 not-hand-coded hashtags. We evaluate the performance of the prediction made by the algorithm on the unlabeled data by comparing a ten-percent sample of the predicted classes to hand-coded classifications of this same sample done by an independent research assistant.

Our classification exercise results in a set of 2,009 hashtags clearly identifying topics of GBV. Next, we order these hashtags by relevance and choose the 62 most relevant hashtags to meet the query restriction of the Twitter API (a total character count of 1,024 characters). We then retrieve all tweets, including retweets, quotes, and replies from the year 2014-2016, filtering on these 62 hashtags. This results in 6,175,643 tweets for 2014, 2,685,019 tweets for 2015 and 2,474,767 tweets for 2016.[7]

Our ultimate goal is to create a dataset, which we can match to the crime-incidence level data we have at hand. There are two possibilities for that. To begin with, we can make use of the high-frequency of both of our datasets and match them at the weekly level. This implies aggregating the data to the weekly level and counting the number of tweets in each week. Figure 2.1 shows the number of tweets over time.

Another possibility is to exploit geographic variations and merge our Twitter and crime data at the state-week level.[8] We rely on users' location information because only a small share of Twitter tweets (approximately 1.5 percent) are geocoded. Around 76.5 percent of tweets identify a user location in their authors' user profile. Still, the location information

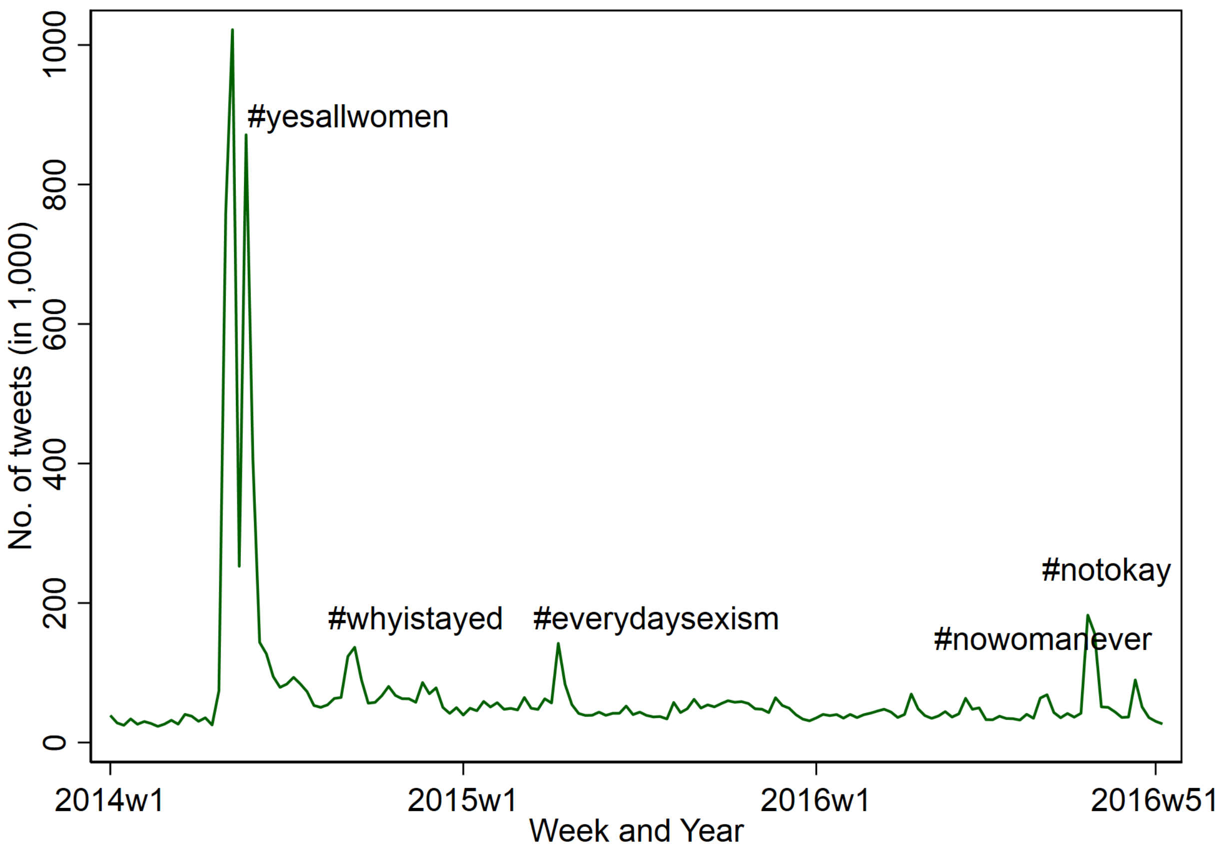
[5]Spot-checking refers to a rapid approach towards the evaluation of several machine learning algorithms. Spot-checking diverges from grid searching or the tuning of algorithms by taking a more hands-on approach. It refers to handpicking a number of algorithms and comparing their performance to each other.

[6]We define our Garbage category as all tweets not relating to the broader topic of Gender Equality. For the details see Annex B.1.

[7]For a detailed overview on how we constructed our Twitter dataset see Annex B.1.

[8]We later account for the different size of federal states by weighting our regressions by the respective population size.

Figure 2.1: Number of weekly Twitter tweets on GBV (2014-2016)



Notes: The figure shows the average number of English language tweets related to GBV generated from our Twitter API via the hashtag based approach for the period 2014 to 2016. The first spike refers to the Twitter movement *#yesallwomen*. For details behind the hashtag-based approach see Appendix B.1. The x-axis shows the respective week in a respective year. The y-axis shows the number of Twitter tweets (in 1,000). The graph does not include the *#metoo* movement, as it only took place in October 2017. Source: Twitter (2014-2016).

is not available in a unified format when accessing the data via an academic account. The administrative level varies largely, ranging from neighborhoods over cities to states. Additionally, some of the users' locations are invented. For this reason, we conduct a location assignment of all Twitter tweets forming part of our dataset. To do this, we match the users' location to administrative data on locations in the US.[9] We can assign a location to 29.9 percent of all tweets in our dataset. To assess if this is a representative share of tweets posted by users in the United States, we rely on information about the location distribution of Twitter users by countries. In 2021, 37.7 percent of Twitter users were from the United States (77.75 out of 206 million users worldwide) (Statista 2022). As our dataset consists of English tweets, the share of US users is probably slightly higher. This would mean that we cannot fully match the complete share of US tweets to a location. Moreover, many US cities have the same name. If Twitter users only indicate the city they live in, it is not possible for us to distinguish duplicated names from each other. In these cases, we assign observations to the city with the largest population.

Although we can only assign 29.9 percent of Twitter tweets a location our final dataset does not contain any missing state by week combinations.[10] Our main data limitation stems from different cities and/or counties having the same name. While using population sizes as a decision criteria reflects a probability distribution, with larger cities and/or counties having a larger probability, the true location of a user might still deviate from our assignment.

2.3.2 Sentiments within Tweets

The impact of Twitter tweets could vary largely depending on what is written. If tweets within the debate mainly disagree with the GBV-related movements, the potential effect on crime reports might be different than in a scenario, in which people agree with them. To investigate this further, we take our processed tweet text and conduct a sentiment analysis for each tweet. A sentiment analysis is a text analysis method that detects the polarity of the underlying text. For this purpose, we rely on the Valence Aware Dictionary and Sentiment Reasoner (VADER), which was explicitly trained on social media data (Hutto and Gilbert 2014).[11] The VADER relies on a pre-defined dictionary, which relates lexical features to emotion intensities. The sentiment score of the final text is the sum of the emotion intensity of each word in the text, and expresses the degree to which a text is positive, neutral, or

[9] For a detailed overview of the location generation see Annex B.2.

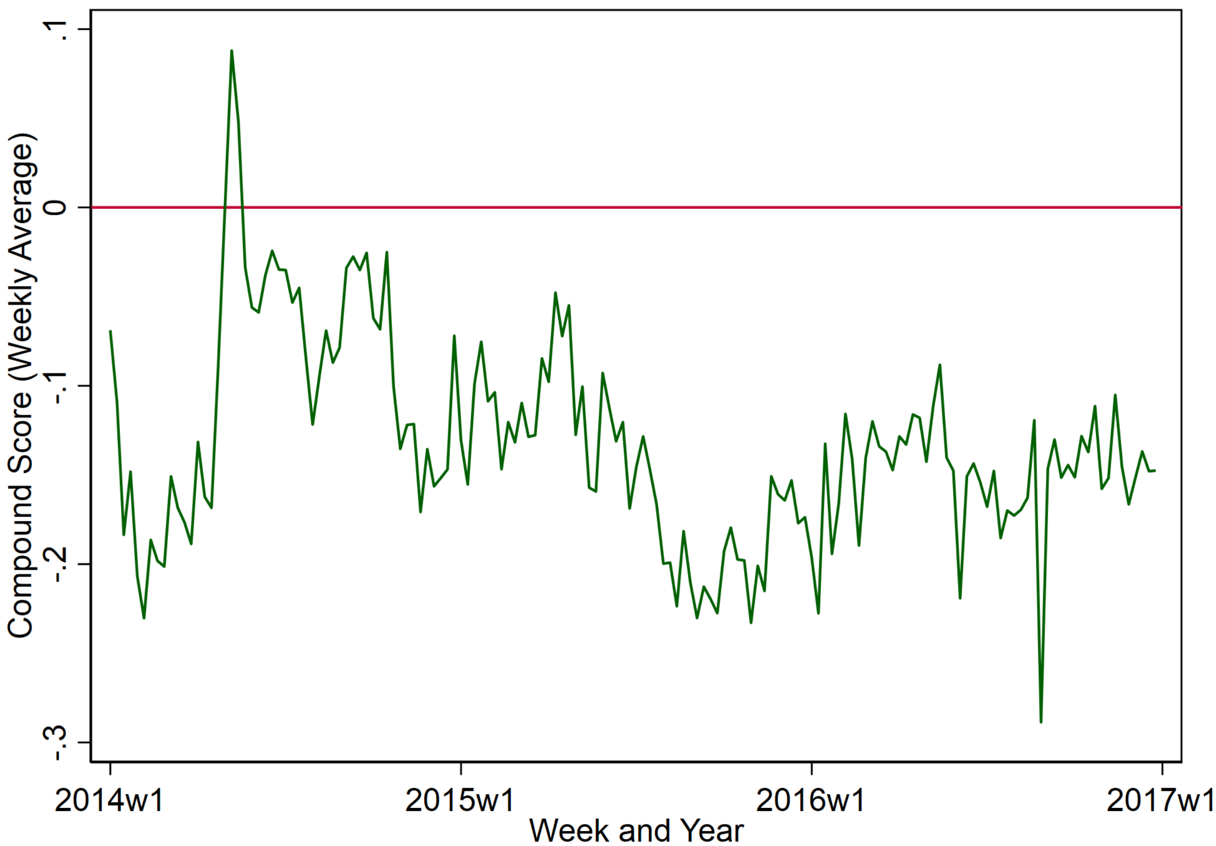
[10] There are on average 52 weeks in each year. This means that our dataset consists of roughly 156 weeks in total, as we consider three different years. When multiplying this by 50 federal states plus DC and Puerto Rico, we would theoretically end up with a dataset of 8,112 observations. We verify that there are indeed 8,112 observations in our dataset. This speaks for the quality of the data at hand.

[11] For the details on the VADER Analysis see the Annex B.3.

negative.

We first identify the sentiment score of each tweet, and then calculate the average sentiment score of each week of each year. Figure 2.2 shows the average weekly sentiment score. The graph indicates that there is fluctuation in the compound score, particularly below 0. This means that tweets had, on average, negative sentiment scores in the majority of weeks. When calculating the average sentiment score for the period 2014-2016, the score is -0.133. Consequently, negative sentiments dominate within the debate.

Figure 2.2: Weekly Sentiment Scores (2014-2016)



Notes: The figure shows weekly average sentiment scores for all tweets included in our dataset. We employ the VADER Sentiment Analysis tool to identify the sentiments within tweets' text and consider data from the period 2014-2016. The x-axis shows the respective week and the y-axis shows the compound score. The red line refers to a compound score of zero, which reflects neutrality. For the details behind the compound score see Appendix B.3. Source: Twitter (2014-2016).

2.3.3 Crime Rates and Arrest per Crime Rates on GBV

To measure crime reports we leverage data generated through the National Incident-Based Reporting System (NIBRS). The NIBRS is an incidence-based reporting system of crimes reported to the police in the US managed by the FBI. The system collects a variety of information of each incidence reported to the police, such as the nature of the offense, characteristics about its victim(s) and offender(s) and the date and place of its occurrence. The system collects information on 22 offense categories consisting of 46 specific crimes. The data is collected through reports made by city, county and state law enforcement agencies to the FBI. The submission is voluntary and takes place monthly.[12]

We make use of the processed data by the Inter-university Consortium for Political and Social Research (ICPSR) (United States Bureau of Justice Statistics 2023). The ICPSR provides the NIBRS data in four different formats, namely at the crime incident, victim, offender, and arrestee level. We take advantage of the dataset available at the crime incident for the period 2014-2016. This dataset consists of one record per incident and a total of 4,919,278 cases in 2014, 4,986,608 cases in 2015, and 5,293,536 cases in 2016. The ICPSR merges information about victim(s) and offender(s) to each incident based on the uniquely identified incidence number, resulting in a total of 390 variables.

We take advantage of crime classifications in the NIBRS to identify crimes relevant to our research question, namely sexual violence (rape, sodomy, sexual assault with an object, fondling, statutory rape), physical violence (murder/intentional manslaughter, aggravated assault, simple assault, kidnapping/abduction) and emotional violence (intimidation). In the case of physical violence, we use information provided on the circumstances of the crime and restrict the cases to those related to an argument or lovers quarrel[13]. Additionally, we restrict physical and emotional violence to cases, in which victim and offender are of opposite sexes, as we are only interested in GBV. We do not classify other crime types, in which the victim and offender are of opposite sexes as GBV. We also consider information on the victim's sex, age, race and residence status, as well as their relationship to the offender and the offender's sex, age and race.

We make use of this data by aggregating it from the crime-incidence to the weekly as well as week by federal state level to derive insights about the crime activity over time as

[12] There are 6,251 law enforcement agencies included in the data for 2014, 6,278 in the data for 2015, and 6,570 in the data for 2016. This shows that there was only a marginal increase in the number of agencies that report to the FBI over time. We are therefore confident that our results are robust to significant changes in the pattern of reporting law enforcement agencies. In total, approximately 30 percent of all law enforcement agencies in the United States report to the NIBRS. There is an estimated total of 18,000 agencies in the US (Link: https://ucr.fbi.gov/nibrs/2019).

[13] While this measure is not perfect, given that it may also include cases of violence between opposite sexes, such as an argument between neighbors, we are confident that it is a good approximation for GBV.

well as at the regional level. Figure 2.3 plots the number of reports on GBV made to the police at the weekly level in the United States. The graph illustrates that crime reports are subject to considerable seasonality. We later take this into consideration by controlling for respective time fixed effects.

The crime data suffers from several data limitations. First of all, reporting to the FBI by federal states is voluntary. As a consequence, not all federal states form part of the underlying dataset.[14] Moreover, similar to other papers making use of crime data, it is not possible for us to distinguish between reporting rates and actual underlying crime rates. Crime rates reported to the police can change due to changes in crime penetration or changes in reporting behavior. The first channel reflects behavioral adjustments by perpetrators. The second channel is based on victims changing their behavior. This is a limitation for the research question at hand, as an identification of channels is crucial for the interpretation of our findings. If social movements lead perpetrators to adapt their behavior, Twitter social movements mirror social pressure. On the contrary, if social movements change victims' behavior, this would be evidence of victims' empowerment. We address these concerns by additional analyses in Section 2.6.

We are additionally interested in the effect of Twitter social movements on behavioral changes by the authorities and therefore investigate the impact of Twitter tweets on GBV-related arrests. While the usage of arrests to measure behavior changes by the authorities is subject to limitations, this indicator has been widely used in the literature for this purpose (examples are LaFree and Drass (1996), Levitt (1998), Bullinger, Carr, and Packham (2021), and Abrams (2021)). Still, it is an empirical limitation that arrests could be driven by pure behavioral changes by the authorities, or by a change in the number of crimes committed. To account for this limitation, we look at arrest per crime rates instead of the absolute number of arrests.

To study the impact of Twitter social movements on arrests, we harness the arrestee-level extract file of the NIBRS data. The arrestee-level extract file contains one record for every arrestee recorded in NIBRS for arrest dates in the respective year, independent of the incident date. There are 3,174,660 records in 2014, 3,174,815 records in 2015, and 3,308,784 records in 2016. We then prepare our variables of interest in a similar manner to the variables from the crime-incident roster of the NIBRS. This means that we uniquely identify arrests

[14] More concretely speaking, only 39 out of 50 federal states form part of the dataset. Additionally, while the dataset covers all weeks (156), several of the week-state combinations are missing from the dataset. While a dataset consisting of 156 weeks and 39 states should result in a total number of 6,084 observations, there are only 5,907 observations in the underlying dataset. This means that 177 week-state combinations are missing. Adding the missing observations from the 11 federal states not reporting their crime data to the FBI, we end up with 1,894 missing week-state combinations.

Figure 2.3: Weekly development of the number of crime reports on GBV (2014-2016)

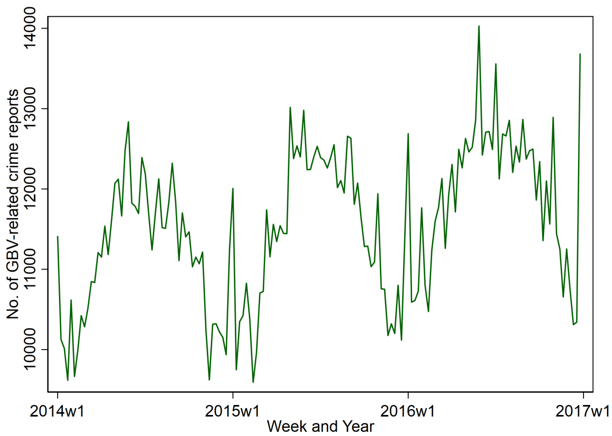
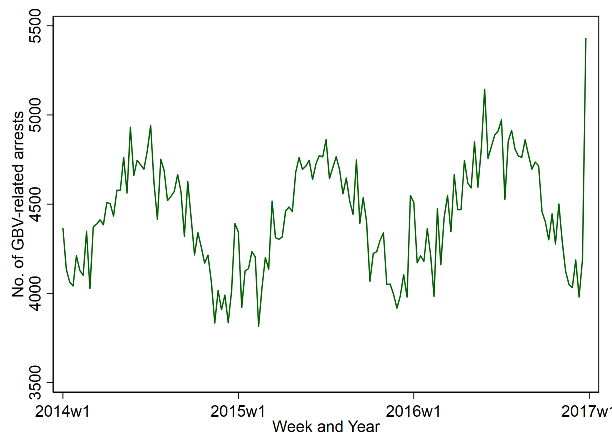


Figure 2.4: Weekly development of the number of arrests on GBV (2014-2016)



Notes: The left panel plots the number of weekly crime reports on GBV-related crimes in the United States reported in the NIBRS for the period 2014-2016. The right panel plots the weekly number of arrests of GBV-related crimes. Source: NIBRS (2014-2016).

related to GBV-related crimes, as well as the respective subcategories (physical, sexual and emotional violence).

We aggregate the number of GBV-related arrests to the week by federal state level. This allows us to combine the data on arrests with our Twitter data. Figure 2.4 shows the number of arrests on GBV made by the police on a weekly level in the United States. The graph shows that in line with the observations on crime reports there is considerable seasonality in the number of arrests on GBV made on a weekly level. We again take this into account by controlling for time fixed effects later in our regressions.

2.3.4 Additional Data Sets

We make use of three additional datasets. First, we leverage data on the population per federal state and year provided by the US Census Bureau (SimpleMaps 2012). We make use of this data by dividing the data on crime reports by population estimates to generate the crime rate at the federal state level. To the best of our knowledge, the population data is only available at the yearly level. Therefore, we divide our high-frequency data on crimes by yearly population estimates to generate the crime rate.

Next, we leverage estimates on the number of people with a cellular data plan for a smartphone or other mobile device. The American Community Survey (ACS) elevates this data at the year by federal state level (US Census Bureau 2021). We then divide the weekly number of Twitter tweets per federal state by these yearly estimates to scale our Twitter tweets by our geographic level.

To shed light on potential mechanisms behind our results, we make use of data from Google Trends (Google Trends 2021). More specifically, we gather data on the weekly search activity on the term *National Domestic Violence Hotline* for each federal state. We then merge this data at the federal state by week level to our Twitter dataset to analyze if Twitter social movements affect the search for informal support networks.

2.4 Empirical Strategy

As detailed in Section 2.3, we make use of crime data and Twitter data to study the effect of online social movements on GBV. We match both datasets at the week by state level. Table 2.1 shows the summary statistics. Each line of data represents one week during the period 2014-2016, resulting in 156 observations. The number of Twitter tweets per week varies between 7,533 and 1.1 million tweets per week, while the number of crime reports ranges from 5,292 to 13,836 reports per week. The significant variation in the number of weekly tweets is in line with observations from Figure 2.1. The figure clearly shows that, while there are very few tweets in many weeks, social movements are large and sudden.

Table 2.1: Summary statistics at the weekly level (2014-2016)

VARIABLES	Mean	Std. Dev.	Min	Max	p25	p75
GBV	11,333.03	1,294.70	5,292	13,836	10,602	12,259
Sexual violence	1,458.03	238.97	583	2,168	1,319	1,567
Physical violence	7,602.82	884.49	3,017	8,987	7,145	8,233
Emotional violence	2,272.18	294.06	751	2,681	2,156	2,455
Non-GBV crime	84,261.34	10,316.90	28,374	97,080	81,194	90,438
No. of tweets	71,000.92	127,427.41	7,533	1,132,676	37,546	58,858

Notes: The table shows the summary statistics of Twitter tweets and crime reports at the weekly level. For each crime type, the variable measuring crime is the number of crime reports in the United States at the weekly level. GBV refers to all crimes related to Gender-Based Violence (sexual, physical, and emotional crime). Sexual violence is defined as rape, sodomy, sexual assault with an object, fondling, and statutory rape. Physical violence includes murder/intentional manslaughter, aggravated assault, simple assault, kidnapping/abduction. Emotional violence is defined as intimidation. In the case of physical violence, we use information provided on the circumstances of the crime and restrict the cases to those related to an argument or lovers quarrel. Additionally, we restrict physical and emotional violence to cases, in which victim and offender are of opposite sexes, as we are only interested in GBV. The period under consideration is 2014 to 2016. Source: NIBRS and Twitter data (2014-2016).

Running regressions at the national level could result in endogeneity concerns, such as reversed causality or simultaneity bias. As an example, more GBV-related crimes could lead to people tweeting more about GBV. Moreover, the level of variation might not be sufficient to truly understand the impact of Twitter tweets on crime reports. Based on these concerns, we dis-aggregate our two datasets to the week by state level. Each line of data then represents a different week in a different federal state.[15]

To account for the fact that states which experience population growth might automatically experience an increase in the number of reported crimes, we scale crime reports by the population in each respective federal state from the US Census Bureau (see Section 2.3). As this would result in very small numbers, we report the crime rate per 100,000 inhabitants. Similarly, we scale the number of Twitter tweets by the number of smartphone internet plans. Through this we account for a potential growth in Twitter tweets driven by an increase in the number of Twitter users.[16]

Table 2.2 shows the summary statistics. The table indicates that there is significant variation in the variables investigated in this paper. The GBV-related crime rate varies between 0 and 21.833 crime reports per 100,000 people in a respective week and federal state. On average, there are 7.233 GBV-related crimes reported to the police per 100,000 people at the week by state level. The average crime rate is lowest in the case of sexual violence and highest for physical violence. In general, the GBV-related crime rate is much lower than the one for non-GBV-related crimes. There are on average 29.364 crime reports on theft and robbery per 100,000 people at the week by state level. Our main explanatory variable, the number of Twitter tweets per 100 cellphone internet subscriptions, varies between 0.003 and 18.95. The average number stands at 0.148 Twitter tweets per 100 cellphone internet subscriptions.

Figure 2.5 plots the aggregate number of GBV-related crime reports over the period 2014-2016 as a share of the population in 2014 at the federal state level. The map depicts significant variation in the aggregated GBV rate across states. The rate varies between close to zero and 0.028. While there seems to be a clear spatial agglomeration of aggregated GBV rates in the Northwest and Center of the country, there is no clear spatial pattern in the Eastern parts of the United States. This could be because political factors, such as police and law enforcement policies, differ by federal states.

[15]Due to the data limitations outlined in Section 2.3 we end up with missing observations. While the combination of approximately 156 weeks and 50 federal states should theoretically lead to 7,800 lines of code, the missing observations in the crime data lead to a dataset of only 5,751 observations. This is due to 2,361 week-state cells missing in the crime data and 156 missing week-state cells in the Twitter data.

[16]While it would be better to scale the number of Twitter tweets by the number of Twitter users, we do not dispose of this data at the state-week level. While one could theoretically generate the number of Twitter users via the API, this would exceed our monthly rate limit as well as storage space. We believe that smartphone internet plans are a good enough proxy for the number of Twitter users.

Table 2.2: Summary Statistics of crime data and Twitter tweets at the week by federal state level (2014-2016)

VARIABLES	Mean	Std. Dev.	Min	Max	p25	p75
GBV	5.961	5.508	0.000	21.833	1.289	10.038
Physical violence	4.000	3.715	0.000	15.638	1.035	6.178
Sexual violence	0.765	0.720	0.000	5.061	0.134	1.269
Emotional violence	1.197	1.418	0.000	6.984	0.100	1.847
Homicides	0.018	0.028	0.000	0.527	0.000	0.030
Violent crimes	0.963	1.026	0.000	4.907	0.148	1.373
Theft and Robbery	23.863	20.903	0.054	76.519	5.220	41.235
Twitter tweets	0.148	0.301	0.003	18.950	0.043	0.141

Notes: The table shows the summary statistics of the main variables of interest at the week by federal state level. For each crime type, the variable measuring crime is the average crime rate per 100,000 inhabitants by calendar year and federal state. GBV refers to all crimes related to Gender-Based Violence (sexual, physical, and emotional crime). We define sexual violence as rape, sodomy, sexual assault with an object, fondling, and statutory rape. We define physical violence as murder/intentional manslaughter, aggravated assault, simple assault, kidnapping/abduction. We define emotional violence as intimidation. In the case of physical violence, we use information provided on the circumstances of the crime and restrict the cases to those related to an argument or lovers quarrel. Additionally, we restrict physical and emotional violence to cases, in which victim and offender are of opposite sexes, as we are only interested in GBV. The Twitter tweets are the number of tweets per 100 cellphone internet subscriptions in a respective year and federal state. The period under consideration is 2014 to 2016. Source: NIBRS, Twitter data, and ACS (2014-2016).

The fact that the agglomerated GBV rate is lower in conservative states, such as Texas, may be unexpected. There could be two reasons for that. First, in conservative settings under-reporting of GBV might be especially high, as victims might be less empowered. Authorities might also be less likely to react to reports on GBV, which could create further disincentives to report these crimes. Similar patterns have been observed in Nigeria and Rwanda for the reporting of IPV (Cullen 2020). Next, the spatial patterns could emerge due to the FBI's data gathering process. As indicated in Section 2.3, the NIBRS relies on voluntary and monthly submissions made on crime reports by law enforcement agencies at the city, county, and state level. It could be that a lower number of law enforcement agencies in those states with lower aggregated GBV rates participate in the FBI's Uniform Crime Reporting Program. A concern would be that the number of law enforcement agencies in certain several states develops differently than in other federal states, which could potential confound our results. To show that the lack of spatial pattern is not related to the GBV categories investigated, we present a similar map for non-GBV categories in Figure B.13. The distribution of aggregated non-GBV-related crime rates is very similar to the one of aggregated GBV-related crime rates. We conclude that these spatial patterns are unlikely driven by the GBV categories.

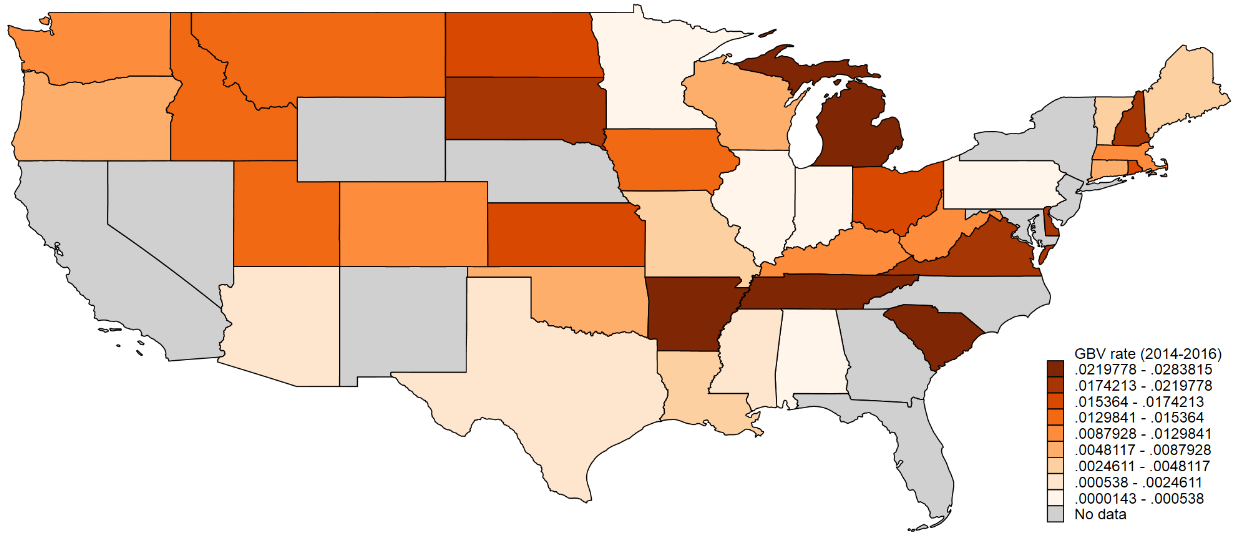
Figure 2.6 shows the aggregate number of Twitter tweets over the period 2014-2016 per 100 cellphone internet subscriptions in 2014. Contrary to Figure 2.5, Figure 2.6 displays a clear spatial pattern. Especially the Southeastern federal states as well as Arizona present low Twitter rates. Structural differences between federal states, such as a lower number of Twitter users or lower social media usage in general, might drive these results. Lower internet connectivity or a higher median age in these federal states might also account for these patterns. We consider these potential structural factors through the inclusion of federal state fixed effects in our regressions.

We run regressions at the week-state level as follows:

$$Y_{ws} = \alpha_0 + \beta_1 * T_{ws} + MY + S + \epsilon \tag{2.1}$$

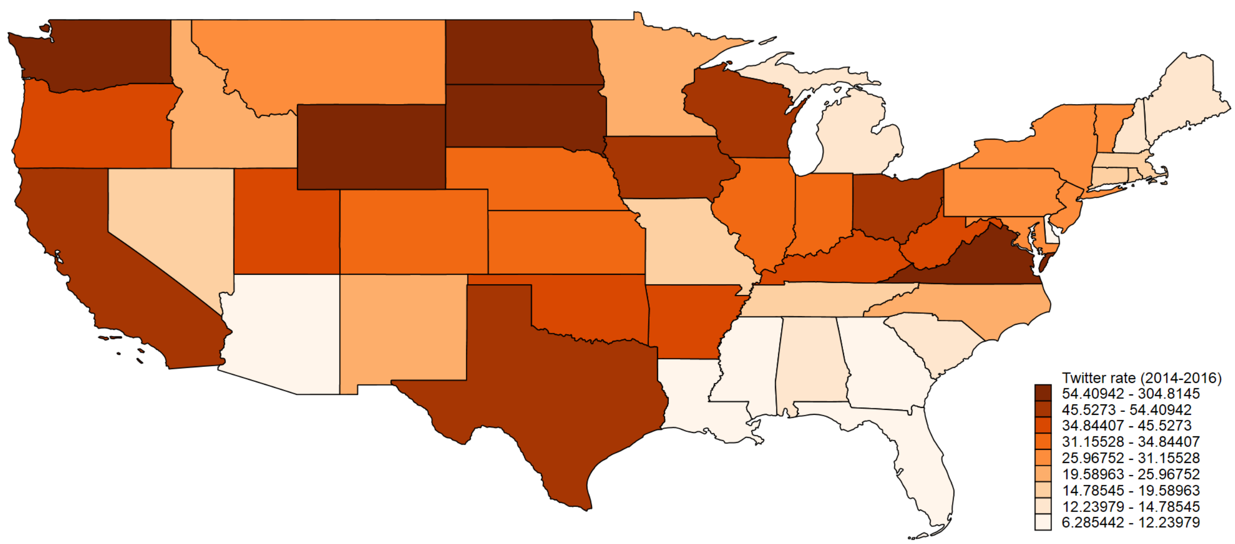
where Y_{ws} is the crime rate on GBV per 100,000 inhabitants, or one of its subcategories, at the week by federal state level. T_{ws} is the number of GBV-related tweets per 100 cellphone internet subscriptions at the week by federal state level. MY are month of the year fixed effects which control for monthly trends, such as holiday seasons, at the national level. Federal state fixed effects (S) control for state characteristics which are constant over time. Examples are population compositions, or internet connectivity. Although our fixed effects model eliminates omitted variable bias from unobservables that are constant over time at the state level, or constant across states at the monthly level, they still leave room for

Figure 2.5: Number of GBV-related crime reports over the total population by federal states (aggregate of 2014-2016)



Notes: The map depicts the aggregate number of GBV-related crimes reported to the police for the years 2014-2016 divided by population estimates from 2014 at the federal state level in the US. The graph excludes Alaska, Hawaii, and Puerto Rico. Darker colors indicate higher aggregated GBV-related crime rates. Source: NIBRS and US Census Bureau.

Figure 2.6: Number of GBV-related tweets over cellphone internet subscriptions by federal states (aggregate of 2014-2016)



Notes: The map depicts the aggregate number of Twitter tweets for the years 2014-2016 divided by the number of cellphone internet subscriptions in 2014 at the federal state level in the US. The graph excludes Alaska, Hawaii, and Puerto Rico. Darker colors indicate higher aggregated GBV-related crime rates. Source: Twitter data and US Census Bureau.

confounding factors that take place at the month by federal state level, such as economic downturns or policy instruments. We cluster standard errors at the federal state by month level to account for within-group dependencies. We weight each cell by the population size of federal states to account for the relative importance of each federal state in the United States.

People might not react immediately to Twitter social movements. They might reflect and think about what they observe online before they internalize this information and change certain behavioral patterns. To account for these potential behavioral delays, we introduce lags of Twitter tweets from the previous weeks as alternative regression specifications. We consider one to two different lags, but do not go further back than one month due to our month of the year fixed effects. The introduction of lagged coefficients can also shed light on the causal interpretation of our estimates. It is unlikely that future tweets affect current or past crime rates as the Twitter movements investigated in this paper emerged suddenly. Consequently, it is unlikely that victims of crime foresee them. We can therefore ensure that our estimates represent the effect of Twitter social movements on GBV and not vice versa.

The introduction of lagged coefficients results in the following final equation:

$$Y_{ws} = \alpha_0 + \beta_1 * T_{ws} + \beta_1 * T_{w-1s} + \beta_1 * T_{w-2s} + MY + S + \epsilon \tag{2.2}$$

T_{w-1s} represents the number of Twitter tweets in the previous week while T_{w-2s} is the number of Twitter tweets two weeks previously to the one investigated.

2.5 The Impact of Social Movements

2.5.1 The Effects on Crime Rates

The following section reports our main results. If social movements increase the social costs of GBV and deter perpetrators from committing these crimes, we would expect to see a negative effect of Twitter tweets on crimes. This would be in line with the model by Becker et al. (1995) described in Section 2.2. At the same time, reporting might increase due to victims being more empowered. If the effect on perpetration outweighs the effect on reporting we expect to see negative overall effects.

Table 2.3 shows that social movements indeed deter GBV-related crimes at the state by week level. In Column 1, we do not include the number of GBV-related Twitter tweets from previous weeks. The reported coefficient on GBV is -0.068 and significant at the 10 percent significance level. This means that social movements decrease the crime rate. One additional tweet per 100 total cellphone internet subscriptions decreases the number

Table 2.3: The effect of social movements on GBV on crime rates per 100,000 inhabitants (GBV)

	(1) GBV	(2) GBV	(3) GBV
Twitter tweets	-0.0608*	-0.0257	-0.00476
	(0.0367)	(0.0376)	(0.0384)
L.Twitter tweets		-0.0698**	-0.0378
		(0.0340)	(0.0277)
L2.Twitter tweets			-0.0842**
			(0.0425)
Constant	5.970***	5.977***	5.987***
	(0.0266)	(0.0271)	(0.0273)
Mean (Dep. Var)	5.961	5.963	5.968
St. Dv. (Dep. Var.)	5.508	5.508	5.512
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on crime rates. The outcome variable is the crime rate per 100,000 inhabitants in a respective week and federal state, considering all GBV-related crimes. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by state. The first column only considers the impact of Twitter tweets on the contemporaneous crime rate. Column 2 adds the Twitter tweets in the previous week, while Column 3 also considers the Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of reported GBV-related crimes by 0.068 per 100,000 people. Compared to the mean, this effect is approximately a 1 percent decrease.

Column 2 and 3 account for a lagging effect of Twitter tweets on GBV. These columns indicate that the crime-deteriorating impact of Twitter social movements on GBV seems to take some time. In contrast to Column 1, the coefficient in Row 1 is insignificant when introducing lags, while the lagged coefficients are significant at the 5 percent significance level. This pattern of results could mean that perpetrators take a couple of weeks before refraining from committing GBV. In terms of magnitude, the coefficients are similar to the one observed in Column 1, but increase over time. The point estimate in Column 3 indicates that the crime rate per 100,000 people decreases by 1.4 percent when compared to its mean.

Table 2.4 reveals another interesting fact about the relationship between Twitter social movements and GBV-related crime rates. The table presents the contemporaneous point coefficients of our regressions including a variety of different fixed effects. Column 1 abstracts from fixed effects, while Column 2 only considers time fixed effects. Column 3 only includes state fixed effects, while Column 4 is our main regression specification, which considers month and state fixed effects. The table clearly shows that seasonality plays a crucial role in the relationship of Twitter social movements and GBV-related crime rates. This is in line with the seasonality observed in Figure 2.3.

Table 2.4 also illustrates the impact of aggregating standard errors at different levels. Up to Column 5 we cluster standard errors at the month-state level. We then analyze the effect of varying the level of clustering. Column 4 to 6 demonstrate that standard errors increase with their level of aggregation. While coefficients are significant at the 10 percent significance level when clustering at the month-state level, they are significant at the 5 percent significance level when clustering at the quarter level. Under a specification that clusters at the quarter-state level, our coefficient is insignificant. Consequently, estimates are sensitive to the level of clustering. We choose the second most conservative specification as our baseline model.

Table 2.4: The effect of social movements on GBV on crime reporting rates per 100,000 inhabitants (GBV)

	(1) GBV	(2) GBV	(3) GBV	(4) GBV	(5) GBV	(6) GBV
Twitter tweets	0.532	0.149^{**}	0.390	-0.0610^{*}	-0.0610^{**}	-0.0610
	(0.501)	(0.0583)	(0.429)	(0.0367)	(0.0145)	(0.0399)
Constant	5.883^{***}	5.939^{***}	5.904^{***}	5.970^{***}	5.970^{***}	5.970^{***}
	(0.363)	(0.0322)	(0.363)	(0.0266)	(0.00214)	(0.0404)
Mean (Dep. Var)	5.961	5.961	5.961	5.961	5.961	5.961
St. Dv. (Dep. Var.)	5.508	5.508	5.508	5.508	5.508	5.508
State fixed effects	No	Yes	No	Yes	Yes	Yes
Month fixed effects	No	No	Yes	Yes	Yes	Yes
Clustered standard errors	Month-State	Month-State	Month-State	Month-State	Quarter	Quarter-State
N	5751	5751	5751	5751	5751	5751

Notes: The table shows the results from a linear regression of the number of Twitter tweets on crime rates under different empirical specifications. The outcome variable is the crime rate per 100,000 inhabitants in a respective week and federal state, considering all GBV-related crimes. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by state. Column 1 abstracts from fixed effects. Column 2 includes state fixed effects and Column 3 month fixed effects. Column 4 controls for both state and month fixed effects. Clustered standard errors are at the month-state level in Column 1 to 4, at the quarter level in Column 5 and at the quarter-state level in Column 6. Clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5.2 The Effect on Arrests per Crimes

To study whether social movements on GBV impact the behavior of the police force, we analyze the impact of GBV-related Twitter tweets on the arrest per crime rate related to GBV. If the social pressure generated by these movements trickles down to the authorities, we would expect a positive and significant impact of Twitter tweets on arrest per crime rates. This is a relevant question, as it has important policy implications for law enforcement.

Table 2.5 shows that there is a significant impact of Twitter social movements on arrest per crime rates two weeks later. The coefficient of the second lag in Column 3 is positive and significant at the 10 percent significance level. In terms of magnitude, the effect is equivalent to a 2.1 percent increase when compared to the mean.

In general, the evidence speaking in favor of significant effects on arrest per crime rates is more limited than the one showing a significant impact on crime rates.

Table 2.5: The effect of social movements on GBV on arrests per crime (GBV)

	(1) Arrests	(2) Arrests	(3) Arrests
Twitter tweets	0.00190	0.00233	0.000266
	(0.00532)	(0.00439)	(0.00487)
L.Twitter tweets		-0.00193	-0.00514
		(0.00360)	(0.00392)
L2.Twitter tweets			0.00796^{*}
			(0.00476)
Constant	0.388^{***}	0.388^{***}	0.388^{***}
	(0.00155)	(0.00167)	(0.00167)
Mean (Dep. Var)	0.388	0.388	0.388
St. Dv. (Dep. Var.)	0.134	0.134	0.134
State-Month fixed-effects	Yes	Yes	Yes
N	5748	5708	5668

Notes: The table shows the results from a linear regression of the number of tweets on the arrest per crime rate. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds the Twitter tweets in the previous week, while Column 3 also considers the Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5.3 Interpretation and Caveats

We find that Twitter social movements lead to a significant decrease in the crime rate of GBV-related crimes. From this, we conclude that online conversations on Twitter have important implications for offline behavior. One possible interpretation is that Twitter facilitates the signaling of shifting social norms. This would be in line with previous studies demonstrating a significant association between social norms and GBV (Linos et al. (2013); Yilmaz (2018)), as well as work by Bursztyn, Egorov, and Fiorin (2020) demonstrating significant effects of Donald Trump's rise in polarity on publicly express xenophobic views. Our findings could mean that perpetrators are increasingly aware of the social pressure resulting from online movements and fear social punishment. They might also become more aware of punishments experienced by other perpetrators due to the increased visibility of these cases on social media platforms. Our estimates can be understood as a lower bound effect, if reporting rates increase in parallel.

A change in perpetrators' behavior as a result of increased social pressure would be in line with the standard economic perspective on crime. From an economic viewpoint, criminal activity varies with the price of conducting a crime. This price increases with peer pressure and neighborhood effects (Falk and Fischbacher 2002). Previous research studying the role of social pressure in the prevention of GBV-related crimes confirm the role of potential social costs. Standish (2014), for example, find that social pressure plays an important role in the prevention of dowry murder. Likewise, a literature review on the role of gender norms in GBV perpetration in forced displacement settings finds that social pressure plays a significant role in GBV perpetration by male youth (Fry, Skinner, and Wheeler 2019). In a slightly different setting, Balestrino (2008) explains that people who normally refrained from committing illegalities became digital pirates as there was no social stigma and, consequently, no social costs attached to it.

Alternatively, one might argue that our results reflect a decrease in the reporting behavior of victims of GBV. This could be due to them having increasing access to informal support networks on Twitter, or potential backlash behavior by perpetrators. Both would result in a decrease in the crime rate. We provide evidence in Section 2.6 which proves that it is unlikely that these channels dominate our findings.

The positive coefficient on arrest per crime rates in Table 2.5 could mean that the social pressure generated via Twitter social movements on GBV does trickle down to the authorities and leads to an increase in the arrests made relative to the crimes reported. Consequently, law enforcement might increase. Nevertheless, the evidence showing significant effects is more limited than in the case of crime rates.

Our results point towards the absence of a potential backlash by the police as a response to Twitter social movements. Backlashes are a concern as less than 13 percent of full-time police officers in the United States are women. Consequently, the police force in the United States consists mostly of male and might consciously or subconsciously feel threatened by social movements about GBV. Backlashes have been observed with respect to gay police officers disclosing their homosexuality (Rumens and Broomfield 2012). In some cases these disclosures have resulted in the reinforcement of traditional notions of masculinity in some police work environments. Similarly, mandatory and preferred arrest policies related to domestic violence cases have resulted in a backlash for victims who were arrested along with their batterers (Finn and Bettis 2006). Police officers explained this behavior by their desire to force the victim to get counseling for their relationship. In a similar fashion, Amaral, Bhalotra, and Prakash (2021) stress that the introduction of women police stations in India might be efficient when aiming at an increased reporting of GBV to the authorities, but might be hampered by backlashes from male police officers. Our results could mean that the social pressure generated by Twitter social movements dominates the potential emergence of backlash movements.

Our findings are interesting from a policy perspective due to several reasons. To start with, entities interested in decreasing GBV can explore the potential of social media platforms to achieve this. They can also explore alternative ways to make use of social pressure and the perception of social costs around these types of crime. Still, they should also be aware of potential backlashes and find ways to protect victims against the impact of those. Moreover, they can assess the potential of social media platforms to develop informal as well as formal support networks of those who have fallen victim of GBV. Through this, they can strengthen the public support network related to GBV and provide victims with spaces where they can talk openly and fearlessly about their experiences, as well as provide them with guidance on how to address GBV.

Our findings should be taken with caution as they might be subject to empirical limitations. First, there might be a simultaneity bias. GBV-related Twitter social movements might be triggered by an increase in GBV-related crimes committed or arrest per crime rates in a respective federal state. Given that our results persist when introducing lagged coefficients, we believe that this is unlikely. We confirm the causality of our findings by an event study design in Section 2.7. Next, our results might be confounded by an omitted variable bias that affects crime rates and Twitter social movements. To investigate this possibility, we conduct placebo regressions in Section 2.7. Lastly, our estimates might be subject to reporting bias. If victims of GBV feel empowered by Twitter social movements, reporting of GBV might increase. Our coefficients might then reflect lower bound estimates of the true underlying effect on crime.

2.6 Drivers behind GBV-related Social Movements

2.6.1 Reporting versus Committing Crimes

There could be several drivers behind the observed decrease in crime rates as a response to Twitter social movements. Based on the economic theory by Becker et al. (1995) our findings would indicate that perpetrators refrain from committing GBV as a response to increased social pressure and costs. Still, there could be alternative explanations for a decreasing crime rate.

Another possibility could be that instead of perpetrators it is the victims who change their behavior. While this seems counter-intuitive, it could well be that people who have fallen subject to GBV found informal support networks online, or alternative ways to express their outrage or pain related to these experiences thanks to online platforms, such as Twitter. They might therefore feel less urged to report their experiences to the authorities. McCart, Smith, and Sawyer (2010) find that only a small fraction of crime victims seek help from formal support networks while many seek help from informal sources. This pattern could have increased through GBV-related Twitter tweets and an easier accessibility to these informal networks as many people identify themselves as victims of GBV.

To investigate this further, we analyze the impact of Twitter social movements on Google search activities on the term "*National domestic violence hotline*". If an increase in the access to informal support networks drives the observed decrease in crime rates, we would expect a positive impact of the number of Twitter tweets on the Google search activity of this term.

Table 2.6 shows that there is no clear pattern on the interaction between Twitter tweets and Google search activities for informal support. While the lagged coefficients presented in Column 3 are significant at the 10 percent significance level, they go into opposite directions. This finding implies that an increase in the number of Twitter tweets per 100 cellphone internet subscriptions first leads to an increase in the search for informal support. One week later, it then results in a decrease. One possible interpretation is that there is no clear impact of Twitter social movements on seeking informal support networks. It is unlikely that this channel is the dominant driver behind our results.

Alternatively, crime reporting by victims could decrease due to them experiencing backlashes. A backlash is a sudden and violent backward movement. The political economic literature has identified backlashes as a response to female political empowerment (example studies are by Gangadharan et al. (2019), Gagliarducci and Paserman (2012)), and several have noted them in response to female economic empowerment (as for example work by Bobonis, González-Brenes, and Castro (2013), Erten and Keskin (2018), Guarnieri and Rainer (2021), Bhalotra et al. (2021b)). Backlashes by male partners might increase as a re-

Table 2.6: The effect of social movements on Google Searches on informal support networks

	(1) Google Trends	(2) Google Trends	(3) Google Trends
Twitter tweets	0.482	-0.245	0.309
	(2.199)	(2.275)	(2.208)
L.Twitter tweets		1.434	2.410*
		(1.563)	(1.364)
L2.Twitter tweets			-2.270*
			(1.294)
Constant	8.931***	8.839***	8.952***
	(0.433)	(0.457)	(0.472)
Mean (Dep. Var)	8.996	9.000	9.012
St. Dv. (Dep. Var.)	19.14	19.17	19.20
State-Month fixed-effects	Yes	Yes	Yes
N	4212	4185	4158

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the google search activity of the term "*National domestic violence hotline*" on Google. The explanatory variable is the number of GBV-related Twitter tweets in a respective week by federal state, per 100 cellphone internet plan subscriptions in a respective year and federal state. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous crime rate. Column 2 adds the Twitter tweets in the previous week, while Column 3 also considers the Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: Google Trends and Twitter. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

sponse to GBV-related social movements on Twitter to protect the status quo. Perpetrators might feel threatened by social movements happening online and, consequently, intimidate their victims even more.

We validate this possibility by testing the impact of a Twitter social movement identified as a backlash movement. If our results are driven by victims reporting GBV-related crimes less to the police based on backlashes, we would expect to see an even more negative effect of Twitter backlash movements on GBV-related crime rates. To investigate this possible channel, we retrieve all tweets using the hashtag *#alphamale* from the Twitter API. We believe that this hashtag embraces traditional gender norms and a traditional understanding of masculinity. We then estimate our regressions using the number of tweets with the hashtag *#alphamale* per 100 cellphone internet subscriptions as our main explanatory variable.

The evidence presented in Table 2.7 may confirm our hypothesis that behavioral changes by perpetrators drive our main findings. The number of Twitter tweets belonging to Twitter backlash movements has no impact on crime rates. While the coefficient in Column 1 is significant at the 10 percent significance level, it becomes insignificant when accounting for lagged coefficients in Column 2 and 3. If at all, the coefficient in Column 1 would be indicative of an increase in GBV, as it is unlikely that reporting would increase as a response to such a movement. Consequently, the results presented in Table 2.7 confirm that our main findings are based on changes in perpetrators' behavior.

To shed further light on the question to which extent reporting behavior plays a role in our overall estimator, we investigate the impact of Twitter tweets on violent crimes, namely homicides and aggravated assault. The underlying idea is that homicides cannot be driven by a change in reporting behavior. In addition, aggravated assaults often involve a relationship worthy of protection, such as a caregiver and a mentally ill person. The victims worthy of protection might also be more unlikely to report crimes by themselves. Therefore, the crime rate of violent crimes might be less subject to reporting bias. Consequently, if there is a significant impact of Twitter social movements on the violent crime rate, it is likely driven by perpetrators changing their behavior and not victims' reporting behavior.

Table 2.8 shows that there is evidence in favor of Twitter tweets decreasing the crime rate of violent crimes. The coefficient on the lagged number of Twitter tweets in Column 2 is significant at the 5 percent significance level. Moreover, when abstracting from lagged coefficients, Column 1 reports a significant estimator at the 1 percent significance level. Importantly, the coefficients indicate that the number of Twitter tweets decreases the violent crime rate. An additional tweet per 100 cellphone internet subscription leads to a decrease of 0.02 to 0.03 violent crimes per 100,000 people. This evidence points towards perpetrators' behavior driving our results.

Table 2.7: The effect of tweets using the hashtag *#alphamale* on crime rates per 100,000 inhabitants (GBV)

	(1) GBV	(2) GBV	(3) GBV
Twitter tweets	19.61*	16.68	12.27
	(10.59)	(14.81)	(15.67)
L.Twitter tweets		3.967	-7.770
		(14.41)	(14.15)
L2.Twitter tweets			19.36
			(12.69)
Constant	5.948***	5.948***	5.951***
	(0.0264)	(0.0264)	(0.0263)
Mean (Dep. Var)	5.961	5.963	5.968
St. Dv. (Dep. Var.)	5.508	5.508	5.512
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all GBV-related crimes. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related Twitter tweets using the hashtag *#alphamale* in a respective week and federal state, per 100 cellphone internet plan subscriptions in a respective year and federal state. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In summary, the evidence provided in this section bolsters the case for changes in perpetrators' behavior driving our results. While victims' reporting behavior might also change as a response to Twitter social movements, this channel unlikely dominates the overall estimator.

Table 2.8: The effect of social movements on GBV on crime rates per 100,000 inhabitants (Violent crimes)

	(1) Violent crimes	(2) Violent crimes	(3) Violent crimes
Twitter tweets	-0.0258***	-0.0105	-0.00652
	(0.00859)	(0.00795)	(0.00889)
L.Twitter tweets		-0.0311**	-0.0249**
		(0.0133)	(0.0118)
L2.Twitter tweets			-0.0158
			(0.0129)
Constant	0.967***	0.970***	0.973***
	(0.00639)	(0.00666)	(0.00670)
Mean (Dep. Var)	0.963	0.964	0.966
St. Dv. (Dep. Var.)	1.026	1.026	1.028
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all violent crimes. We define violent crimes as homicides and aggravated assault, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related Twitter tweets in a respective week and federal state, per 100 cellphone internet plan subscriptions in a respective year and federal state. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.6.2 The Role of Stigma and Tabooing

To better understand what drives our results we analyze to which extent social stigmatization and tabooing play a role in our findings. This is an important question, as it can provide policymakers with further guidance on how to best tackle the underlying drivers of GBV. To this end, we conduct a more granular analysis distinguishing between sexual, physical and emotional GBV.[17] We divide by these sub-forms of GBV, as the degree of stigmatization might differ by type of GBV.[18]

Our hypothesis would be that stigmatization and tabooing is largest for sexual violence, followed by emotional violence, and lastly physical violence.[19] If social stigmatization and tabooing are important drivers in GBV, we would then expect to see a lower effect of social movements on sexual GBV. A larger impact on sexual GBV, on the other hand, could be due to many of the Twitter social movements investigated in this paper focusing on sexual violence. Additionally, previous reporting rates of sexual violence might have been especially low. We start by exploring these channels for crime rates. In a second step, we investigate these possible drivers for arrest per crime rates.

Crime Rates and the Role of Social Stigma

Table 2.9 presents our findings on the impact of Twitter tweets on sexual violence. The table shows a clear negative impact of Twitter social movements on the crime rate in the same week, as well as one and two weeks after the Twitter social movements took place. These findings are stable across model specifications and point towards perpetrators committing these crimes to a lesser extent. The coefficient in Row 1 and Column 3 is -0.0242. In terms of magnitude, an increase in the number of Twitter tweets per 100 cellphone internet subscriptions decreases sexual violence per 100,000 people by approximately 3.158 percent when compared to the mean value. Two weeks later, the crime-deteriorating impact increases to 5.979 percent when compared to the average.

Interestingly, the coefficient on the first lag reported in Row 2 is significant and positive. One possible interpretation is that social stigmatization and tabooing is likely higher around sexual violence than other forms of GBV. Hence, the reporting effect triggered by Twitter social movements could be especially large in this case and might dominate the crime deteriorating effect. In addition, many of the Twitter social movements investigated in this paper focused on sexual violence. Our results could point towards online social movements potentially counteracting stigma and tabooing around sexual violence. Thus, they might empower victims to report these types of crimes more often. This line of thought would bol-

[17] While one might argue that there is no clear definition of emotional violence, we base our definition on data gathered by the FBI. This means that emotional violence is intimidation between opposite sexes. While this might not represent the full universe of emotional violence, we believe that it is a close enough approximation to capture its occurrence.

[18] To date, there is only limited evidence of the degree of stigma by type of GBV. Work by Harris (2017) demonstrates that the type of violence does not alter the relationship between stigma and reporting GBV in the case of homosexual men.

[19] Scholars from other fields have shown that social stigma around sexual violence is especially high (see for example Delker et al. (2020)).

ster a potential interpretation of Twitter social movements serving as a signaling mechanism for shifting social norms.

Table 2.9: The effect of social movements on GBV on crime rates per 100,000 inhabitants (Sexual violence)

	(1) Sexual violence	(2) Sexual violence	(3) Sexual violence
Twitter tweets	-0.0258***	-0.0356***	-0.0242**
	(0.00810)	(0.0116)	(0.0104)
L.Twitter tweets		0.0231*	0.0407***
		(0.0119)	(0.0138)
L2.Twitter tweets			-0.0456***
			(0.0123)
Constant	0.768***	0.765***	0.768***
	(0.00442)	(0.00425)	(0.00440)
Mean (Dep. Var)	0.765	0.763	0.764
St. Dv. (Dep. Var.)	0.720	0.717	0.717
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all crimes related to sexual violence. We define crimes related to sexual violence as rape, sodomy, sexual assault with an object, fondling, statutory rape, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In the case of physical violence, Twitter social movements significantly decrease the crime rate one week later. The coefficients reported in Row 2 of Table 2.10 are significant at the 1 percent significance level for all model specifications. Therefore, while Twitter tweets do not affect crime rates in the same week, they decrease the number of physical GBV-related crimes in the following week. The coefficient in Row 2 is close to -0.07, meaning that one more tweet per 100 cellphone internet subscriptions leads to a decrease of 0.07 physical crimes per 100,000 people. In terms of magnitude, the point estimator mirrors a decrease of 1.705 percent when compared to the average value of physical violence per 100,000 people. The effect then seems to fade out in the following weeks.

Table 2.10: The effect of social movements on GBV on crime rates per 100,000 inhabitants (Physical violence)

	(1) Physical violence	(2) Physical violence	(3) Physical violence
Twitter tweets	-0.0132	0.0215	0.0236
	(0.0269)	(0.0302)	(0.0312)
L.Twitter tweets		-0.0714***	-0.0683***
		(0.0271)	(0.0259)
L2.Twitter tweets			-0.00897
			(0.0279)
Constant	4.002***	4.009***	4.013***
	(0.0178)	(0.0182)	(0.0181)
Mean (Dep. Var)	4.000	4.002	4.005
St. Dv. (Dep. Var.)	3.715	3.716	3.718
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all crimes related to physical violence. We define physical violence as crimes related to murder/intentional manslaughter, aggravated assault, simple assault, kidnapping/abduction, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are repoted in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Lastly, Table 2.11 demonstrates a lagged effect of Twitter tweets on crime rates in the case of emotional GBV. The point estimates presented in Table 2.11 fluctuate between -0.02 and -0.03 and are significant at the 10 and 5 percent significance level respectively. A one unit increase in the number of tweets per 100 cellphone internet subscriptions decreases the crime rate per 100,000 people by 0.02 in the case of emotional violence. These estimates present a 2.469 percent increase for the second lag in Column 3. In conclusion, similarly to our results on sexual and physical violence, social movements have a significant effect on emotional GBV.

Table 2.11: The effect of social movements on GBV on crime rates per 100,000 inhabitants (Emotional violence)

	(1) Emotional violence	(2) Emotional violence	(3) Emotional violence
Twitter tweets	-0.0218*	-0.0116	-0.00416
	(0.0112)	(0.0110)	(0.0113)
L.Twitter tweets		-0.0216**	-0.0102
		(0.00985)	(0.0101)
L2.Twitter tweets			-0.0296**
			(0.0125)
Constant	1.200***	1.203***	1.206***
	(0.00791)	(0.00798)	(0.00801)
Mean (Dep. Var)	1.197	1.198	1.199
St. Dv. (Dep. Var.)	1.418	1.419	1.420
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all crimes related to emotional violence. We define emotional violence as intimidation, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In summary, the overall impact on GBV reported previously persists for all three forms of GBV-related crimes investigated in this paper. The effect is largest in the case of sexual violence. In addition to decreasing the prevalence of these crimes, Twitter social movements seem to be especially effective in altering the reporting rate of sexual violence. These results speak for stigmas being more persistent in the case of sexual violence, and social movements addressing these stigmas openly. This insight aligns with the fact that many of the social movements investigated in this paper had a strong focus on sexual violence. Overall, our findings show that stigmatization and tabooing play an important role when analyzing the effect of GBV-related social movements on GBV-related crime rates.

Consequently, policymakers interested in decreasing the prevalence of GBV should address stigmatization, tabooing and silencing surrounding it. It is recommended to design interventions addressing harmful gender norms.

Arrest per Crime Rates and the Role of Social Stigma

We next analyze the impact of Twitter social movements on arrest per crime rates by subtype of GBV. This can shed light on to which extent social stigma and tabooing play a role in law enforcement. If arrests made by the police are affected by these factors and Twitter social movements impact them, we would expect a varying effect of Twitter social movements on arrest per crime rates.

We do not find compelling evidence that Twitter social movements alter the arrest per crime rate in the case of physical nor emotional violence. The coefficients reported on physical violence in Tables 2.12 and on emotional violence in Table 2.13 are mostly insignificant. While one of the lagged coefficients in Row 2 of Table 2.12 and Table 2.13 is significant at the 10 percent significance level, this finding is not stable across model specifications. Based on this evidence, we conclude that significant effects of social movements on arrest per crime rates are less likely in the case of physical as well as emotional violence.

When looking at the results on arrest per crime rates on sexual violence presented in Table 2.14 a slightly different picture emerges. The coefficient on the lagged impact of Twitter tweets reported in Column 3 is significant at the 10 percent significance level. In addition, the estimator reported on the second lag is significant at the 5 percent significance level. Interestingly, the coefficient on the first lag in Row 2 is negative, while the one on the second lag in Row 3 is positive. This means that there is first a decrease in the arrest per crime rate on sexual violence in response to Twitter social movements. This could be driven by the authorities making less arrests due to backlash behavior or less sexual violence taking place. They then, as a response to what is observed in Table 2.9 - namely an increase in reporting behavior - increasingly arrest perpetrators committing sexual violence.

The diverging results on arrest per crime rates with respect to sexual, physical, and emotional violence indicate that tabooing and social stigma play a significant role in the behavior of the police. Moreover, the fact that many of the Twitter social movements investigated in this paper had a special focus on sexual violence could explain why we find

Table 2.12: The effect of social movements on GBV on arrests per crime (Physical violence)

	(1) Arrests	(2) Arrests	(3) Arrests
Twitter tweets	0.000269	-0.00526	-0.00832
	(0.00644)	(0.00668)	(0.00807)
L.Twitter tweets		0.00998*	0.00520
		(0.00562)	(0.00346)
L2.Twitter tweets			0.0118
			(0.0115)
Constant	0.473***	0.472***	0.471***
	(0.00224)	(0.00227)	(0.00238)
Mean (Dep. Var)	0.473	0.473	0.473
St. Dv. (Dep. Var.)	0.161	0.160	0.160
State-Month fixed-effects	Yes	Yes	Yes
N	5732	5692	5652

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the arrest per crime rate. The outcome variable is the respective arrest per crime rate, considering all crimes related to physical violence. We define physical violence as crimes related to murder/intentional manslaughter, aggravated assault, simple assault, kidnapping/abduction, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.13: The effect of social movements on GBV on arrests per crime (Emotional violence)

	(1) Arrests	(2) Arrests	(3) Arrests
No. of Twitter tweets	-0.00463	-0.000605	-0.00219
	(0.00767)	(0.00641)	(0.00680)
L.No. of Twitter tweets		-0.00830	-0.0105*
		(0.00622)	(0.00635)
L2.No. of Twitter tweets			0.00557
			(0.00848)
Constant	0.215***	0.216***	0.216***
	(0.00328)	(0.00346)	(0.00355)
Mean (Dep. Var)	0.214	0.215	0.215
St. Dv. (Dep. Var.)	0.204	0.204	0.205
State-Month fixed-effects	Yes	Yes	Yes
N	5480	5445	5408

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the arrest per crime rate. The outcome variable is the respective arrest per crime rate per week and federal state, considering all crimes related to emotional violence. We define emotional violence as intimidation, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.14: The effect of social movements on GBV on arrests per crime (Sexual violence)

	(1) Arrests	(2) Arrests	(3) Arrests
Twitter tweets	0.00899	0.0128	0.00788
	(0.00608)	(0.00837)	(0.00765)
L.Twitter tweets		-0.00814	-0.0156*
		(0.00717)	(0.00868)
L2.Twitter tweets			0.0188**
			(0.00751)
Constant	0.187***	0.188***	0.187***
	(0.00309)	(0.00302)	(0.00306)
Mean (Dep. Var)	0.188	0.189	0.189
St. Dv. (Dep. Var.)	0.183	0.183	0.183
State-Month fixed-effects	Yes	Yes	Yes
N	5431	5394	5360

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the arrest per crime rate. The outcome variable is the respective arrest per crime rate per 100,000 inhabitants per week and federal state, considering all crimes related to sexual violence. We define crimes related to sexual violence as rape, sodomy, sexual assault with an object, fondling, statutory rape, in which the perpetrator and victim are of opposite gender. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

a significant impact for sexual violence but neither physical nor emotional violence. Our findings have important implications for the training of the police force. When authorities are trained, it is crucial to raise awareness about social stigmas and tabooing surrounding GBV. Policymakers should enforce environments that are free of harmful gender norms and do not permit backlash cultures.

2.6.3 Analyzing the Text of Tweets

We next explore to which extent the content of tweets' text plays a role in our findings. The impact of tweets in favor of GBV-related movements might differ from tweets opposing them. In addition, the polarity of written text might also play a role. More extreme tweets, for example, might have a more significant effect than less extreme tweets. This is a relevant research question as it has important policy implications. If the content of tweets within social movements matters, one might worry about potential backlash behavior emerging as part of social movements in favor of gender equality.

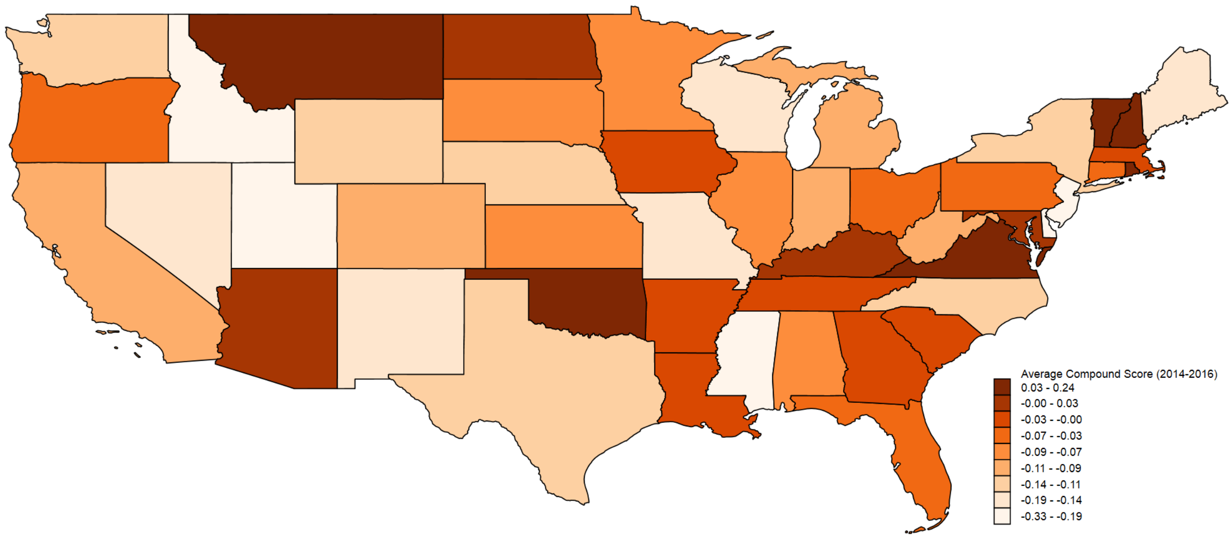
To investigate this in more detail, we analyze how the average polarity of what is written impacts crime rates as well as arrest per crime rates. We measure the average polarity of the tweets' content by a sentiment analysis. This means that we apply text analysis methods and deduce the average compound sentiment score of all tweets in a respective week and federal state.[20] Figure 2.7 shows that there is significant variation in the average compound score for the period 2014-2016 at the federal state level. We then estimate our main regression specification using the average compound score in a respective week and federal state as our main explanatory variable and the crime rate as well as arrest per crime rate as our outcome variables. If the polarity of Twitter tweets matter, we expect to see a significant coefficient.

The results presented in Table 2.15 indicate that the polarity of tweets do not play a significant role in their impact on crime rates. All coefficients reported in the table are insignificant. This is the case for all sub-types of GBV investigated in this paper (see Appendix B.5.1 for the detailed results). This means that the pure magnitude of Twitter social movements matters more than their content. Similarly, Table 2.16 demonstrates that the polarity of what is written on Twitter does not play a role in altering the arrest per crime rate. All coefficients presented in the table are insignificant.

Consequently, backlash behavior within social movements in favor of gender equality are less of a concern. The mere existence and magnitude of these movements seem to be more important. However, our findings solely focus on the content of social movements on Twitter that are in favor of gender equality. Therefore, our analysis abstracts from the potential

[20]For a detailed description see Section B.3.

Figure 2.7: Average Compound Score for the period 2014-2016 at the federal state level



Notes: The map depicts the average compound score at the federal state level for the period 2014-2016 in the US. We derive sentiment scores by applying the VADER Sentiment Analysis Tool (see Appendix B.3 for details). The graph excludes Alaska, Hawaii, and Puerto Rico. Darker colors indicate higher compound scores. Source: Twitter data and US Census Bureau.

emergence of social movements against gender equality as a response to social movements in favor of gender equality.

Table 2.15: The effect of the polarity of GBV-related tweets on crime rates per 100,000 inhabitants (GBV)

	(1) Crime rate	(2) Crime rate	(3) Crime rate
Compound Score	-0.000648	-0.0214	-0.0210
	(0.000976)	(0.0312)	(0.0357)
L.Compound Score		0.0210	0.0327
		(0.0315)	(0.0271)
L2.Compound Score			-0.0121
			(0.0344)
Constant	5.962***	5.963***	5.968***
	(0.0258)	(0.0259)	(0.0258)
Mean (Dep. Var)	5.961	5.963	5.968
St. Dv. (Dep. Var.)	5.508	5.508	5.512
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the average polarity of Twitter tweets on the crime rate. We deduce the average polarity of Twitter tweets by employing a VADER Sentiment Analysis. This text analysis method is a social media sentiment analysis method and approximates the average polarity of social media text by a compound score. The compound score is a score with values ranging from -1 to 1. A value of -1 represents text in complete disagreement while a value of 1 represents text in complete agreement. For methodological details of the VADER Sentiment Analysis see Appendix B.3. The outcome variable is the respective crime rate per 100,000 inhabitants per week and federal state, considering all GBV-related crimes. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of opposite gender. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.16: The effect of the polarity of GBV-related tweets on GBV on arrest per crime rates (GBV)

	(1) Arrests	(2) Arrests	(3) Arrests
Compound Score	0.000314	-0.00389	-0.00198
	(0.000500)	(0.0118)	(0.0137)
L.Compound Score		0.00423	0.0129
		(0.0120)	(0.0111)
L2.Compound Score			-0.0107
			(0.0131)
Constant	2.288***	2.289***	2.291***
	(0.00853)	(0.00859)	(0.00863)
Mean (Dep. Var)	2.289	2.289	2.291
St. Dv. (Dep. Var.)	2.025	2.025	2.025
State-Month fixed-effects	Yes	Yes	Yes
N	5750	5710	5670

Notes: The table shows the results from a linear regression of the average polarity of Twitter tweets on the arrest per crime rate. We deduce the average polarity of Twitter tweets by employing a VADER Sentiment Analysis. This text analysis method is a social media sentiment analysis method and approximates the average polarity of social media text by a compound score. The compound score is a score with values ranging from -1 to 1. A value of -1 represents text in complete disagreement while a value of 1 represents text in complete agreement. For methodological details of the VADER Sentiment Analysis see Appendix B.3. The outcome variable is the arrest per crime rate in a respective week by federal state, considering all GBV-related crimes. We define GBV-related crimes as physical, sexual, and emotional crimes, in which the perpetrator and victim are of opposite gender. The unit of analysis is the week federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.7 Robustness Checks

We next validate our findings by conducting several robustness tests. We start by analyzing the existence of potential unobservable confounding factors through placebo regressions. Next, we investigate if Twitter users are systematically different from victims reporting GBV to the police. Furthermore, we conduct an event study to validate the causality of our results. Lastly, we allow for spatial spillovers between federal states and conduct spatial regressions.

2.7.1 Placebo Regressions

Our main findings could be subject to potential endogeneity concerns, such as omitted variable biases or reversed causality. We therefore estimate placebo regressions, in which our main outcome variable is the number of crimes per week-state not related to GBV over 100,000 inhabitants: theft and robberies. If our results are driven by unobserved factors which drive the number of all committed crimes we would expect to find significant effects of Twitter tweets on thefts and robberies.

Table 2.17 shows that GBV-related Twitter social movements do not impact the theft and robbery crime rate significantly. All point estimator reported in the table are insignificant. We are therefore confident that our empirical strategy is robust to potential confounding factors that drive both Twitter social movements and crime rates.

Similarly to our robustness tests on crime reports, we estimate placebo regressions using non-GBV-related arrest per crime rates at the week by state level as our main outcome variable. We consider the same placebo outcomes, namely theft and robberies. The evidence on potential unobservable confounding factors driving our results is weak. Only the estimator on the second lag reported in Row 3 of Table 2.18 is significant at the 10 percent significance level. When combining this with results from Table 2.19 on all non-GBV-related arrest per crime rates, we conclude that unobservable events taking place at the same time as the Twitter social movements studied in this paper unlikely affect our main findings. None of the estimators reported in Table 2.19 are significant.

Table 2.17: The effect of social movements on GBV on crime rates per 100,000 inhabitants (Theft and Robbery)

	(1) Theft and Robbery	(2) Theft and Robbery	(3) Theft and Robbery
Twitter tweets	-0.170	-0.165	-0.103
	(0.167)	(0.125)	(0.133)
L.Twitter tweets		-0.0219	0.0668
		(0.147)	(0.123)
L2.Twitter tweets			-0.235
			(0.160)
Constant	23.89***	23.91***	23.94***
	(0.128)	(0.129)	(0.131)
Mean (Dep. Var)	23.86	23.89	23.90
St. Dv. (Dep. Var.)	20.90	20.92	20.92
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the crime rate. The outcome variable is the crime rate of thefts and robberies per 100,000 inhabitants per week and federal state. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous crime rate. Column 2 adds Twitter tweets in the following week, while Column 3 also considers Twitter tweets two weeks later. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.18: The effect of social movements on GBV on arrests per crime (Theft and Robbery)

	(1) Arrests	(2) Arrests	(3) Arrests
Twitter tweets	0.00163	0.000260	-0.000589
	(0.00240)	(0.00242)	(0.00248)
L.Twitter tweets		0.00236	0.00110
		(0.00183)	(0.00183)
L2.Twitter tweets			0.00328*
			(0.00193)
Constant	0.199***	0.199***	0.199***
	(0.00109)	(0.00108)	(0.00109)
Mean (Dep. Var)	0.200	0.200	0.200
St. Dv. (Dep. Var.)	0.0745	0.0745	0.0745
State-Month fixed-effects	Yes	Yes	Yes
N	5750	5710	5670

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the arrest per crime rate. The outcome variable is the arrest per crime rate of thefts and robberies per 100,000 inhabitants per week and federal state. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.19: The effect of social movements on GBV on arrests per crime (All crime but GBV)

	(1) Arrests	(2) Arrests	(3) Arrests
No. of Twitter tweets	0.0131	0.0112	0.00784
	(0.0108)	(0.00774)	(0.00796)
L.No. of Twitter tweets		0.00299	-0.00190
		(0.00855)	(0.00701)
L2.No. of Twitter tweets			0.0125
			(0.00878)
Constant	0.626^{***}	0.626^{***}	0.625^{***}
	(0.00206)	(0.00248)	(0.00263)
Mean (Dep. Var)	0.628	0.628	0.628
St. Dv. (Dep. Var.)	0.237	0.237	0.237
State-Month fixed-effects	Yes	Yes	Yes
N	5750	5710	5670

Notes: The table shows the results from a linear regression of the number of Twitter tweets on the arrest per crime rate. We define non-GBV-related crimes as all crimes besides GBV-related crimes. The explanatory variable is the number of GBV-related tweets in the federal state during the week, divided by 100 cellphone internet plan subscriptions in the federal state in that year. The unit of analysis is the week by federal state. The first column only considers the impact of Twitter tweets on the contemporaneous arrest per crime rate. Column 2 adds Twitter tweets in the previous week, while Column 3 also considers Twitter tweets two weeks previously. We weight each cell by the population size of each federal state in the respective year. We control for month of the year and state fixed effects. Month by state level clustered standard errors are reported in parenthesis. Source: NIBRS, Twitter and ACS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.7.2 Characteristics of Twitter Users

Twitter users engaging in the GBV-related debate might not be representative of the population in the US. To investigate this further, we compare the average characteristics of those reporting GBV to the police to those engaging in GBV-related Twitter tweets. In order to do so, we employ the *DeepFace* framework developed by Serengil and Ozpinar (2020) to Twitter users' profile pictures. This framework is a lightweight face recognition and facial attribute analysis package in Python.[21] It allows to retrieve the age, gender, emotion, and race from profile pictures. We apply this code to a 10 percent random sample of Twitter users tweeting, retweeting, or quoting tweet text in our dataset.

Our results show that the average user tweeting in our sample is 31.5 years old. Users are mainly White (63.8 percent). Around 9.2 percent are Black. 59.6 percent of Twitter users are female.[22]

When analyzing the basic socioeconomic characteristics of victims of GBV for the period 2014-2016, a large share is female (76.8 percent) (see Table 2.20). Moreover, victims of GBV are relatively young. The average age is 31.8 years, and only one quarter of victims is above 40 years old. The majority of victims is White (67.7 percent), followed by Black (27.1 percent). Only a small share is American Indian or Alaska Native (1.0 percent), Asian (1.0 percent), and Native Hawaiian (0.02 percent). 3.4 percent do not report any race. 9.4 percent of all victims are Hispanic.[23]

In conclusion, victims of GBV and those tweeting about GBV are similar in terms of age and ethnicity, but a lower share of tweet authors in our sample than victims of GBV reporting to the police is female. Overall, we conclude that Twitter users and victims of GBV are sufficiently close to each other on observable characteristics.

[21] Its accuracy is above 97.53 percent (Serengil and Ozpinar 2020).

[22] We apply the *GenderGuesser* tool to the first name of Twitter users to detect the gender of tweet authors. Appendix B.4 presents the details of this open-source Python package.

[23] Ethnicity is reported independently of race in the crime-incidence level reporting system. Most Hispanics are classified as White (95.9 percent), followed by Black.

Table 2.20: Descriptive statistics of victims of GBV (2014-2016)

VARIABLES	Mean	Std. Dev.	Min	Max	p25	p75
Age	31.81744	14.53767	0	99	22	41
White	0.67708	0.46759	0	1	0	1
Black	0.27078	0.44436	0	1	0	1
Hispanic	0.06969	0.25462	0	1	0	0
Female	0.76760	0.42236	0	1	1	1

Notes: The table shows descriptive statistics of characteristics of victims of GBV during the period 2014 to 2016 in the United States. *Age* is in years, while the rest of variables reports the share of those belonging to the respective group. While the *White* and *Black* variable is based on information gathered on victims' race, the Hispanic variable is drawn from information gathered on victims' ethnicity. Source: NIBRS (2014-2016).

2.7.3 Event Study

Our main regressions rely on several hashtags related to GBV, and while we believe in its validity, it is certainly not the only way one could measure social movements about GBV on Twitter. In order to test the robustness of our results, we conduct an event study. Through this, we can also further establish the causality of our estimates. Based on our findings from the main model specifications, we expect to find negative effects.

Event studies originally developed within the finance and accounting literature to estimate the impact of certain events on stock prices, starting with Brown and Warner (1985) and Dolley (1933), among others. Since then, their application has spread to other fields. Event studies have also been used to study the effects of civil unrest.[24] We follow this literature and define an event as a social movement started by one specific hashtag. We then measure the event by counting the number of English tweets using the hashtag under consideration.

For this purpose, we only consider the hashtag *#yesallwomen*. The *#yesallwomen* movement was a response to the 2014 Isla Vista killings, a series of killings of misogynistic nature taking place in May, 2014, close to the Santa Barbara Campus in California. The movement also partly emerged as a response to the hashtag *#notallmen*. Appendix B.1 shows that this movement was even larger than the *#metoo* movement when considering the number

[24]To name a few examples, Dave et al. (2020) use an event study design to study the effect of Black Lives Matter protests on risk avoidance, while Chernin and Lahav (2014) analyze the impact of social protests on the financial market.

of tweets using this hashtag during the first month of the movement. Additionally, it might outperform the *#metoo* movement in the context of the underlying research question as the general awareness about GBV was likely lower in 2014 than 2017. Therefore, one important empirical assumption of the event study, namely that there are no other confounding events taking place at the same time as the event under consideration, might be more likely.

One important underlying assumption of event studies is that the event occurs unexpectedly. Figure 2.8 shows that the number of tweets using *#yesallwomen* increased sharply on the 24th of May in 2014, the day that officially marked the start of this social movement on Twitter.

We estimate our event study at the federal state by week level. This allows us to exploit the fact that the hashtag *#yesallwomen* was trending in different states at different points in time. We determine the treatment status of a state in a respective week by its relative ranking in the Twitter tweet rate when compared to all other states. As soon as a state ranks in the upper third of states with the highest tweet rate, we consider the state as treated. We consider the first week a state ranks in the upper third as the moment the respective state starts to be exposed to the Twitter social movement. We then calculate the relative event time relative to this cutoff date and stag all event times to zero. We follow Clarke and Tapia-Schythe (2021) and implement our event study design through estimating the following regression:

$$Y_{gt} = \beta_0 + \sum_{j=2}^{J} \beta_j \times (Leadj)_{gt} + \sum_{k=1}^{k} \gamma_k \times (Lagk)_{gk} + \mu_g + \lambda_t + \epsilon_{gt} \tag{2.3}$$

In the above equation, g is the federal state, t is the week of the year fixed effect, μ_g are federal state fixed effects and λ_t are week of the year fixed effects. We consider 21 pre-treatment periods, as the *#yesallwomen* movement started in week 21 of 2014 and our sample starts in week one of 2014. We consider 52 post-treatment periods in order to include one full year after treatment exposure into our analysis. This means that $J = 11$ and $K = 52$.

Figure 2.9 shows the event study graph for Twitter tweet exposure to tweets using the hashtag *#yesallwomen*. The graph confirms our findings from previous analyses. There is a crime deteriorating effect of Twitter social movements on GBV with a significant drop in the coefficients after week 11. Importantly, the event study assumption is satisfied, as there is no clear pretrend in the event study graph.

It is worth mentioning that our main results from the methodology detailed in Section 2.4 and the event study analysis are not fully comparable. While we look at short-term effects in our main analysis, the event study design considers the weekly impact of Twitter tweets on crime rates up to one year later. Moreover, the event study graph controls for week of

Figure 2.8: Weekly number of tweets with the hashtag *#yesallwomen* (2014-2016)

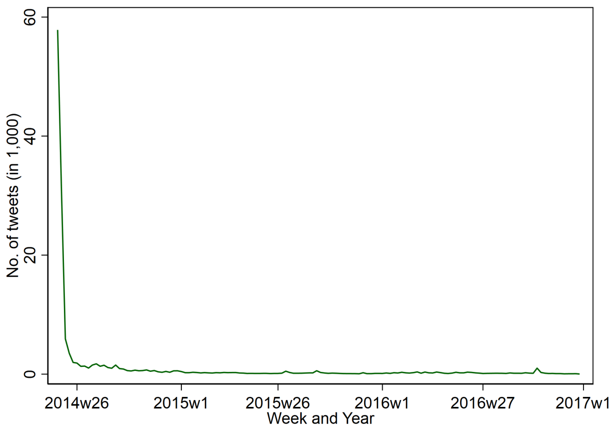
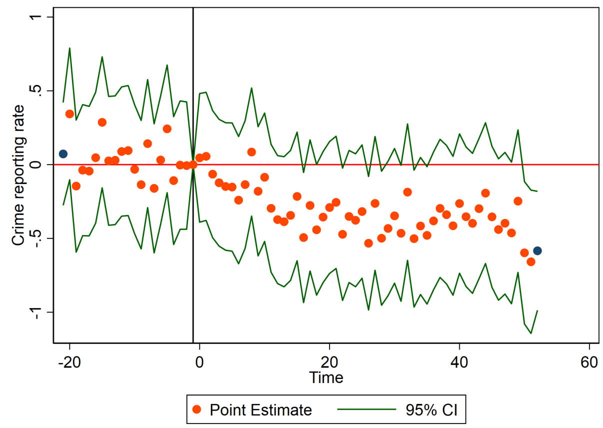


Figure 2.9: Event study graph at the week by state level



Notes: The figure on the left reports the weekly number of Twitter tweets using the hashtag *#yesallwomen*. The figure on the right shows the event study graph for Twitter tweet exposure of the hashtag *#yesallwomen* at the state by week level. The vertical black line indicates the last period, in which states are not treated. Treated states are those states ranking in the upper third of states with the highest Twitter tweet rate at least for one period prior to the week under consideration. We control for federal state and week of the year fixed effects. We consider 21 pre-treatment periods, as the *#yesallwomen* movement started in week 21 of 2014 and our sample starts in week one of 2014. We consider 52 post-treatment periods in order to include one full year after treatment exposure. For details on the estimation procedure see Clarke and Tapia-Schythe (2021). Source: Twitter, NIBRS, and ACS.

year fixed effects, while our main specification relies on month of year fixed effects.

2.7.4 The Role of Geography on Twitter

Our empirical strategy relies on the assumption that geography plays a significant role in the way in which information spreads on Twitter. While previous research shows that geographic networks play an important role on Twitter (Comito (2021); Hawelka et al. (2014)), information spreads quickly across regions. This might confound our empirical strategy, which relies on geographic variations. Consequently, our results could be subject to spillover effects of Twitter tweets between federal states. To account for this possibility, we conduct spatial regressions and investigate if our results hold when allowing for spillovers between neighboring states. More concretely speaking, we follow Lee and Yu (2010) and apply a spatial autoregressive model for panel data of the form:

$$y_{fw} = \lambda W y_{fw} + c_f + u_{fw} \tag{2.4}$$

$$u_{fw} = \rho M u_{fw} + v_{fw} \tag{2.5}$$

, where y_{fw} is the crime rate or arrest per crime rate in week w and federal state f, c_f is the area fixed effect, u_{fw} is the spatially lagged error and v_{fw} is an error term, which is assumed to be independent and identically distributed. M and W are spatial weighting matrices. We assume a random effects model and include a spatial lag of our independent variable, which is the number of Twitter tweets in the respective week and federal state.

The findings from spatial regressions presented in Table 2.21 confirm our main results. The impact of Twitter tweets on GBV-related crime rates is negative. Importantly, while the direct effect of Twitter tweets on crime rates in the same federal state is insignificant, the spillover effect is significant at the 1 percent significance level. The overall effect reported in the third row of the Table is significant and larger than the direct effect presented in the first row. Consequently, an increase in the number of Twitter tweets at the state-week level is associated with a decrease in the GBV-related crime rate.

Table 2.22 shows that - in line with the results from the main regression specification - coefficients remain insignificant in the case of arrest per crime rates. While the coefficient p-values on direct effects reported in Column 2 is slightly below the 10 percent significance level, the p-values on the indirect and total effect of Twitter tweets are larger than the common significance levels. Consequently, the overall impact of Twitter tweets on GBV-related arrest per crime rates is insignificant. These findings are in line with our main model specification, showing an insignificant impact on the contemporaneous coefficient.

To sum up, accounting for spatial spillover effects of Twitter tweets across federal states confirms our findings. In the case of crime rates, employing spatial regressions even reinforces

Table 2.21: Spatial Regression of Twitter tweets on Crime Rates at the state-week level

	Coefficient	P-Values
direct		
Twitter tweets	-.0006908	0.985
indirect		
Twitter tweets	-.1745702	0.001
total		
Twitter tweets	-.175261	0.000
Observations	5616	

Notes: The table shows the results from a spatial regression of the Twitter tweet rate on GBV-related crime rates. The unit of analysis is the week by federal state. The direct effect presents results for the impact of Twitter tweets in the same federal state, while the indirect effect presents the results on spatial lags. The total effect presents results for the combination of both. We control for state fixed effects and month fixed effects. Column 1 presents the regression coefficient, while Column 2 presents the coefficient p-value. Source: Twitter data (2014-2016).

Table 2.22: Spatial Regression of Twitter tweets on Arrest per Crime Rates at the state-week level

	Coefficient	P-Values
direct		
No. of Twitter tweets	.0050275	0.091
indirect		
No. of Twitter tweets	-.0003302	0.936
total		
No. of Twitter tweets	.0046973	0.190
Observations	5460	

Notes: The table shows the results from a spatial regression of the Twitter tweet rate on GBV-related arrest per crime rates. The unit of analysis is the week by federal state. The direct effect presents results for the impact of Twitter tweets in the same federal state, while the indirect effect presents the results on spatial lags. The total effect presents results for the combination of both. We control for state fixed effects and month fixed effects. Column 1 presents the regression coefficient, while Column 2 presents the coefficient p-value. Source: Twitter data (2014-2016).

our previous findings. The overall impact of Twitter social movements on GBV-related crime rates is larger when accounting for spillover effects between neighboring federal states. For arrest per crime rates, our results remain insignificant.

2.8 Conclusion

This paper examines whether Twitter social movements impact crime rates and arrest per crime rates of gender-based violence (GBV). We utilize text analysis and machine learning methods to create a novel dataset measuring the extent of online conversations on GBV on Twitter. We take advantage of the high frequency of our data and conduct regressions at the state by week level in the US, introducing a number of fixed effects to account for potential confounding factors. We also include lagged coefficients to allow for potential adjustment times of human behavior and to establish a causal interpretation of our results.

We find that Twitter social movements lead to a decrease in GBV-related crime rates of about 1 percent. We provide evidence showing that behavioral changes in perpetrators of GBV most likely drive our results. If, in addition, victims feel empowered and are more likely to report GBV, our coefficients are lower bound estimates of the true underlying effect. Moreover, we show that the impact of Twitter tweets on crime rates is most pertinent in the case of sexual violence. In this case, the reporting effect seems to outweigh the crime deteriorating effect. This pattern of results could mean that stigmatization, tabooing and silencing around sexual violence were especially persistent, and that Twitter social movements most likely addressed some of these barriers to the reporting of sexual violence.

Furthermore, analyzing the tweets' text via a sentiment analysis shows that the polarity of what is written does not play a significant role. Consequently, the pure magnitude of social movements seem to be more important than their content. Moreover, we find significant effects of social movements on GBV-related arrest per crime rates. The police force possibly increases its law enforcement as a result of rising social pressure. The increase in arrest per crime rates is solely driven by sexual violence, further establishing the potential importance of social stigma and tabooing.

We conduct an event study to shed further light on the causality of our main findings. The analysis confirms that Twitter social movements have a crime deteriorating impact. Moreover, we run placebo regressions to investigate the potential existence of unobservable confounding factors. Our results are robust to using non-GBV related crime rates and arrest per crime rates as outcome variables. Lastly, there is no evidence of selection bias with respect to observable personal characteristics of Twitter users and victims of GBV.

One important limitation of this paper is that there might be spillover effects of Twitter tweets between states. While previous research shows that geographic networks play an important role in Twitter tweets (Comito (2021); Hawelka et al. (2014)), information spreads quickly across regions. This might confound our empirical strategy, which relies on geographic variations. Still, allowing for spillover effects between neighboring states confirms our main findings, and enforces them for crime rates.

The pattern of results presented in this paper makes the case for Twitter platforms facilitating the signaling of social norms and having significant effects on offline behaviors. The evidence in this paper points towards Twitter tweets increasing social pressure and costs. These findings have important policy implications. Entities interested in decreasing the prevalence of GBV should explore the potential of social media platforms to do so. They can take advantage of these platforms to create informal support networks for those who experience GBV or signal social norms. Moreover, our paper generates novel insights on stigmatization, tabooing and silencing playing an important role in the reporting and arrests of GBV. Hence, policymakers should design strategies to address these barriers and facilitate the reporting and conversation on GBV. Lastly, although we do not find evidence in favor of backlash behavior, institutions should still secure environments safe of backlash cultures, especially within the police, to impede GBV.

Further research should investigate how to fully disentangle the effect on reported crime rates between crime perpetration and crime reporting. Moreover, it would be interesting to study if our results, which solely focus on social movements that take place on Twitter, differ from other social media platforms, such as Facebook. Future research could also explore alternative levels of variation instead of geographic variation. Lastly, future studies could investigate the degree to which traditional media coverage of Twitter social movements influences our results.

Chapter 3

Automation and Immigration: Migration Dynamics and the Adoption of New Technologies

"I am one of those who think like Nobel, that humanity will draw more good than evil from new discoveries." - Marie Curie

This chapter is based on work with Yvonne Giesing.

3.1 Introduction

Migration is one channel through which firms address skill shortages (Cameron 2011). It is also one important mechanism in the mitigation of local labor market shocks (Blanchard and Katz 1992). Technological change, such as the employment of industrial robots or the adoption of artificial intelligence (AI), play an important role in the emergence of local labor market shocks. The economic literature shows that AI potentially creates productivity as well as complementarity effects (Acemoglu et al. 2020). At the same time, new technologies have the potential to substitute automatable tasks and can result in displacement of workers (Acemoglu et al. 2020). To date, there is no consensus on the overall labor market impact of AI in the economic literature. While Webb (2019) finds that AI might replace high-skilled tasks and reduce wage inequality, Acemoglu et al. (2020) do not find any significant effects on employment nor on wages. For industrial robots Acemoglu and Restrepo (2018b) demonstrate negative labor market effects in the US. Similarly, Dauth et al. (2021) find displacement effects for industrial workers in Germany. Based on these findings, the question arises if migration is a tool to cover skill shortages and to mitigate labor market effects of emerging technologies.

In this paper, we study the impact of automation on migration patterns.[1] Answering this question sheds light on the mechanisms behind labor market effects of task automation. We first consider automation of high-skilled tasks. In particular, we investigate if AI adoption leads to an increase in immigrant inflows from abroad as well as an increased probability to migrate between German counties. AI might play a particular role here, as it creates highly specialized tasks (Zanzotto 2019). Germany is an ideal setting for the research questions investigated, as it is one of the leading countries in AI adoption and at the same time one of the main migration destinations. Given that AI-related skills are highly specialized, we expect to see a positive effect of AI adoption on internal migration inflows. At the same time, AI adoption might displace certain automatable tasks, potentially leading to migration outflows by those workers facing displacement effects. Consequently, we estimate the impact of AI adoption on net internal migration inflows.

The analysis reveals that AI adoption increases net internal migration inflows by German citizens. Importantly, this effect takes place across the skill distribution. For foreign citizens, on the other hand, we find a decrease in net internal migration flows. This applies to the low-, middle-, and high-skilled foreign population. There is no evidence in favor of significant immigrant inflows from abroad for none of the skill groups.

We follow Acemoglu and Restrepo (2018b) and use a local labor market approach and

[1] For related work see Rude and Giesing (2022) and Giesing and Rude (2022).

a Bartik (1991) shift-share instrument to investigate the effect of AI adoption on migration patterns. We use matched employer-employee data from Germany to measure migration patterns and labor market outcomes. For AI adoption, we rely on the share of job vacancies requiring at least one AI-related skill in each county calculated from Burning Glass Data. We focus on the period 2014-2019 due to data availability. Simple linear regressions would require that AI adoption at the county level is exogenous and not correlated with local labor demand. However, the adoption of AI could be subject to domestic industry-specific demand shocks. To address this endogeneity concern, we employ an instrumental variable strategy and instrument AI adoption in Germany with AI adoption in Switzerland at the industry level. We then proxy a local labor market's exposure to AI by exploiting the industry structure in each county.

Recent dynamics in migration policies in Switzerland build the case for AI adoption in Switzerland satisfying the exclusion restriction. Switzerland is neither part of the European Union nor the European Economic Area and follows its own migration policies, at least regarding migration from outside the European Union. In addition, in 2014, the referendum "against mass immigration" took place in Switzerland and was accepted by 50.3 percent of the electorate. The main goal of the referendum was to limit immigration through a quota system. While the Swiss parliament only implemented a softer version of this referendum, it still imposed several restrictions on immigration from European countries.

The lower net internal migration inflow by foreign citizens is in contrast to the standard economic theory predicting that foreigners are more likely to migrate internally, given that they already absorbed high fixed costs of previous movements (Borjas 2001). Our results could mean that foreigners are not perfect substitutes for natives to absorb AI-related skill demands. We investigate this further by analyzing the impact of AI adoption on displacement and wages. We do not find a significant impact of AI adoption on unemployment rates nor daily wages, neither in the case of German nor foreign citizens. This is in line with findings from the US, where AI adoption did not lead to any significant aggregated labor market effects (Acemoglu et al. 2020). We conclude that labor market effects are unlikely behind the fact that AI adoption affects foreigners' net internal migration inflow differently from natives'. In addition, our findings illustrate that AI creates very different labor market effects when compared to alternative automation technologies, such as industrial robots, for example (Dauth et al. 2021).

To shed further light on potential drivers behind the interaction of migration dynamics and automation, we investigate if differences in the probability to switch economic sectors could explain the diverging impact of AI on foreigners' and natives' cumulative net internal migration inflows. We find that there is indeed a difference between foreign and native citizens in the probability to switch sectors. This result points towards foreigners being less adaptive in response to automation. The effect is only significant for the high- and middle-skilled. These differences could emerge due to language barriers (Lochmann, Rapoport, and Speciale 2019) or lower access to important social networks (Martén, Hainmueller, and Hangartner 2019). We leave a detailed investigation of potential drivers to future research.

We next compare the impact of automation of high-skilled tasks to effects of automation of low-skilled tasks. We proxy automation of low-skilled tasks by the operational stock of industrial robots. Industrial robots are predominantly low-skilled automation technologies while AI has a greater relevance for high-skilled tasks (Acemoglu and Restrepo 2018a). Our analysis confirms that labor market effects of automation vary with the skill type technologies are likely to replace. In contrast to our results on AI adoption, industrial robots lead to diverging effects on wages by citizenship. While native citizens benefit from automation of low-skilled tasks, foreign citizens' daily wages decrease significantly across the skill distribution. Similar to our findings on automation of high-skilled tasks, there are no significant effects of industrial robots on cumulative immigrant inflows from abroad. In contrast, we find significant effects on cumulative net internal migration inflows. In the case of low-skilled task automation, foreigners' cumulative net internal migration inflows increase, while natives' inflows decrease.

We implement several robustness tests to validate our shift-share instrument, following empirical estimation prodecures proposed by Goldsmith-Pinkham, Sorkin, and Swift (2020) as well as Borusyak, Hull, and Jaravel (2022). These analyses reveal that industry shares are significantly correlated with county characteristics. The share component might therefore suffer from unobservable confounding factors. Shocks, on the other hand, seem to satisfy the exclusion restriction. Given that orthogonality in shocks is sufficient for the exclusion restriction to hold (Borusyak, Hull, and Jaravel 2022), these analyses reveal important limitations but overall confirm the robustness of our results. We also test if our results are confounded by urbanization. For this purpose, we restrict our sample to urban counties. The main results hold under this specification. Lastly, our results hold under different empirical specifications.

Our paper contributes to the literature studying labor market effects of automation technologies. On the effect of AI, Acemoglu et al. (2020) find that AI has not yet had any significant aggregate labor market effects, while Webb (2019) predicts a decrease in inequality through replacement effects on the high-skilled. In contrast to that, Felten, Raj, and Seamans (2019) show that AI could exacerbate current levels of inequality as it leads to an increase in wages of high-skilled occupations. Finally, Alekseeva et al. (2021) document an increased skill demand of AI in the US and a wage premium for these jobs. We contribute to this literature by studying if AI could potentially result in skill shortages, which are covered through migration, either from abroad or within Germany. In addition, our study provides insights regarding the extent to which internal migration can mitigate the impact of emerging technologies. Additionally, we explore heterogeneity of labor market effects with respect to the origin of employees, differentiating between natives and foreigners. This contribution sheds further light on underlying drivers behind labor market effects of technological change.

This paper is closely connected to several papers tying the topic of technological change to migration economics. Most closely, recent work by Hanson (2021) illustrates that foreign-born workers have accounted for more than half of the job growth in AI-related occupations since 2000. Our paper sets apart by studying the reversed causality of this question. Additionally, several other papers study the interaction between different forms of technologies and migration. Basso, Peri, and Rahman (2020) study the effect of computerization on immigration. They show that newly arrived immigrants specialize in manual service occupations and immigrants attenuate job and wage polarization facing natives owing to computerization. Hanson (2021) shows that an increase in the supply of high-skilled immigrants leads to an increase of AI in local labor markets. Work by Beerli, Indergand, and Kunz (2021) study the effect of ICT adoption in local labor markets on immigrant inflows in Switzerland. They demonstrate that a higher exposure to ICT leads to a significant inflow of high-skilled immigrants. For Germany, Danzer, Feuerbaum, and Gaessler (2020) analyze the effect of immigrant inflows on innovation, and find that they reduce innovation (measured by automation related patents). This especially applies to industries with many low-skilled workers.

Our work also speaks to the literature studying the effect of migration on innovation. Hunt and Gauthier-Loiselle (2010) find that immigrants patent at double the rate of natives, Peri and Sparber (2011) show that immigration influences the specialization of the native population and research by Lewis (2011) suggests that firms could see low skilled foreigners and automation machinery as substitutes.

This paper is connected to the literature studying drivers and consequences of internal migration. Work by Piyapromdee (2021), for example, demonstrates that internal migration is a mitigation mechanism for immigration inflows. Internal migration is also a mitigation mechanism as a response to economic downturns (Cadena and Kovak 2016).

Our findings are relevant for policymakers. First, retraining towards AI-related skills might have beneficial welfare effects if they address skill shortages. Second, our diverging results for natives and foreigners have important equity implications. Policymakers should devote special attention to the foreign population when designing mitigation policies in response to technological change; this would avoid further increases in inequality between foreigners and natives. There are several possible mechanisms to avoid these possibly negative equity implications. Policymakers should make sure that foreigners have equal access to information about the need to adapt their skill-set in response to technological change. In addition, equal access to retraining could also decrease potential diverge effects of technological change. On a general note, our results speak against foreigners and natives being skill-type perfect substitutes.

The paper is structured as follows. Section 3.2 provides an overview of recent trends in automation and migration in Germany. Section 3.3 presents theoretical rationales. Section 3.4 describes the datasets used and outlines our empirical strategy. Section 3.5 presents our main results. Section 3.6 looks at the underlying mechanisms behind these results. Section 3.7 provides a comparison to low-skilled automation. Lastly, Section 3.8 presents several robustness checks and Section 3.9 concludes.

3.2 AI Adoption and Migration: the case of Germany

The following section provides an overview of recent trends in AI adoption as well as immigration. Recent developments highlight the relevance of the underlying research questions. The descriptive statistics also show that Germany is the ideal setting for the research question at hand.

3.2.1 Recent Trends in AI

Figure 3.1 plots the number of AI-related patent families and scientific publications by publication year over time. The figure highlights the exponential increase observed for this technology over time. Especially since 2014, AI technologies have been on the rise. When analyzing AI adoption by region, the number of AI-related patent applications has increased for the three economic players over time, with China catching up with the US by 2014 (see Figure 3.2). Since 2000 the number of AI-related patent applications increased by a factor of 3.5 for the US and 2.9 for Europe. For China, the number of AI-related patent applications in 2014 was more than 23 times the observed value in 2000.

The growth in the number of AI-related patents was associated with an increase in the demand for AI-related skills. We follow a combined keyword list by Acemoglu et al. (2020) and Chiarello et al. (2021) to identify AI-related skills (Section 3.4 outlines the details of this measure). We measure the labor demand for these skills by relying on online job vacancy data from Burning Glass. The increase in absolute terms was greatest for Germany, followed by France (see Figure 3.3). To measure the relative demand for AI-related skills, we rely on the share of AI-related skill demand in all skill demand. Figure 3.4 plots the relative

Figure 3.1: Number of AI-related patent families and scientific publications over time

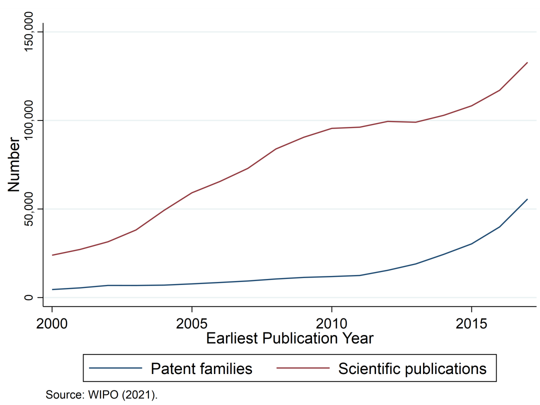
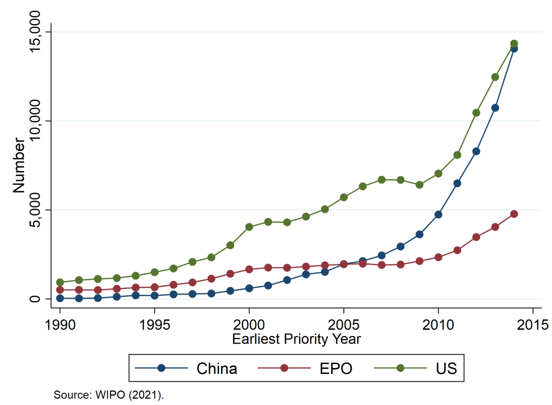


Figure 3.2: Number of AI-related patent applications in three patent offices over time



Notes: The left panel plots the number of AI-related patent families (in blue) and scientific publications (in red) over time by earliest publication year for the period 2000 to 2017. A patent family is a conglomeration of related patents. The right panel plots the number of AI-related patent applications by earliest priority year over time for three patent offices. The blue line represents China, the green line the United States (US), and the red line the European Patent Office (EPO). A priority year is the year of a patent's first patent filing. Source: WIPO (2021).

demand for AI-related skills in selected European countries for the period 2014 to 2019. The share of AI-related skills is low with around 0.1 percent across all countries under consideration. Moreover, German-speaking countries report the highest share, together with the Netherlands. Switzerland is in the lead.

The rise in the demand for AI-related skills comes along with several researchers observing skill shortages with respect to these skills. Metz (2017) note that Big Tech companies pay huge salaries for scarce AI talent. In addition, a report by Anderson, Viry, and Wolff (2020) concludes that Europe faces a dearth of AI talent and demonstrates that firms in Germany spent on average six months filling tech positions. Combining the growth in AI-related skill demand and the evidence pointing towards shortages of these skills underscores the relevance and timeliness of tying labor market effects of AI adoption to migration dynamics.

Figure 3.3: AI-related skill demand over time by selected European countries

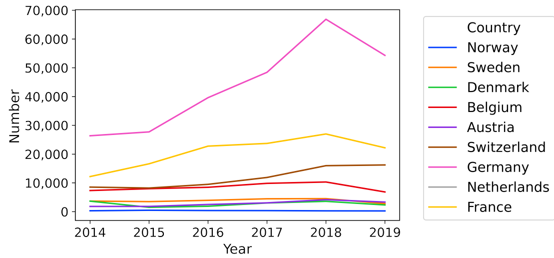
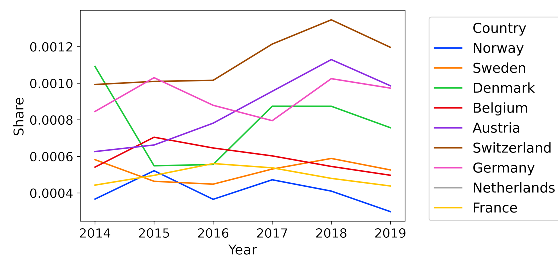


Figure 3.4: Share of AI-related skill demand over time by selected European countries



Notes: The left figure plots the absolute number of AI-related skills mentioned in online job vacancies for nine European countries with available data for the period 2014 to 2019. The right panel presents the share of AI-related skill demand in all skill demand for nine European countries with available data for the period 2014 to 2019. Source: Burning Glass Data (2014-2019).

3.2.2 Germany's Role in Automation

Germany is the ideal setup for the underlying research question, as it is a highly automated country. Germany is the fourth largest economy in the world when measured by GDP. Germany's industrial sector accounts for 26.5 percent of GDP, while the service and primary sectors comprise 63.3 and 0.7 percent, respectively (The World Bank 2021). Along with the importance of the industrial sector for the German economy, there is a long history of automation. In fact, Germany is the most automated economy in Europe, when measured by industrial robots. Figure 3.5 shows that Germany is among the top five countries worldwide in terms of installed industrial robots. In 2019 alone, Germany installed more than 22,000 industrial robots. In comparison, the US installed around 33,000 and China 139,859 industrial robots in the same year. It is the major player among European countries, even when measuring the stock of industrial robots per employees (see Figure 3.6).

A similar picture emerges when analyzing Germany's role in the production of artificial intelligence. Figure 3.7 shows that Germany is among the Top 10 Artificial Intelligence producers in 2017, when measured by the number of patent filings (OECD 2021). The country filed 400 AI patents in 2017, and was the largest player in the European market until 2016, when the UK caught up with Germany (see Figure 3.8). In comparison, the US filed 6,728 patents in 2017 and China 1,674. Therefore, Germany is a highly relevant country in the realm of artificial intelligence. This further bolsters the case for studying the underlying research question in German labor markets.

Figure 3.5: Operational stock of robots in 2019, Top 15 economies

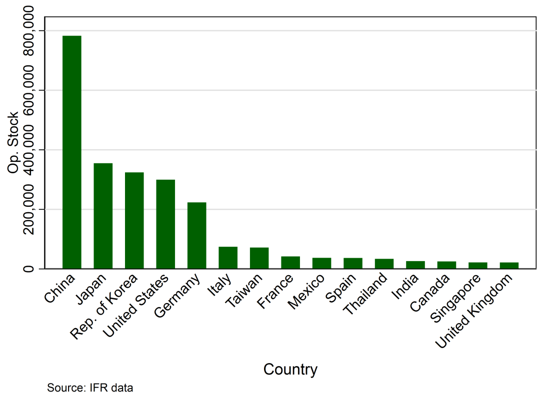
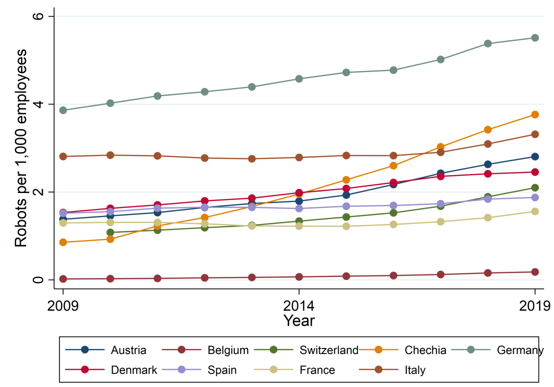


Figure 3.6: Operational stock of robots (per 1,000 employees) over time in European countries



Notes: The left panel plots the operational stock of industrial robots by country in 2019 for the 15 countries with the highest operational stock in robots. The right panel plots robot exposure over time for the period 2009-2019 in nine European countries. Robot exposure is the number of installed robots per 1,000 employees. Source: IFR data.

Figure 3.7: AI patent filings in 2017, Top 15 economies

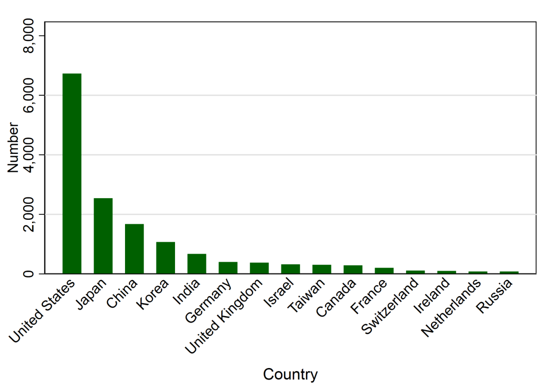
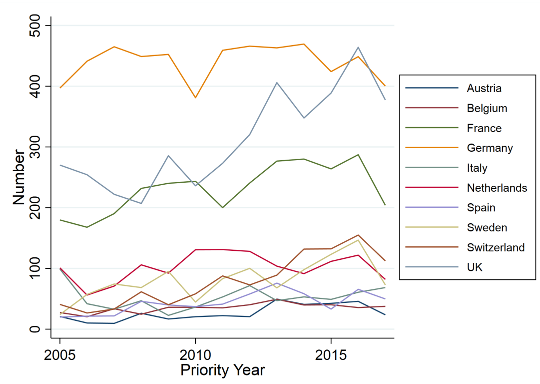


Figure 3.8: AI patent filings in European countries over time



Notes: The left panel plots the number of AI-related patent filings in 2017 for the top 15 countries in 2017. The right panel plots the number of AI-related patent filings in European countries over time for the period 2005 to 2017. Source: OECD patent data (2021).

3.2.3 Germany's Recent Trends in Migration

In what follows, we descriptively analyze Germany's recent patterns in migration flows. Germany is a migration recipient country and has been for many years. The annual influx of foreign citizens to Germany was over half a million between 2000 and 2013, and surpassed one million for the period 2013 to 2019 (Federal Statistical Office, 2021). Figure C.1 plots the immigrant inflow over time. At the same time, Germany has been subject to constant outflows of foreign citizens, but also German citizens (see Figure C.2 for details). For most years, the country's migration balance has been largely positive, with a balance fluctuating between 127,000 and 1.1 million since 2010 (Federal Statistical Office, 2021). Germany has been the main migrant-receiving country among OECD countries, overtaking the US in 2012, in terms of yearly inflows (see Figure C.3).

Figure 3.9: Immigrant inflow by skill group

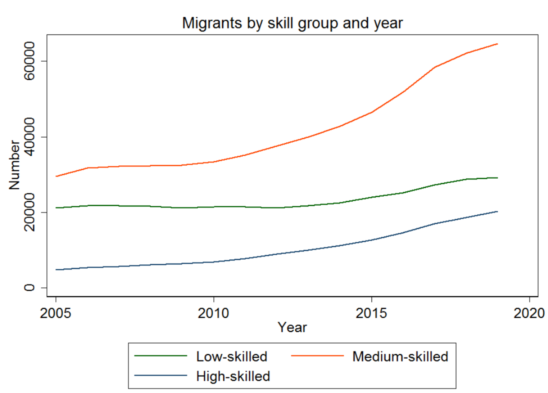
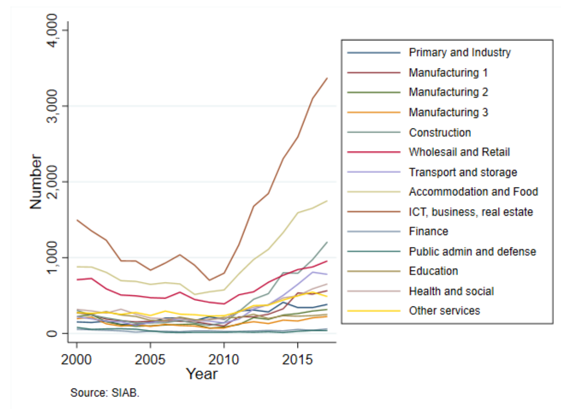


Figure 3.10: Immigrant inflow by sector



Notes: The graphs shows the immigrant inflow by skill groups and economic sectors. We follow the IAB's definition of migrants, which is based on citizenship. This means that a person is identified as a migrant as soon as they do not hold a German citizenship. The inflow is identified via the first observation of a foreigner in the SIAB data. Skill level are based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. The sector variable captures the economic activity in accordance with the WZ08 classification, grouped into 14 broad categories. *Manufacturing sector (1)* refers to the manufacture of food products, manufacture of beverages, retail sale of food, beverages and tobacco in specialised stores, wood and wood products, as well as other manufacturing. *Manufacturing sector (2)* refers to the manufacture of coke and refined petroleum products, chemical and pharmaceutical products, as well as rubber and plastics, and of basic metals and fabricated metal products. *Manufacturing sector (3)* refers to the manufacture of computer, electronic and optical products, electrical equipment, mechanical engineering, and vehicle manufacturing. The graphs show data from the SIAB, which is a 2 percent sample of all individuals with an entry in the Integrated Employment Biographies (IEB).

Figure 3.9 plots the immigrant inflow to Germany over time by skill group. There has been a constant increase of immigrants for each of the skill groups, but the increase has been largest for the medium-skilled. Moreover, when looking at the immigrant inflow by economic sector, Figure 3.10 shows that the inflow has been most significant for the ICT and Business Services Sector. The second largest increase of foreigners is to the Accommodation and Food Service Sector. Importantly, the migrant share, measured as the number of foreigners relative to the whole population registered in the SIAB data (Section 3.4 gives a detailed overview of the SIAB data), is above five percent for all sectors except the Financial Services Sector and the Public Administration and Defense sector (see Figure C.4). The migrant share is greatest for the Accommodation and Food sector (nearly 30 percent), followed by the Construction and the ICT, Business and Real Estate as well as Transport and Storage sector (all above 15 percent) (SIAB, 2021).

3.3 Theoretical Rationale

This paper is based on different theoretical rationales. First, we rely on insights from a task-based model developed by Acemoglu et al. (2020). They derive four effects of AI adoption on labor market outcomes. First of all, AI leads to a displacement effect, which evolves due to the replacement of certain tasks traditionally performed by labor through technology. The replacement of tasks results in a contraction in the number of tasks, leading to downward pressure on wages. At the same time, there is a productivity effect. The productivity effect evolves due to expensive labor being substituted by cheaper capital. Firms thus increase their demand for labor in tasks that are not yet automated, a so-called complementarity effect. Lastly, there is a reinstatement effect, meaning a creation of new tasks generated by AI adoption.

The overall labor market effect of high-skilled automation depends on whether the productivity, complementarity, and reinstatement effect outweighs the displacement effect. In addition, according to the framework developed by Acemoglu et al. (2020) there might be additional labor market effects induced by labor market frictions. These frictions might evolve as a response to the rapid employment of technologies and the need for new skills. Retraining employees, on the other hand, takes time.

When applying this framework to the migrant and native population, one should also consider additional theoretical rationales. First, migrants can substitute or complement natives (Battisti et al. 2018). An inflow of low-skilled migrants, for example, might complement native workers from a certain skill group (such as the high-skilled), but substitute those from the same skill group (such as the low-skilled native population). That is another rationale for this paper, as the question arises whether firms substitute cheap labor from abroad with cheap automated technologies. If this was the case, we would expect to see a decline in the immigrant inflow in those skill groups most affected by automated technologies investigated in this paper. In addition, most OECD countries exhibit a wage gap between migrants and natives due to differences in labor productivity as well as outside options of natives (Battisti et al. 2018). We build upon this evidence by asking if these differences in labor productivity or outside options create a differing impact of technological change on migrants versus natives. Moreover, ex ante differences in labor productivity could increase through automation. We investigate these potential channels empirically in this paper.

Additionally, similarly educated migrants might not be perfect substitutes for natives from the same skill group as they have different levels of experience (Borjas 2003). The fact that natives and foreigners might not be perfect substitutes is another rationale for our research question. It motivates a differentiation between foreigners and natives when studying labor market effects of technological change. Moreover, foreigners might be more willing to migrate internally than natives (Borjas 2001), and thus mitigate the effects of technological change on natives. The greater willingness to migrate internally stems from the fact that migrants already assumed the high fixed costs of migration in the past. We take this into account by looking at how technological change influences the probability to migrate, differentiating between foreign and German citizens.

3.4 Empirical Strategy

3.4.1 Data Sets in Use

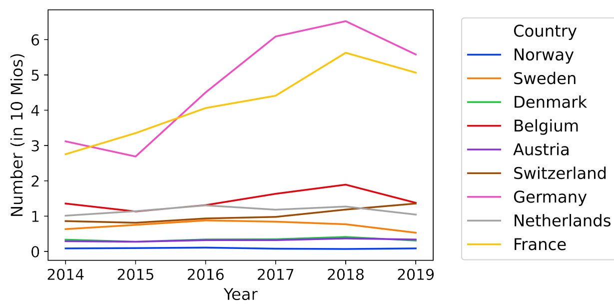
We make use of several different datasets in order to address the underlying research question. First, we harness Online Job Vacancy (OJV) data provided by Burning Glass. The data is available for ten European countries for the period 2014 to 2019.[2] For each job vacancy we have information about the NUTS-3-region the job add refers to, the respective economic sector (at the 2-digit-level), the occupation (at the 4-digit-level) as well as all skills mentioned in the job vacancy (at the ESCO-level-3). We also have the official description of each of these skills provided by the European Commission. Our analysis is based on a total number of 48.5 million job vacancies in Germany for the period 2014 to 2019. The data covers virtually the entire spectrum of OJVs in Germany, as it also extracts information from the country's public employment agencies.

Figure 3.11 shows that skill demand in Germany registered in OJV has nearly doubled over time, from 31.9 million in 2014 to 55.8 million in 2019. This could be due to economic

[2]These countries are Austria, Belgium, Denmark, France, Germany, Luxembourg, the Netherlands, Norway, Sweden and Switzerland.

growth but also due to job adds being increasingly transferred to the virtual space. It could also mean that jobs have become more complex over time and require a larger variety of skills. The number of AI-related skill demand registered in online job vacancies increased by 106 percent between 2014 and 2019, from 26,381 to 54,320 (see Figure 3.3). The share of AI-related skill demand in all skill demand is still extremely low at 0.1 percent in 2019.

Figure 3.11: Number of skills in demand over time by selected European countries



Notes: The figure plots the absolute number of skills mentioned in OJV for nine European countries with available data for the period 2014 to 2019. Source: Burning Glass Data (2014-2019).

We measure labor market outcomes of immigrants and German citizens in Germany using administrative individual-level spell data provided by the Institute for Employment Research (IAB) (Antoni et al. 2021). We use the Sample of Integrated Labor Market Biographies (SIAB). The SIAB is a two percent sample of the population of the Integrated Employment Biographies (IEB) of the IAB.[3]

The SIAB contains information on the following individuals: employees covered by social security (including marginal part-time employees from 1999 onwards), benefit recipients, job-seekers, as well as participants in active labor market policies. The SIAB covers all white-and blue-collar workers as well as apprentices as long as they are not exempt from social security contributions. This means that civil servants, self-employed persons and regular students are not recorded in the SIAB in principle (Cramer 1985). The dataset comprises information on the following topics: the employee history, benefit recipient history, unemployment benefit recipient history, the job seeker history and information on participation in employment and training measures. We prepare the SIAB dataset closely following the methodology proposed by Dauth and Eppelsheimer (2020) in order to create a dataset in panel-format with annual observations per individual.

We follow the SIAB's definition of an immigrant, which is based on the citizenship an individual holds. Based on this definition, somebody is a foreigner as soon as they do not

[3]The SIAB encompasses the employment histories of 1,940,69 individuals, and their employment biographies are documented in a total of 72,225,126 lines of data. Of these, 12.7 percent of observations (a total of 7.5 million data entries) are related to non-German nationalities.

hold German citizenship. We construct the immigrant inflow per year and county by creating a dummy variable which is equal to one if a foreign citizen is observed for the first time in the SIAB data. We then aggregate the dummy variable at the county by year level. To measure internal migration inflows, we create a dummy variable, which is equal to one, as soon as a county of residence diverges from the one observed in the previous year. By taking the sum of this dummy variable at the county-year-level, we measure yearly internal migration inflows into a certain county. Similarly, we measure internal migration outflows by a dummy variable, which is equal to one if the county of residence in the next year is different from this year's county of residence. We rely on the county of residence to also include movements of unemployed people.

We identify the skill level of an individual through the imputed educational variable included in the SIAB. The low-skilled are those individuals who have neither vocational training nor a university degree. The middle-skilled have vocational training and the high skilled hold a degree from a university or university of applied science. The sector variable captures the economic activity in accordance with the WZ08 classification, grouped into 14 broad categories. We construct several control variables on basic labor market characteristics based on data from the SIAB as well as Eurostat and Comtrade.

Table C.14 provides an overview of the main variables considered in this paper at the yearly and county level.

Table 3.1: Descriptive table of main variables of interest (at the year by county level)

	N	Mean	Standard Dev.	Min	Max
Share of women	2416	0.4869	0.0398	.25	.7142857
Share of low-skilled	2416	0.1212	0.0361	0	.2577207
Share of middle-skilled	2416	0.7341	0.0655	.4914844	.8672269
Share of high-skilled	2416	0.1447	0.0560	.0505618	.3878738
Share of <35	2416	0.2990	0.0366	0	.4284507
Share of 35-54	2416	0.4583	0.0276	.25	.7142857
Share of >54	2416	0.2427	0.0374	.152381	.75
Share of part-time	2416	0.3814	0.0497	0	.572549
Share in manufacturing	2416	0.2248	0.0995	.0199161	.75
Share in ICT	2416	0.0196	0.0189	0	.1444393
Yearly immigrant inflow	2416	32.3142	213.5609	0	7030
Yearly immigrant inflow (High-skilled)	2416	6.1461	34.4985	0	1066
Yearly immigrant inflow (Middle-skilled)	2416	17.4739	136.4566	0	4583
Yearly immigrant inflow (Low-skilled)	2416	8.3228	38.7852	0	1218
Yearly immigrant inflow of employed	2416	22.9230	44.4356	0	680
Yearly immigrant inflow of employed (High-skilled)	2416	N/D	N/D	N/D	N/D
Yearly immigrant inflow of employed (Middle-skilled)	2416	22.9230	44.4356	0	680
Yearly immigrant inflow of employed (Low-skilled)	2416	6.6676	11.2911	0	137
Internal migration inflow (Foreign)	2412	8.0112	12.9106	0	183
Internal migration inflow (Low-skilled, foreign)	2412	2.0659	3.3281	0	33
Internal migration inflow (Middle-skilled, foreign)	2412	4.2330	6.2186	0	71
Internal migration inflow (High-skilled, foreign)	2412	1.7123	4.2396	0	83
Internal migration inflow (German)	2412	44.7745	61.3829	0	801
Internal migration inflow (Low-skilled German)	2412	7.3682	11.8177	0	161
Internal migration inflow (Middle-skilled German)	2412	25.1111	26.7578	0	301
Internal migration inflow (High-skilled German)	2412	12.2952	25.4105	0	364
Internal migration outflow (Foreigner)	2412	8.1712	12.4391	0	183
Internal migration outflow (Low-skilled foreigners)	2412	2.2388	3.5224	0	51
Internal migration outflow (Middle-skilled foreigners)	2412	4.3188	6.3792	0	87
Internal migration outflow (High-skilled foreigners)	2412	1.6136	3.3402	0	45
Internal migration outflow (German)	2412	44.6269	57.5917	0	733
Internal migration outflow (Low-skilled German)	2412	9.1414	10.7540	0	134
Internal migration outflow (Middle-skilled German)	2412	24.7301	28.7270	0	365
Internal migration outflow (High-skilled German)	2412	10.7554	20.1409	0	245
Migrant Share	2416	0.0864	0.0489	0	.3024255
Migrant Share (Low-skilled)	2412	0.1844	0.0934	0	.4957265
Migrant Share (Middle-skilled)	2416	0.0698	0.0442	0	.2870346
Migrant Share (High-skilled)	2416	0.0874	0.0437	0	.3734777
Unemployment rate (German)	2412	0.0194	0.0064	0	.0525
Unemployment rate (Low-skilled, German)	2412	0.0256	0.0184	0	.1698113
Unemployment rate (Middle-skilled, German)	2412	0.0192	0.0069	0	.0583942
Unemployment rate (High-skilled, German)	2412	0.0162	0.0113	0	.0810811
Unemployment rate (Foreign)	2412	0.0276	0.0219	0	.1875
Unemployment rate (Low-skilled, Foreign)	2412	0.0070	0.0090	0	.0638298
Unemployment rate (Middle-skilled, Foreign)	2412	0.0023	0.0027	0	.0347222
Unemployment rate (High-skilled, Foreign)	2412	0.0022	0.0041	0	.0344828
Daily wage (Foreign)	2412	63.9314	12.5593	23.16956	148.0571
Daily wage (Low-skilled, Foreign)	2393	46.1626	11.2783	4.325736	91.29745
Daily wage (Middle-skilled, Foreign)	2412	59.6657	12.0398	15.76131	121.2609
Daily wage (High-skilled, Foreign)	2398	105.3805	41.9916	1.14	477.0361
Daily wage (German)	2416	83.6386	16.2671	47.17413	172.3029
Daily wage (Low-skilled, German)	2412	38.8737	7.6860	15.50581	81.87044
Daily wage (Middle-skilled, German)	2416	77.8539	11.8270	44.16252	136.3969
Daily wage (High-skilled, German)	2416	144.2587	31.2408	64.93137	282.3608
Yearly labor earnings (Foreign)	2412	22440.2613	4850.8194	1241.672	52002.17
Yearly labor earnings (German)	2416	30534.2296	6079.6233	3683.819	62167.03
Yearly labor earnings (Low-skilled, Foreign)	2393	16091.9354	4338.3855	1278.313	35927.73
Yearly labor earnings (Middle-skilled, Foreign)	2412	20957.6991	4836.7190	1115.3	44869.53
Yearly labor earnings (High-skilled, Foreign)	2402	37102.6597	15053.8868	227.12	157923.2
Yearly labor earnings (Low-skilled, German)	2412	14301.3140	2858.4751	3754.3	29868.91
Yearly labor earnings (Middle-skilled, German)	2416	28435.7044	4580.4299	3364.082	49335.31
Yearly labor earnings (High-skilled, German)	2416	52599.0438	11372.8790	5964.068	101825.5
Trade exposure	2417	3843853.6600	1814267.4933	0	2.48e+07
AI skill demand (in 0,01 percent)	2297	0.0531	0.0789	0	2.173913
AI skill demand (IV, in 0,01 percent)	2417	0.0576	0.0088	0	.106961
Observations	2419				

Notes: The table shows summary statistics of the underlying dataset at the county (NUTS-3) by year level for the period 2014 to 2019. We follow the IAB's definition of immigrants, which is based on citizenship: a person is a foreigner as soon as they do not hold German citizenship. Immigrant inflows are identified via the sum of foreigners' first observations in the SIAB data for a given year by county cell. Skill levels are based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. The sector variable captures the economic activity in accordance with the WZ08 classification, grouped into 14 broad categories. The shares refer to the respective shares in the overall SIAB population. Trade exposure is constructed via the shift-share instrument (see Section 3.4.2 for details) and reported in Euros. The original trade variable is at the national-industry level and represents the trade value in Euros. AI is the share of AI-related skills in all skills mentioned in OJV. The instrumented AI-related skill demand uses data from Switzerland (see Section 3.4.2 for details). Source: SIAB, Burning Glass, and Comtrade.

3.4.2 Empirical Strategy

We follow Acemoglu and Restrepo (2018b) and study the effect of AI at the local labor market level. For this purpose, we aggregate the SIAB data to the county level. We consider the period of 2014 to 2019 due to data availability. We measure a local labor market's adoption of artificial intelligence by constructing a variable equal to a local labor markets' demand for AI-related skills. We conduct a keyword search on terms relevant to AI to detect all online job vacancies demanding AI-related skills in the Burning Glass dataset described above. We rely on keywords defined by Acemoglu et al. (2020) and Chiarello et al. (2021). Having a look at the list of keywords illustrates that the terms refer to high-skilled tasks.[4] As soon as one of these keywords forms part of an ESCO-skill or its description, we assign it a value of one. We then calculate the share of these skills within all skills in demand in a local labor market.

We run a stacked-difference regression at the local labor market for two sub-periods of three years, 2014-2016 and 2017-2019. Our main explanatory variable is the percentage point change in the share of online job vacancies requiring at least one AI-related skill in a local labor market during the stacked time period of three years:

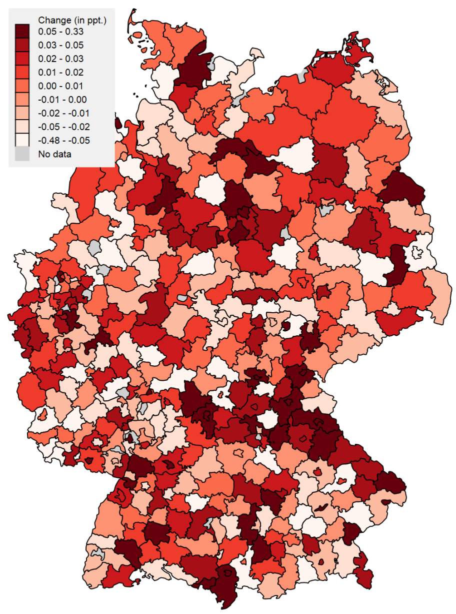
$$AI_{rj} = (AI_{r,t1} - AI_{r,t0}) * 100 \tag{3.1}$$

In the above equation, j is a three-year period (e.g. 2014-2016), r is a local labor market, $t1$ is the last year of the three-year period under consideration (e.g. 2016) while $t0$ is the first year (e.g. 2014). Figure 3.12 shows the percentage point difference in the share of AI-related skill demand for the complete period under consideration (2014-2019). While some counties report negative growth, others have experienced a difference in the share of AI-related skill demand of up to 0.33 percentage points. Overall, changes are small. Therefore, we scale AI adoption by 100. A one unit change in the main explanatory variable is then equivalent to a 0.01 percentage point increase.

We estimate the impact of AI adoption on the percentage change in the migrant share, unemployment rates and daily wages. We construct these variables as follows:

[4] These terms are Artificial Intelligence, Machine Learning, Decision Support System, Speech Recognition, Natural Language Processing, Computational Linguistics, Speech Recognition, Virtual Machine, Deep Learning, Biometrics, Neural Networks, Computer Vision, Machine Vision, Virtual Agents, Image Recognition, Data Mining, Pattern Recognition, Object Recognition, AI ChatBot, Text Mining, Support Vector Machines, Unsupervised Learning, Image Processing, Mahout, Recommender Systems, SVM, Random Forests, Latent Semantic Analysis, Sentiment Analysis, Opinion Mining, Latent Dirichlet Allocation, Predictive Models, Kernel Methods, Keras, Gradient boosting, OpenCV, Xgboost, Libsvm, Word2Vec, Chatbot, Machine Translation, Sentiment Classification.

Figure 3.12: Difference in the share of AI-related skill demand at the county level (2014 to 2019)



Notes: The map shows the percentage point difference in the demand for AI-related skills between 2014 and 2019 in a respective county. We define AI-related skill demands as the share of AI-related skills (defined by keywords from Acemoglu et al. (2020) and Chiarello et al. (2021)) in all skills in demand. The unit of analysis is the county. Darker colors indicate that demand for AI-related skills increased, while brighter colors indicate that demand for AI-related skills decreased. Source: Burning Glass Data (2014-2019).

$$V_{rj} = \frac{V_{r,t1} - V_{r,t0}}{V_{r,t0}} * 100 \tag{3.2}$$

Moreover, we investigate the impact of AI adoption on the cumulative immigrant inflow from abroad and cumulative net internal migration flows. We measure the cumulative immigrant inflow by taking the sum of the dummy variables described in Section 3.4 for the two three-year periods investigated in this paper for each local labor market. We calculate the cumulative net internal migration inflow by subtracting the cumulative migration inflow by the cumulative migration outflow. Table 3.2 gives an overview of the main variables of interest.[5]

We control for demographic characteristics at the county-level in the pre-treatment period in 2014, namely the female share, the overall share of different skill groups, the share of workers belonging to different age groups, the share of workers employed in the manufacturing sector as well as in the ICT sector. We also control for regional dummies at the federal state (NUTS-1) level and cluster our standard errors at the county level (the NUTS-3 level). We additionally control for the difference in trade exposure at the local labor market relying on Comtrade data. We measure a county's trade exposure via a shift-share instrument. The original trade variable is at the national-industry level and represents the trade value in Euros. This is an important control variable as changes in trade patterns might confound our results.

We weight our regression by the SIAB population observed in each local labor market to account for the level of aggregation and give more weight to those local labor markets with a greater SIAB population. The equation below shows our baseline regression specification:

$$Y_{rj} = \alpha X'_{r2014} + \beta_1 \times AI_{rj} + \beta_2 \times \hat{\text{trade}}_{rj} + \varphi_{REGrj} + \epsilon_r \tag{3.3}$$

r is a local labor market and j a three-year period. AI_{rj} is the main explanatory variable of interest, namely the percentage point change in the share of AI-related skill demand in all skill demand. $\hat{\text{trade}}_{rj}$ is a shift-share instrument of trade exposure at the local labor market and φ_{REGrj} are regional dummies. X'_{r2014} is a vector of pre-treatment labor market characteristics.

Simply analyzing the impact of an increased AI adoption through linear regressions could bias our estimates, as locations with more AI adoption might systematically differ on unobservable characteristics from those with less exposure to AI. Consequently, to estimate

[5]The table reveals an important data limitation. The unemployment variable suffers from missing observations. Especially in the case of foreigners' unemployment rates many of the stacked time period by county cells do not report values.

Table 3.2: Descriptive table of main variables of interest (stacked-differences)

	N	Mean	Standard Dev.	Min	Max
Ppt. change in AI	762	.1269391	9.815061	-217.3913	33.25943
Ppt. change in AI (IV)	805	.0309671	1.140967	-5.357176	2.036067
Cum. immigrant inflow	806	96.86228	586.1769	0	12768
Cum. immigrant inflow (High-skilled)	806	18.42308	95.42616	0	1964
Cum. immigrant inflow (Middle-skilled)	806	52.37841	372.8536	0	8222
Cum. immigrant inflow (Low-skilled)	806	24.94789	107.8545	0	2297
Pct. change in migrant share	804	22.18748	24.15737	-35.3166	198.6842
Pct. change in migrant share (Low-skilled)	795	25.19114	56.20718	-100	709.1771
Pct. change in migrant share (Middle-skilled)	804	26.96228	29.66026	-40.91448	208.8097
Pct. change in migrant share (High-skilled)	797	33.72741	63.96545	-100	665.4508
Cum. internal migration inflow (German)	806	133.9901	168.7659	0	2277
Cum. internal migration inflow (High-skilled, German)	806	36.79404	71.48767	0	1007
Cum. internal migration inflow (Middle-skilled, German)	806	75.1464	70.14644	0	838
Cum. internal migration inflow (Low-skilled, German)	806	22.04963	32.71351	0	432
Cum. internal migration inflow (Foreign)	806	23.97395	35.907	0	463
Cum. internal migration inflow (High-skilled, Foreign)	806	5.124069	11.57313	0	171
Cum. internal migration inflow (Middle-skilled, Foreign)	806	12.66749	16.8681	0	203
Cum. internal migration inflow (Low-skilled, Foreign)	806	6.182382	8.819388	0	89
Cum. internal migration outflow (Foreign)	806	24.45285	31.30499	0	374
Cum. internal migration outflow (Low-skilled, Foreign)	806	6.699752	8.608774	0	91
Cum. internal migration outflow (Middle-skilled, Foreign)	806	12.92432	15.65634	0	191
Cum. internal migration outflow (High-skilled, Foreign)	806	4.828784	8.280365	0	92
Cum. internal migration outflow (German)	806	133.5484	152.759	0	2026
Cum. internal migration outflow (Low-skilled, German)	806	27.35608	28.04988	0	352
Cum. internal migration outflow (Middle-skilled, German)	806	74.0062	74.8815	0	1018
Cum. internal migration outflow (High-skilled, German)	806	32.1861	54.03392	0	656
Cum. internal migration inflow (Net)	806	.4416873	111.4395	-1000	953
Cum. internal migration inflow (Net, high-skilled German)	806	4.60794	42.2444	-284	580
Cum. internal migration inflow (Net, middle-skilled German)	806	1.140199	52.75752	-569	270
Cum. internal migration inflow (Net, low-skilled German)	806	-5.306452	25.56695	-147	231
Cum. internal migration inflow (Net, Foreign)	806	-.4789082	21.59897	-219	192
Cum. internal migration inflow (Net, high-skilled, Foreign)	806	.2952854	6.984786	-37	92
Cum. internal migration inflow (Net, middle-skilled, Foreign)	806	-.2568238	11.10122	-122	60
Cum. internal migration inflow (Net, low-skilled, Foreign)	806	-.5173697	6.31839	-60	40
Pct. change in unemployment rate (German)	802	1.931488	43.31786	-69.23077	432.5736
Pct. change in unemployment rate (Low-skilled, German)	725	15.07751	109.3914	-100	1151.005
Pct. change in unemployment rate (Middle-skilled, German)	801	2.553877	47.58144	-82.41913	441.1192
Pct. change in unemployment rate (High-skilled, German)	695	7.93704	93.98106	-100	680.1242
Pct. change in unemployment rate (Foreign)	649	3.967753	89.49014	-100	823.1155
Pct. change in unemployment rate (Low-skilled, Foreign)	434	-.4476843	108.3289	-100	726.6667
Pct. change in unemployment rate (Middle-skilled, Foreign)	541	16.86311	110.0161	-100	710.084
Pct. change in unemployment rate (High-skilled, Foreign)	270	-27.60009	89.63057	-100	366.2162
Pct. change in daily wage (High-skilled, Foreign)	793	14.64613	113.0566	-96.5956	2047.75
Pct. change in daily wage (Middle-skilled, Foreign)	804	5.687203	16.97443	-60.68026	229.8272
Pct. change in daily wage (Low-skilled, Foreign)	793	12.62555	48.59167	-85.00084	1003.584
Pct. change in daily wage (High-skilled, German)	805	1.192976	8.268082	-30.43513	41.06249
Pct. change in daily wage (Middle-skilled, German)	805	3.953283	2.949958	-3.737607	56.60109
Pct. change in daily wage (Low-skilled, German)	804	8.007221	10.24891	-30.44722	82.10068
Pct. change in yearly earnings (Low-skilled, Foreign)	793	14.55203	38.61651	-71.60979	481.4861
Pct. change in yearly earnings (Middle-skilled, Foreign)	804	8.324711	50.59907	-65.2952	1329.527
Pct. change in yearly earnings (High-skilled, Foreign)	796	15.21516	144.0415	-89.80944	3590.499
Pct. change in yearly earnings (Low-skilled, German)	804	8.423333	10.37758	-29.51702	87.07076
Pct. change in yearly earnings (Middle-skilled, German)	805	4.245331	2.563048	-3.766606	34.45712
Pct. change in yearly earnings (High-skilled, German)	805	1.567271	8.107592	-31.47783	41.14194
Pct. change in Trade exposure	805	-4.2796	19.16233	-78.67512	417.7786
Pct. change in share in ICT	792	6.509606	43.03759	-100	381.4149
Observations	806				

Notes: The table presents summary statistics of the underlying dataset at the county (NUTS-3) level for the two stacked three-year periods (2014-2016 and 2017-2019). We follow the IAB's definition of immigrants, which is based on citizenship: a person is a foreigner as soon as they do not hold German citizenship. The immigrant inflow is the sum of foreigners' first observation in the SIAB data. Skill levels rely on the imputed educational variable included in the SIAB. The low skilled have neither vocational training nor a university degree. The middle skilled have vocational training and the high skilled hold a degree from a university or university of applied science. AI and AI (IV) are reported in percentage point differences and indicate the share of AI-related skills in demand as a share of all skills in demand in a given county and three-year period cell. Immigrant flows and internal migration flows are cumulative flows. Unemployment rates, daily wages and yearly earnings are reported in percentage changes. Trade exposure is in percentage changes. Source: SIAB, Burning Glass, and Comtrade.

the effect of AI on immigration flows as well as labor market outcomes of foreigners versus natives, we follow the approach by Acemoglu and Restrepo (2018b) and apply a shift-share instrument. More concretely, to account for potential endogeneity concerns, we instrument the AI-related skill demand in Germany with the one in Switzerland. We are confident that both the exclusion and relevance conditions hold for this instrument due to the following reasons.

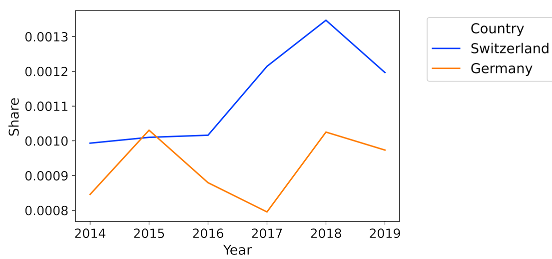
We choose Switzerland as it is the only country among the ten countries, for which we have data for 2014 to 2019, and which is neither part of the European Union nor the European Economic Area. Switzerland therefore follows its own migration policies, at least regarding migration from outside the European Union. In addition, in 2014, the referendum "against mass immigration" took place in Switzerland and was accepted by 50.3 percent of the electorate. The main goal of the referendum was to limit immigration through a quota system. While the Swiss parliament only implemented a softer version of this referendum, it still imposed several restrictions on immigration from European countries. These restrictions are based on employers' obligation to prioritize Swiss over foreign workers in areas with unemployment levels above the average. Furthermore, the Foreign Nationals Act regulates immigration from non-EU/EFTA countries. In this case, admissions are limited to high-skilled individuals and via yearly quotas. Consequently, migration dynamics in Switzerland are unlikely affected by the same potential unobservable drivers which could affect both AI adoption and migration outcomes in Germany. We are therefore confident that AI adoption in Switzerland satisfies the exclusion restriction and is a valid instrument. We test the validity of the instrumental variable in Section 3.8.

Additionally, Switzerland is among the ten leading countries in artificial intelligence worldwide, according to the Nature Index 2021. Figure 3.13 shows the share of AI-related skill demand over time in both countries. From the figure it becomes clear that Switzerland has a higher share of AI-related skill demand than Germany. Figure 3.14 plots the AI-related skill demand at the industry by year level in Germany against the one in Switzerland. There is a positive relationship between AI adoption at the industry by year level in both countries. The figure also reveals that there is significant heterogeneity in AI adoption by industry and year.

Table 3.3 shows the first-stage results of running a regression at the year by industry level. The outcome variable is the AI-related skill demand in Germany and the explanatory variable is the AI-related skill demand in Switzerland. The coefficient reported in the table is positive and significant and the F-statistic is over ten, providing support for the strength and relevance of the instrument.

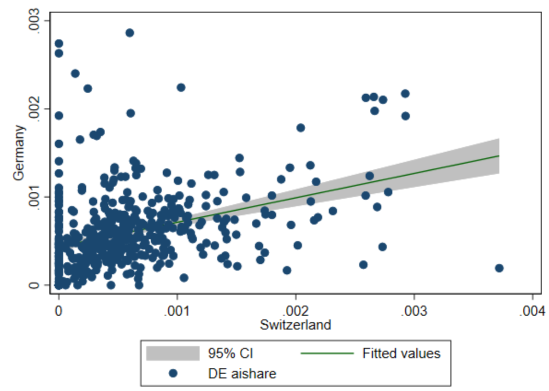
Based on these results, we exploit the local industry structure of labor markets and

Figure 3.13: AI-related skill demand in Germany and Switzerland over time



Notes: The graph plots the AI-related skill demand in Switzerland and Germany by year. AI-related skill demand is reported as a share in all skills in demand in a given year.

Figure 3.14: AI-related skill demand in Germany and Switzerland at the industry by year level



Notes: The graph plots the AI-related skill demand in Switzerland and Germany by industry and year. AI-related skill demand is reported as a share in all skills in demand in a given industry and year. The green line presents the fitted line.

Table 3.3: First-stage: AI-related skill demand by industry

	AI-related skill demand (Germany)
AI-related skill demand (Switzerland)	0.277***
	(0.0317)
Constant	0.000437***
	(0.0000266)
Adj. R-squared	0.137
F-statistic	76.55
N	483

Notes: The table shows the results for a first-stage regression of AI-related skill demand in Switzerland on AI-related skill demand in Germany. The unit of analysis is the year by industry level. Standard errors are in parentheses. Source: BGD data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

construct our shift-share instrument as detailed below:

$$\hat{\mathrm{AI}}_{rj} = \sum_{i\in I}\Big(\frac{\mathrm{emp}_{irt1}}{\mathrm{emp}_{rt1}} \times AI_{it1} - \frac{\mathrm{emp}_{irt0}}{\mathrm{emp}_{rt0}} \times AI_{it0}\Big)\times 100 \text{ , with I=84} \tag{3.4}$$

emp_{irt1} is the number of employees in industry I, German labor market r and year t1. There are 84 different industries. Similarly, emp_{rt1} is the number of employees in a German labor market r and year t1. AI_{it1} is the share of AI-related skill demand in all skills in demand in industry i and year t1 in Switzerland. We then construct the same variable for t0 and subtract both values from each other. Multiplying the resulting difference by 100 gives us the percentage point difference.

We estimate the following regression:

$$Y_{rj} = \alpha X'_{r2014} + \beta_1 \times \hat{\mathrm{AI}}_{rj} + \beta_2 \times \hat{\mathrm{trade}}_{rj} + \varphi_{REGrj} + \epsilon_r \tag{3.5}$$

, where r is a local labor market and j a three-year period (2014-2016 and 2017-2019). $A\hat{I}_{rj}$ is the main explanatory variable of interest, namely the shift-share instrument for AI-related skill demand in a local labor market. $\hat{\mathrm{trade}}_{rj}$ is a shift-share instrument of trade exposure at the local labor market and φ_{REGrj} are regional dummies. X'_{r2014} is a vector of pre-treatment labor market characteristics. We conduct our analysis for the population as a whole, but also for three different skill groups: the high-, medium- and low-skilled workers.

In the case of cumulative net internal migration inflows, percentage changes in wages, and percentage changes in unemployment rates we introduce an interaction term as follows:

$$Y_{rjf} = \alpha X'_{r2014} + \beta_1 \times \hat{\mathrm{AI}}_{rj} + \beta_2 \times \hat{\mathrm{Foreign}}_{rjf} + \beta_3 \times \hat{\mathrm{Foreign}}_{rjf} \times \hat{\mathrm{AI}}_{rj} + \beta_4 \times \hat{\mathrm{trade}}_{rj} + \varphi_{REGrj} + \epsilon_r \tag{3.6}$$

The regression specification above is at the county by 3-year period by foreign citizenship level. $\hat{\mathrm{Foreign}}_{rjf}$ is a dummy variable which is equal to one for foreign citizens, and zero otherwise. β_1 is the coefficient of AI-related skill demand in a local labor market and stacked long-difference period for the native population. $\beta_1 + \beta_3$ is the effect of AI-related skill demand in a local labor market and stacked long-difference period for the foreign population (when the dummy variable is equal to one). Importantly, all variables but the dummy variable for citizenship assume the same value for a given county by 3-year period by foreign citizenship (rjf) combination. They only vary across county by 3-year periods (rj). Therefore, I simplify their subscripts.

3.5 The Effect of AI Adoption

In the following we present our results. We first present our findings on migration inflows from within Germany. Doing so allows us to investigate if AI creates skill shortages which employers cover by recruiting workers with these skills from other internal labor markets. We next report our results on migration inflows from abroad. This analysis can shed light on potential skill shortages in AI-related skills which cannot be covered by the German labor supply. We additionally investigate if migration responses differ by citizenship. Making this distinction is interesting as foreigners might be more willing to migrate internally (Borjas 2001). If they are more likely to move between counties as a response to technological change, they might absorb otherwise negative labor market impacts of automation.

3.5.1 Internal Migration and Potential Skill Shortages

Past research shows that internal migration is one important mechanism in the mitigation of local labor market shocks (Blanchard and Katz 1992). To investigate if this is the case in response to labor market shocks induced by high-skilled task automation, we estimate the impact of AI on net internal migration inflows from other German counties. If individuals migrate in response to AI, we expect to find a positive effect. We choose the county of residence and not the county of individuals' workplace to also capture a movement between counties for the unemployed.

Table 3.4 shows that AI increases the cumulative net internal migration flow for German citizens. A 0.01 percentage point increase in the share of AI-related skill demand in all skill demand increases the cumulative internal migration inflow of German citizens from other counties by 293.7 people. The magnitude is approximately 2.6 standard deviations. Looking at IV coefficients, the increase is largest for the middle-skilled (1.7 standard deviations), followed by the high-skilled (1.5 standard deviations) and low-skilled (1.4 standard deviations). The estimates reported in Table 3.4 are consistent with the hypothesis that people migrate in response to technological shocks.

The evidence presented indicates that cumulative internal migration could serve as a mitigation mechanism of the labor market effects of high-skilled automation technologies. Our findings coincide with previous research showing that internal migration is a mitigation mechanism for immigration inflows (Piyapromdee 2021) or economic downturns (Cadena and Kovak 2016). They are in contrast to work by Faber, Sarto, and Tabellini (2022) showing that robot adoption, which mainly reflects automation of low-skilled tasks, leads to lower in-migration at the commuting-zone level in the US.

Our results should be taken with some caution. Differently from coefficients generated by IV regression estimations, OLS coefficients presented in Table 3.4 are insignificant. The difference could be due to lacking inconsistency in OLS and IV regressions, or because of omitted variables confounding our results from simple linear regressions. Moreover, IV coefficients are up to 100 times larger. These differences emerge from the fact that AI adoption measured via the shift share instrument assumes lower values on average than AI adoption observed directly from online job vacancies (see Table 3.2). Consequently, a one unit change in instrumented AI adoption is a more significant change than a one unit change in observed AI adoption. Nevertheless, the differences in magnitude could also imply that the shift-share instrument has important empirical weaknesses.

The impact on cumulative net internal migration inflows seems to be driven by an increase in internal migration inflows. There is a significant effect on cumulative internal migration inflows, but not cumulative internal migration outflows (see Table C.1 and Table C.2). The positive effect on cumulative internal migration inflows persists when put in relation to the number of employed people in a given county (see Table 3.5). A possible interpretation of our results is that the reinstatement and productivity effect outweigh the displacement effect of high-skilled task automation, leading to economic growth and job creation across the skill distribution. Our findings hint towards technological change having different effects than labor market shocks induced by import competition which leads to lower population growth (Greenland, Lopresti, and McHenry 2019).

Row 3 in Table 3.4 presents the results for those with foreign citizenship. The coefficients in Row 3 of Column 1 and 2 indicate that the cumulative net internal migration flow is lower for foreign citizens compared to natives. The analysis by skill groups illustrates that coefficients on cumulative net internal migration flows is negative for IV regressions. Consequently, the impact of AI on internal migration dynamics differs by citizenship. This is interesting, as it is in contrast to the economic theory (Borjas 2001). AI seems to impact migration decisions by foreigners and natives differently than other external shocks. Cadena and Kovak (2016), for example, find that low-skilled Mexican-born immigrants' location choices respond stronger to changes in local labor demand than natives' location choices. Our results are also relevant for the literature studying if foreigners and natives are substitutes. Borjas (2003) shows that similarly educated foreigners with different levels of experience might not be perfect substitutes for natives from the same skill group. Our evidence is in line with these findings.

We validate our results by estimating a linear probability model at the individual level. For this purpose, we take advantage of the panel-data structure of the SIAB and follow individuals over time. We explore their probability to migrate to counties with a higher AI adoption. We do so by constructing a dummy variable which is equal to one as soon

Table 3.4: AI skill demands and cumulative net internal migration inflows by skill groups and citizenship at the county level

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	-8.898	94.24	-19.06***	12.29	10.59	64.59*	-0.428	17.36*
	(7.987)	(61.10)	(6.197)	(17.04)	(9.302)	(36.43)	(2.816)	(9.314)
AI	1.984	293.7**	0.702	63.58**	0.820	91.33***	0.462	36.40***
	(1.614)	(146.5)	(0.547)	(24.75)	(0.717)	(32.60)	(0.370)	(12.14)
Foreign*AI	-1.658	-233.4**	-0.702	-71.15**	-0.571	-121.9**	-0.384	-40.36**
	(1.319)	(111.4)	(0.529)	(35.81)	(0.515)	(57.36)	(0.302)	(18.60)
Constant	-43.97	-4394.3	107.7	1.837	-131.1***	-23.00	-20.51	-13.39
	(82.76)	(4728.3)	(89.99)	(13.56)	(46.75)	(21.39)	(30.72)	(8.239)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.00886	.	0.0795	.	0.0628	.	0.0283	.
N	1524	1524	1524	1524	1524	1524	1524	1524

Notes: The table presents stacked long-difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period-citizenship cells. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of skills mentioning AI-related expressions in all skills. The cumulative net internal migration inflow is the cumulative inflow into a specific county minus the cumulative outflow for the two 3-year periods investigated. We identify inflows and outflows via changes in the county of residence between different years and then aggregating these changes at the unit of analysis. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Clustered standard errors (at the NUTS-3 level) are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5: AI skill demands and cumulative net internal migration inflows as a share of employed (German) at the county level

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	0.000366	0.0297***	0.0000879	0.00720***	0.000199	0.0166***	0.0000788	0.00591***
	(0.000446)	(0.00971)	(0.000113)	(0.00234)	(0.000245)	(0.00533)	(0.0000920)	(0.00208)
Constant	-0.0475	-0.748	-0.0154	-0.212	-0.0234	-0.415	-0.00865	-0.121
	(0.0373)	(0.705)	(0.0193)	(0.171)	(0.0229)	(0.393)	(0.0147)	(0.143)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0140	.	0.0323	.	0.0263	.	0.105	.
N	762	762	762	762	762	762	762	762

Notes: The table reports the results for a stacked long-difference estimation on the cumulative net internal migration inflow of German citizens. We consider two three year periods, namely 2014-2016 and 2017-2019. The units of analysis are county-period cells. We measure the internal migration inflow by creating a dummy variable which is equal to one if the current county of residence is not equal to the county of residence in the previous year. We then sum all observations at the county level for the periods investigated. In this case, we relate the net internal migration inflow to the number of employed in a respective county. AI skill demands are measured via the share of AI-related skill demand in all skill demand. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population of each county, and the share of young- and old-aged workers. We also create a shift-share instrument for trade exposure, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective SIAB population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

as a county of residence differs from the county of residence observed in the previous year. We establish that AI adoption increases the probability to migrate between counties at the individual level for German citizens (see Table C.3). Table C.3 also reveals that this is not the case for foreign citizens. Consequently, individual-level regressions confirm our results on cumulative net internal migration inflows.[6]

[6]Importantly, the comparison is limited by individual regressions being at the yearly level, while regressions on immigrant inflows rely on stacked long-difference regressions at the county level.

Table 3.6: AI skill demands and cumulative immigrant inflows from abroad by skill groups and at the county level

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	0.478	48.88	0.110	17.19	0.228	20.95	0.139	10.19
	(0.463)	(45.00)	(0.142)	(15.87)	(0.213)	(19.97)	(0.115)	(8.887)
Constant	2147.6*	-565.7	724.3*	-152.4	876.5	-308.0	538.0*	-96.30
	(1208.1)	(1347.9)	(378.9)	(461.3)	(537.3)	(590.5)	(290.4)	(288.4)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.864	.	0.881	.	0.852	.	0.844	.
N	762	762	762	762	762	762	762	762

Notes: The table presents stacked long-difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period cells. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of skills mentioning AI-related expressions in all skills. The outcome variable is the cumulative immigrant inflow into a specific county. We identify the immigrant inflow by aggregating the number of times a foreigner appears for the first time in the SIAB per stacked long-difference. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Clustered standard errors (at the NUTS-3 level) are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.5.2 Immigrant Inflows from abroad and Skill Shortages

We next estimate the impact of AI adoption on immigrant flows from abroad. This question is interesting, as it has important implications for the global competitiveness of firms. If there are skill shortages in the national labor supply, firms might want to recruit the best talent possible from abroad. Their ability to do so might have important implications for productivity, similar to the rationale by Acemoglu et al. (2020) on labor market frictions. If AI adoption indeed leads to skill shortages, which employers cannot cover by recruiting from national labor markets, we expect to find a positive effect on immigrant inflows from abroad.

Table 3.6 indicates that AI adoption has no significant impact on immigrant inflows for none of the skill groups. Column 1 and 2 illustrate that AI adoption does not lead to an increase in the overall cumulative immigrant inflow. All coefficients are insignificant at the common significance levels. The following columns report the results by skill groups. The estimates are positive but none of them is significant. Based on these findings, we conclude that cumulative immigrant inflows remain unaffected by AI-related skill demands at the local labor market level, independently of the educational level.

We confirm our findings by restricting the outcome variable to those immigrants who are immediately employed after their arrival (see Table C.4). Doing so allows us to better control for the 2015 refugee crisis. The 2015 refugee crisis could have generated an external shock to local labor markets. If this shock took place in a similar fashion as AI adoption, these dynamics might confound our main findings. Restricting immigrant inflows to those foreigners who are employed immediately after their arrival serves as a robustness test as refugees are unlikely to be employed immediately after their arrival. The coefficients reported in Table C.4 demonstrate that our main findings seem to be robust to the 2015 refugee crisis.

Table 3.7: AI skill demands and cumulative immigrant inflow (share) by skill groups at the county level

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	-0.00000247	0.000428	-0.00000256	0.000306	-0.000000274	0.000182	0.00000382	-0.0000680
	(0.0000360)	(0.000904)	(0.00000831)	(0.000323)	(0.0000218)	(0.000507)	(0.0000167)	(0.000214)
Constant	0.0614	-0.114**	0.0462***	0.000800	0.00118	-0.0924***	0.0142	-0.0218
	(0.0505)	(0.0494)	(0.0116)	(0.0129)	(0.0295)	(0.0279)	(0.0167)	(0.0154)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.609	0.360	0.747	0.450	0.448	0.214	0.548	0.360
N	762	762	762	762	762	762	762	762

Notes: The table reports the results for a stacked long-difference estimation on the immigrant inflow. We consider two three year periods, namely 2014-2016 and 2017-2019. The units of analysis are county-period cells. We measure the immigrant inflow via the first time a person with a foreign citizenship is observed in the SIAB in a given year and county. In this case, we measure the immigrant inflow in relation to the employed population in a given county. AI skill demands are measured via the share of AI-related skill demand in all skill demand. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population of each county, and the share of young- and old-aged workers. We also create a shift-share instrument for trade exposure, using data from COMTRADE. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective SIAB population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.8: AI skill demands and percentage change in migrant shares by skill groups and at the county level

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	-0.115	-1.270	-0.214	-2.442	-0.121	-1.810	-0.130	-1.243
	(0.0853)	(0.810)	(0.134)	(2.140)	(0.135)	(1.212)	(0.194)	(1.707)
Constant	72.57*	133.2***	40.09	74.51	20.44	109.6	64.77	92.33
	(40.14)	(49.23)	(119.0)	(110.2)	(51.00)	(68.51)	(84.23)	(90.66)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.270	0.0509	0.0573	.	0.169	.	0.126	0.0615
N	762	762	755	755	762	762	753	753

Notes: The table presents stacked long-difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period cells. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of skills mentioning AI-related expressions in all skills. The explanatory variable is the 0.01 percentage change in migrant shares at the county level. A foreigner is by definition of the SIAB a foreign citizen. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Clustered standard errors (at the NUTS-3 level) are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Additionally, we validate our results by measuring immigrant inflows relative to the number of employed in a local labor market (see Table 3.7). The coefficients remain insignificant.

Even when abstracting from newly arrived foreigners and just investigating the impact of AI adoption on migrant shares, there is no significant effect (see Table 3.8). Interestingly, while the coefficients on absolute cumulative immigrant inflows are positive, they are negative in the case of migrant shares. This further supports our findings from Table 3.4 illustrating that there is a decrease in the net internal migration inflow in the case of foreigners. Similar to observations from previous tables, IV coefficients reported in Table 3.6 to Table 3.8 are larger than OLS coefficients. In some cases, the IV estimates are more than 10 times larger than the OLS estimates. These differences are due to differences in the magnitude of directly observed AI adoption and the shift-share instrument (see Table 3.2).

Our findings are in contrast to previous research showing that technological change can lead to an inflow of immigrants (Beerli, Indergand, and Kunz 2021). Furthermore, Hanson

(2021) shows that an increase in the supply of high-skilled immigrants leads to a rise in AI in local labor markets. We demonstrate that this does not apply conversely, meaning that an increase in AI does not lead to a rise in immigrants.

3.6 Mechanisms

There could be several reasons for the diverging effects of AI on net internal migration flows by citizenship. In the following, we explore several possible mechanisms. First, we analyze if foreigners and technologies could be substitutes. Second, we ask if AI adoption creates different productivity effects for foreigners and natives. Third, we test if spillover effects could play a role. Lastly, we investigate the impact of AI adoption on the probability to switch sectors.

3.6.1 Foreigners and Technological Change as Substitutes

Employers might see migrants and technological change as substitutes (Lewis 2011). If this was the case, we would expect to find displacement effects on foreigners. To investigate this further, we analyze the impact of AI adoption on unemployment rates at the local labor market level by skill groups and citizenship. The evidence presented in Table 3.9 speaks against displacement effects. All coefficients presented in Row 2 and 3 are insignificant at the common significance levels. AI adoption does not result in displaced workers, neither among German citizens nor among foreign citizens.

These findings could mean that reinstatement, productivity, complementarity and displacement effects cancel each other out. The fact that there are no significant differences in the impact of AI by citizenship hints against companies perceiving AI adoption and immigrants as substitutes. In fact, the lack in displacement effects could be due to internal migration acting as a mitigation mechanism for technological change. Internal migration would then be a mechanism impeding otherwise negative effects of AI adoption. An alternative explanation for the absence of displacement effects is the still negligible share of AI-related skill demand in all skill demand in local labor markets.

Table 3.9: AI skill demands and percentage change in unemployment rates by skill groups and citizenship at the county level

	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	6.364**	10.78*	1.735**	2.455	3.494***	5.653***
	(2.600)	(5.809)	(0.755)	(1.552)	(0.918)	(2.043)
AI	-0.0122	4.486	0.00509	-0.503	0.0432	0.690
	(0.0467)	(3.448)	(0.0110)	(0.907)	(0.0273)	(1.534)
Foreign*AI	-0.0145	-9.756	0.0465	-1.571	-0.0729	-4.883
	(0.176)	(7.934)	(0.0414)	(2.027)	(0.0624)	(3.115)
Constant	259.2	10.52	-77.90*	5.303***	-33.80	7.462**
	(213.4)	(6.960)	(46.91)	(1.853)	(37.51)	(3.162)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0279	.	0.0346	.	0.0725	.
N	1516	1516	1524	1524	1513	1513

Notes: The table presents stacked long-difference regressions for two periods of three years, 2014-2016 and 2017-2019. The unit of analysis are county-period-citizenship cells. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of skills mentioning AI-related expressions in all skills. The outcome variable is the percentage change in unemployment rates. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as Burning Glass data is only available from 2014 onwards. Clustered standard errors (at the NUTS-3 level) are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.6.2 Potential Diverging Productivity Effects

Another potential driver behind the diverging impact of AI on internal migration flows by citizenship could be a difference in productivity effects. We investigate this possibility by analyzing the effect of AI adoption on wages. Analyzing the development of wages allows us to shed light on potential productivity effects induced by AI adoption. If there is a productivity effect, we expect to see an increase in wages. If productivity effects differ by citizenship, this could explain the diverging impacts on cumulative net migration inflows.

There is no evidence in favor of wage increases as a result of AI adoption. Table 3.10 reports our results on the impact of AI adoption on yearly earnings. The coefficients presented in Row 2 and 3 are insignificant at the common significance levels. The table illustrates that AI adoption does not raise productivity, independently of the education level. Moreover, the coefficients presented raise some concerns about the consistency of IV and OLS estimates. In the case of high-, and middle-skilled workers they go into opposite directions. For the low-skilled, both OLS and IV estimates is positive. In addition, we also observe similar differences in the magnitude of OLS and IV regression coefficients as in previous tables.

We confirm our findings by using daily wages as an alternative measure for productivity. Results from these regressions mostly validate our previous findings on yearly earnings (see Table C.5). In contrast to our findings on yearly earnings, there is some evidence on wage losses among low-skilled foreigners. Row 3 in Column 6 reports a negative and significant coefficient. A 0.01 percentage point increase in the share of AI-related skills in all skills

Table 3.10: AI skill demands and percentage change in yearly earnings by skill groups and citizenship at the county level

	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	6.364**	10.78*	1.735**	2.455	3.494***	5.653***
	(2.600)	(5.809)	(0.755)	(1.552)	(0.918)	(2.043)
AI	-0.0122	4.486	0.00509	-0.503	0.0432	0.690
	(0.0467)	(3.448)	(0.0110)	(0.907)	(0.0273)	(1.534)
Foreign*AI	-0.0145	-9.756	0.0465	-1.571	-0.0729	-4.883
	(0.176)	(7.934)	(0.0414)	(2.027)	(0.0624)	(3.115)
Constant	259.2	10.52	-77.90*	5.303***	-33.80	7.462**
	(213.4)	(6.960)	(46.91)	(1.853)	(37.51)	(3.162)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0279	.	0.0346	.	0.0725	.
N	1516	1516	1524	1524	1513	1513

Notes: The table presents stacked long-difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period-citizenship cells. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of AI-related skills in the total number of skills. The outcome variable is the percentage change in yearly earnings. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. Skill levels are based on an imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to information and communication technology as well as trade, using data from Eurostat as well as Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as Burning Glass data is only available from 2014 onwards. Clustered standard errors (at the NUTS-3 level) are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

demanded in a local labor market leads to a decrease in daily wages of 5.6 percent. This coefficient is only significant at the 10 percent significance level. Combining this result with evidence from Table 3.10 showing insignificant effects on yearly earnings, we conclude that a significant impact of AI adoption on low-skilled foreigners' productivity levels is unlikely.

Consequently, neither displacement nor productivity effects seem to drive the different response in internal migration of foreigners and natives to AI adoption. Our findings are in line with findings from the US, where AI did not lead to any significant aggregated labor market effects (Acemoglu et al. 2020). The authors show that there are neither productivity nor complementarity effects of high-skilled task automation yet. This result might be mainly due to the yet low share of AI-related skills in all skills in demand in local labor markets.

3.6.3 Spillover Effects

We next examine if spillover effects between the most and least exposed industries could play a role in our results on migration dynamics. We analyze the effect of AI on the most exposed sectors, namely the ICT sector, public administration and defense, mining and quarrying as well as professional, scientific and technical activities. The results presented in Appendix C.3 discard that spillover effects could play a role in our results. The findings for the most- and least-exposed sectors are in line with our main findings for the full sample.

3.6.4 The Probability of Switching Sectors

We next explore whether the adverse effects on foreigners' and natives' net internal migration flows are due to them being less adaptive than natives. We measure the likelihood to adapt to AI adoption via the probability to switch economic sectors. For this purpose, we follow individuals registered in the SIAB over time and create a dummy variable which is equal to one as soon as an individual switches economic sectors. We then take advantage of the panel-data structure of the SIAB and run individual level regressions.

We find that foreigners are less likely to switch sectors as a response to AI adoption than natives, but the effect is only significant for the high- and middle-skilled and IV regressions (see Column 4 and 6 in Table 3.11). Importantly, coefficients are only significant at the 10 percent significance level. The lower probability of switching sectors could drive the diverging impact on net internal migration flows by citizenship. Still, this interpretation should be taken with caution, given that there are inconsistencies in IV and OLS estimates for middle-skilled foreigners. The sign of these coefficients go into opposite directions. In addition, there is an increase in the probability of switching sectors for low-skilled foreigners (see Columns 7 and 8). Furthermore, the probability of switching sectors also decreases for middle-skilled German citizens (see Column 5 and 6). This pattern of results indicates that there might be other drivers leading to the observed differences in internal migration patterns as a response to AI adoption.

Table 3.11: AI-related skill demands and the probability to switch sectors

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	0.0563***	-0.0248	0.0714***	0.187***	0.0593***	0.134***	-0.00941**	-0.0802
	(0.00177)	(0.0641)	(0.00458)	(0.0334)	(0.00171)	(0.0287)	(0.00310)	(0.0644)
AI	-0.00688***	-0.802**	0.00273	0.0917	-0.00933***	-1.058***	-0.0106**	-0.666
	(0.00169)	(0.258)	(0.00335)	(0.266)	(0.00173)	(0.255)	(0.00327)	(0.398)
Foreign*AI	0.0190***	1.037	-0.00462	-0.677*	0.0151	-2.910*	0.0284***	1.662
	(0.00465)	(0.991)	(0.00898)	(0.273)	(0.0104)	(1.310)	(0.00773)	(1.457)
Constant	-0.735***	0.361***	-0.899***	0.267***	-0.676***	0.355***	-0.734***	0.402***
	(0.154)	(0.0247)	(0.239)	(0.0403)	(0.134)	(0.0186)	(0.155)	(0.0288)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0951	.	0.121	.	0.0951	.	0.0709	.
N	9865642	4177551	1557376	757587	6958648	2885919	1349617	534044

Notes: The table presents yearly regressions for the period 2014-2019. Regressions are at the individual by year level. The explanatory variable is AI adoption. AI adoption is measured via the share of job skills mentioning AI-related expressions in all skill demand. The outcome variable is a dummy variable equal to one if there is a difference in the economic sector between years. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The skill level is based on the imputed educational variable included in the SIAB. The low skilled have neither vocational training nor a university degree. The middle skilled have vocational training and the high skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

There could be several reasons for a lower likelihood to switch sectors in the case of foreign citizens, such as language barriers, or less access to relevant networks or labor market institutions. Work by Martén, Hainmueller, and Hangartner (2019) demonstrates the importance of social networks for integrating refugees into the economy. Similarly, Lochmann, Rapoport, and Speciale (2019) provide evidence of the positive effect of language training on labor force participation. Consequently, language barriers could explain why foreigners are less reactive to technological change than natives. Further research could explore this point in more detail.

3.7 Comparison to Low-Skilled Task Automation

We next analyze to which extent the skill level automation is most likely to replace plays a role in our results. A comparison of this type can shed further light on the impact of automation across the skill distribution. We measure low-skilled task automation via the operational stock of industrial robots. For this purpose, we leverage data from the Industrial Federation of Robotics (IFR) and apply a shift-share instrument similarly to the one detailed in Section 3.4. For the details behind the empirical methodology see Appendix C.5.

Table 3.12 presents the results on the impact of AI adoption on cumulative internal migration inflows. The coefficients in Row 2 report our findings for native citizens. There is a decrease in the net internal migration inflow. The coefficient is significant at the 10 percent significance level for the high-skilled (Column 4) and at the 5 percent significance level in the case of middle- and low-skilled (Column 6 and 8). Similar to our findings on the impact of AI adoption, significance only holds for IV but not OLS estimations. Row 3 presents the results for foreign citizens. In contrast to our findings on the native population, cumulative net internal migration inflows seem to increase in response to robot exposure. The estimates reported in Row 3 are significant at the 10 percent significance level for IV regressions and insignificant for OLS regressions. Based on this result, there is some slight evidence in favor of migration responses to automation of low-skilled tasks. The dynamics are different from the ones observed on high-skilled task automation.

Table 3.13 reports the results from regressions on the impact of robot exposure on cumulative immigrant inflows. We first present coefficients from running simple linear regressions, and then the ones from our instrumental variable specification. We start by showing results for all immigrants and then for high-, middle- and low-skilled immigrants. The table illustrates that an increase in robot exposure does not lead to a significant increase in immigrant inflows at the county level. This is the case for all skill groups. From this, one could conclude that the installment of industrial robots does not lead to skill shortages, which firms cover by recruiting these skills from abroad. These results are in line with our findings on AI adoption. Consequently, skill types of automation technologies do not seem to play a significant role in its impact on immigration inflows from abroad.

We next investigate if robot adoption affects labor market outcomes of foreigners differ-

Table 3.12: Robot exposure and cumulative net internal migration inflows by skill type and citizenship at the county level

	All (OLS)	All (IV)	High-skilled (OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	-322.9***	-384.8***	-168.0***	-196.9***	-96.59***	-113.2***	-58.33	-74.67*
	(123.8)	(138.2)	(59.11)	(65.33)	(26.34)	(33.14)	(42.86)	(44.73)
Robot exposure (Op. Stock)	-43.54	-629.7	-16.52	-119.7*	-9.056	-101.1**	-17.97	-104.8**
	(105.4)	(424.9)	(53.97)	(69.44)	(22.43)	(47.50)	(36.11)	(48.14)
Foreign*Robots	185.2	519.6*	71.01	227.1*	41.13	131.1*	73.06	161.4*
	(187.6)	(281.1)	(91.48)	(129.2)	(37.59)	(78.15)	(66.03)	(83.87)
Constant	2757.3**	1688.3	1466.6**	119.2***	739.4***	110.9***	551.3**	22.59
	(1132.4)	(1095.9)	(725.8)	(33.54)	(251.6)	(18.48)	(262.2)	(26.18)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.524	0.490	0.523	0.408	0.501	0.429	0.469	0.372
N	802	802	802	802	802	802	802	802

Notes: The table shows long-difference estimations for the effect of robot adoption on net internal migration inflows for the period 2005-2018. The unit of analysis are county-citizenship cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. Robot exposure (Op. Stock) is the exposure of a county to industrial robots measured via the shift share instrument. The net internal migration inflow is the cumulative internal migration inflow minus the cumulative internal migration outflow into a respective county during the period 2005 and 2018. We identify internal migration flows via changes in counties of residence between different year and then aggregating all observations for the period under consideration at the county level. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to information and communication technology as well as trade, using data from EUROSTAT as well as COMTRADE. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. Source: IFR Robotics data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.13: Robot exposure and cumulative immigrant inflow by skill type and citizenship at the county level

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Robot exposure (Op. Stock)	-10.23	-1763.7	-4.626	-464.1	-9.490	-732.5	4.041	-550.8
	(204.0)	(1749.5)	(46.73)	(445.8)	(91.33)	(728.9)	(64.67)	(559.1)
Constant	10867.4**	6264.1	2236.6*	1030.4	4955.9**	3057.9	3583.5**	2127.0
	(5101.0)	(4395.9)	(1263.8)	(1083.6)	(2204.9)	(1923.7)	(1599.2)	(1372.9)
Mean (Dep. Var)	1003.3	1003.3	214.2	214.2	502.1	502.1	269.7	269.7
St. Dv. (Dep. Var.)	1713.4	1713.4	423.1	423.1	854.5	854.5	405.4	405.4
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.908	0.873	0.915	0.876	0.924	0.900	0.842	0.779
N	401	401	401	401	401	401	401	401

Notes: The table shows long-difference estimations for the effect of robot adoption on immigrant inflows from abroad for the period 2005-2018. Robot exposure (Op. Stock) is the exposure of a county to industrial robots measured via the shift share instrument. The unit of observation is the county (NUTS-3). The immigrant inflow is the cumulative immigrant inflow into a respective county during the period 2005 and 2018. We identify the immigrant inflow by aggregating the number of times a migrant appears for the first time in the SIAB per year. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to information and communication technology as well as trade, using data from EUROSTAT as well as COMTRADE. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. Source: IFR Robotics data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

ently from natives. To do so, we interact the robot adoption in a respective county with a dummy variable which is equal to one if the outcome variable refers to the foreign population, and zero otherwise. Table 3.14 shows that robot adoption significantly decreases the unemployment rate of medium-skilled citizens when running simple linear regression (see Column 5). The observed decrease in the unemployment rate of middle-skilled foreign citizens could be explained by skill complementarities of technological change. This means that the adoption of robots creates new tasks, such as their supervision or operation. Our results suggest that this task creation has positive effects on middle-skilled foreign citizens' employment share. Still, the effect becomes insignificant when applying a shift-share instrument (see Column 6). The largely insignificant overall effects on unemployment in Germany are in line with findings by Dauth et al. (2021). When compared to the impact of AI, results are in line. Neither of the two automation technologies investigated affect unemployment rates.

Table 3.14: Robot exposure and percentage change in unemployment rates by skill type and citizenship at the county level

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	-1.483	-1.438	31.59	39.21	47.08***	48.75***	12.20	11.92
	(2.107)	(2.744)	(19.98)	(25.94)	(6.540)	(8.175)	(6.333)	(7.336)
Robot exposure (Op. Stock)	0.132	9.387	2.194	-68.93	-13.48	-20.81	16.41	-10.15
	(4.298)	(12.47)	(22.30)	(100.6)	(11.11)	(35.36)	(13.27)	(32.66)
Foreign*Robots	-6.360	-6.523	36.90	-14.17	-25.74**	-34.68	-26.06	-25.51
	(4.170)	(9.843)	(71.75)	(74.68)	(9.789)	(21.44)	(21.75)	(34.66)
Constant	74.75	100.0	126.8	30.67	-191.1	-223.6	693.9**	636.3**
	(81.55)	(86.02)	(304.6)	(306.5)	(249.2)	(285.1)	(212.2)	(228.2)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.107	0.102	0.179	0.154	0.164	0.162	0.0906	0.0838
N	727	727	431	431	688	688	642	642

Notes: The table shows long-difference estimations for the effect of robot adoption on immigrant inflows from abroad for the period 2005-2018. Robot exposure (Op. Stock) is the exposure of a county to industrial robots measured via the shift share instrument. The unit of observation is the county-citizen combination. This means that there are two lines of data for each county, one showing the average change in the unemployment rate for foreign citizens in the manufacturing sector in a respective county during the period 2005 and 2018, and the other one showing the average change in the unemployment rate for natives in the manufacturing sector. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. Source: IFR Robotics data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.15 shows that robots have adverse effects on wages by citizenship. While robot adoption increases the wage of German citizens for all skill groups, it decreases it for foreign citizens of all skill groups. These results could be driven by increases in labor productivity (see Graetz and Michaels 2018 or Acemoglu and Restrepo 2018b), which then result in the observed wage increase for German citizens. Interestingly, high-skilled natives do not benefit from wage increases. These differences could be due to low-skilled task automation having limited relevance for high-skilled tasks. Low- and middle-skilled workers might be more affected by potentially positive spill-over effects resulting from the application of robots.

Our results suggest that foreign citizens, on the other hand, do not benefit from the same wage increases as natives. The coefficients reported in Row 3 of Table 3.15 are negative and significant across the board. There could be several reasons for these wage decreases faced by foreigners. First, foreign citizens could have less access to information about the need to adapt their skill-set as a response to technological change. This result could be due to language barriers, or less access to local networks (Martén, Hainmueller, and Hangartner (2019); Lochmann, Rapoport, and Speciale (2019)). Others demonstrate that there are discriminatory effects in job applications as a response to headscarves, for example (Weichselbaumer 2016). These discriminatory effects could explain some of the negative impact of robots on foreign citizens' wages. Future research should study potential drivers in more detail.

Additionally, even without considering technological change, scholars have shown that immigrants are subject to downskilling, also in Germany (Elsner and Zimmermann 2016). Technological change could worsen this trend. Moreover, firms might consider foreign citizens as cheap alternatives to local labor costs (Walia 2010). The same applies to robots (Acemoglu and Restrepo 2018b). The increasing adoption of robots could then lead to increased competition between foreign citizens and robots, driving the observed decrease in wages for foreign citizens due to robot adoption.

Table 3.15: Robot exposure and percentage change in daily wage by skill type and citizenship at the county level

	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	10.76*	13.72**	-0.631	0.283	7.477*	11.75**
	(5.902)	(5.749)	(3.117)	(3.633)	(4.178)	(5.568)
Robot exposure (Op. Stock)	8.950	-1.816	12.73***	9.492**	22.61***	12.79*
	(7.759)	(11.70)	(4.665)	(4.816)	(5.780)	(7.307)
Foreign*Robots	-21.29***	-41.31**	-17.15***	-22.10**	-27.89***	-50.72***
	(6.894)	(16.66)	(5.285)	(10.84)	(6.812)	(17.53)
Constant	22.74	15.16	6.350	-0.118	-239.1	5.384
	(216.3)	(15.63)	(97.33)	(3.075)	(171.6)	(4.797)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0569	0.0198	0.0961	0.0897	0.253	0.235
N	741	741	796	796	782	782

Notes: The table shows long-difference estimations for the effect of robot adoption on immigrant inflows from abroad for the period 2005-2018. The table shows long-difference estimations for the effect of robot adoption on percentage changes in daily wages for the period 2005-2018. Robot exposure (Op. Stock) is the exposure of a county to industrial robots measured via the shift share instrument. The unit of observation is the county-citizen combination. This means that there are two lines of data for each county, one showing the average percentage change of daily wages for foreign citizens in a respective county during the period 2005 and 2018, and the other one showing the average percentage change of daily wages for German citizens. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. Source: IFR Robotics data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Our results indicate that the skill types automation technologies are likely to replace do not seem to play a role in displacement effects nor immigrant inflows from abroad. Neither high-skilled nor low-skilled task automation significantly impact these outcomes. In contrast, we demonstrate systematic differences in the impact of automation by skill type on wages. This hints towards varying effects on the productivity of workers along the skill distribution.[7]

[7]There could be other reasons for the diverging impact, such as the fact that AI adoption is lower in

Lastly, while both the automation of high- and low-skilled tasks affect internal migration dynamics, the effects go in opposite directions. The automation of high-skilled tasks increases cumulative net internal migration inflows for natives, but decreases it for foreigners. The opposite is true for the automation of low-skilled task automation. Consequently, foreigners might react to negative labor market shocks but are unable to mitigate them, as they experience negative wage effects. Still, their migration dynamics might potentially mitigate otherwise negative effects on natives. Future research should analyze this possibility in more detail.

3.8 Robustness Checks

3.8.1 Testing the Validity of the Shift-Share Instrument

Our findings rely on the validity of the shift-share instrument at use in this paper. It is crucial to analyze the robustness of our empirical strategy to the underlying assumptions of the shift share instrument. There are currently two main contributions on the robustness of shift share instruments. First of all, Goldsmith-Pinkham, Sorkin, and Swift (2020) make the case that the instrument requires exogeneity in industry shares at the local labor market level. Next, Borusyak, Hull, and Jaravel (2022) stress that the shift-share instrument relies on exogeneity in shocks. In the following, we investigate these assumptions in more detail. We abstract from an analysis of parallel trends as proposed by Goldsmith-Pinkham, Sorkin, and Swift (2020) given that AI adoption did take place gradually and not via a sudden shock.

Exogeneity of Initial Shares

Our research design relies on several important assumptions. First, initial shares should be exogenous to changes in the error terms (Goldsmith-Pinkham, Sorkin, and Swift 2020). A potential identification threat would then be that initial shares $\frac{\text{emp}_{ir}}{\text{emp}_r}$ are directly correlated with changes in outcome variables (e.g. the cumulative net internal migration inflow during the period 2014 to 2019). This is a concern as the Bartik instrument is the equivalent of using initial shares as multiple instruments in a weighted generalized methods of moment estimation. Consequently, one can understand initial shares as instruments, and industry growth rates as weights. If initial shares were correlated with changes in the outcome variable, this would violate the exclusion restriction of instrumental variables. Exclusion restrictions state that instrumental variables only affect outcome variables indirectly through the treatment

terms of magnitude than robot adoption. Impacts might also vary given the different time periods under consideration for both technologies.

variable. One example of a potential violation in the setting studied in this paper would be if the initial employment structure is correlated with the unobservable tendency to discriminate against foreign citizens, e.g. through being a more conservative county. This initial tendency to discriminate against foreign citizens could then affect cumulative net internal migration inflows and bias our estimates.

We test if our research design satisfies the exclusion restriction of industry shares. We follow Goldsmith-Pinkham, Sorkin, and Swift (2020) and test if industry shares in 2014 are correlated with average county characteristics in 2014. Table 3.16 shows the results. The pattern of results illustrates that most of the industry shares are significantly correlated with at least one control variable. To give an example, the third aggregated manufacturing sector, which covers the vehicle manufacturing sector, is concentrated in counties with a larger share of high-educated workers. This pattern emphasizes that other trends that could potentially affect cumulative net internal migration flows in the aggregated manufacturing sector in more educated areas could confound our results. Similar patterns emerge when correlating the industry shares in 2014 with percentage changes in control variables over the period 2014-2019 (see Table 3.17).

Table 3.16: Relationship between industry shares and initial county characteristics

	Primary Sector	Manufacturing Sector (1)	Manufacturing Sector (2)	Manufacturing Sector (3)	Construction	Education	Other service activities
Female Share	-0.0838***	0.161***	-0.0106	-0.102	-0.00315	0.0114	0.745***
	(0.0271)	(0.0477)	(0.0669)	(0.0751)	(0.0274)	(0.0252)	(0.106)
Part-time Share	-0.0607***	-0.107***	0.00358	0.0290	-0.0925***	0.0570***	-0.955***
	(0.0156)	(0.0274)	(0.0384)	(0.0431)	(0.0158)	(0.0145)	(0.0608)
Migrant Share	0.0155	-0.0520	0.0331	0.0289	-0.0105	-0.0785***	0.199**
	(0.0198)	(0.0349)	(0.0489)	(0.0549)	(0.0201)	(0.0184)	(0.0773)
Share of low-skilled	0.0716***	0.176***	0.115**	-0.306***	0.0698***	-0.244***	-0.108
	(0.0235)	(0.0413)	(0.0580)	(0.0651)	(0.0238)	(0.0218)	(0.0917)
Share of middle-skilled	-0.00801	0.143**	0.333***	-0.453***	-0.0653*	-0.229***	0.602***
	(0.0352)	(0.0620)	(0.0869)	(0.0975)	(0.0356)	(0.0328)	(0.137)
Share of <35	-0.127***	0.0876	-0.308***	0.202*	-0.138***	0.196***	-0.179
	(0.0379)	(0.0667)	(0.0935)	(0.105)	(0.0384)	(0.0353)	(0.148)
Share of >54	0.0281	-0.0714	0.129	-0.0808	-0.0887	0.249***	-0.509**
	(0.0554)	(0.0975)	(0.137)	(0.153)	(0.0561)	(0.0515)	(0.216)
Share of leased	-0.0288	-0.237***	0.0389	0.254*	-0.114**	0.209***	0.715***
	(0.0475)	(0.0837)	(0.117)	(0.132)	(0.0481)	(0.0442)	(0.186)
Share of temporary	0.0851***	0.0348	0.134**	-0.127*	0.0605**	0.182***	0.252**
	(0.0264)	(0.0465)	(0.0653)	(0.0733)	(0.0267)	(0.0246)	(0.103)
Constant	0.116***	-0.159**	-0.0640	0.245**	0.154***	0.0980***	0.958***
	(0.0355)	(0.0626)	(0.0877)	(0.0985)	(0.0360)	(0.0331)	(0.139)
R-squared	0.579	0.580	0.634	0.700	0.723	0.734	0.852
N	329	329	328	328	329	329	329

Notes: Each column reports results of a single regression of a 2014 industry share on 2014 county characteristics. Each column presents a separate regression of the industry shares in a respective county and the controls presented in the table. *Manufacturing sector (1)* refers to the manufacture of food products, manufacture of beverages, retail sale of food, beverages and tobacco in specialised stores, wood and wood products, as well as other manufacturing. *Manufacturing sector (2)* refers to the manufacture of coke and refined petroleum products, chemical and pharmaceutical products, as well as rubber and plastics, and of basic metals and fabricated metal products. *Manufacturing sector (3)* refers to the manufacture of computer, electronic and optical products, electrical equipment, mechanical engineering, and vehicle manufacturing. We weight each county by its population in 2014. Standard errors are in parentheses. Source: SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.17: Relationship between industry shares and percentage changes in county characteristics

	Primary Sector	Manufacturing Sector (1)	Manufacturing Sector (2)	Manufacturing Sector (3)	Construction	Education	Other service activities
Female share	0.0735**	0.00709	0.0324	0.202	0.0308	0.0801*	-0.0896
	(0.0361)	(0.0633)	(0.102)	(0.138)	(0.0390)	(0.0443)	(0.217)
Part-time share	0.0235	-0.0327	0.0389	-0.0450	0.0175	0.00798	0.165
	(0.0225)	(0.0394)	(0.0636)	(0.0856)	(0.0243)	(0.0276)	(0.135)
Migrant share	0.0210***	0.0105*	-0.0180*	-0.0396***	0.0216***	0.00666	0.0391*
	(0.00351)	(0.00616)	(0.00994)	(0.0134)	(0.00380)	(0.00431)	(0.0211)
Share of low-skilled	0.145***	0.461***	0.623***	0.484***	0.322***	-0.201***	-2.445***
	(0.0338)	(0.0592)	(0.0955)	(0.129)	(0.0365)	(0.0415)	(0.203)
Share of middle-skilled	0.0331***	0.124***	0.135***	0.0778*	0.0913***	-0.0839***	-0.533***
	(0.0122)	(0.0214)	(0.0346)	(0.0466)	(0.0132)	(0.0150)	(0.0732)
Share of high-skilled	0.00449	0.0605**	0.0414	0.153***	0.0483***	-0.0710***	-0.437***
	(0.0135)	(0.0237)	(0.0383)	(0.0516)	(0.0146)	(0.0166)	(0.0812)
Share of <35	-0.202**	-0.179	-0.681**	-0.654*	-0.260**	0.0352	0.737
	(0.0960)	(0.168)	(0.271)	(0.365)	(0.104)	(0.118)	(0.575)
Share of 35-54	-0.234**	-0.191	-0.850**	-0.696	-0.315**	0.0149	1.013
	(0.118)	(0.207)	(0.334)	(0.450)	(0.128)	(0.145)	(0.708)
Share of >54	-0.0571	-0.0476	-0.279***	-0.228*	-0.0859**	-0.0292	0.109
	(0.0361)	(0.0632)	(0.102)	(0.137)	(0.0390)	(0.0443)	(0.216)
Share of leased employees	0.000337	-0.00351	-0.00519	-0.00733	-0.000399	0.000539	0.00872
	(0.00230)	(0.00403)	(0.00650)	(0.00876)	(0.00249)	(0.00282)	(0.0138)
Share of temporary workers	0.0245**	0.0468**	0.0970***	0.0802*	0.0203*	-0.0565***	-0.127*
	(0.0113)	(0.0198)	(0.0321)	(0.0433)	(0.0122)	(0.0138)	(0.0676)
Share in ICT	0.0120	0.0325*	0.0241	0.146***	0.0208*	0.0172	-0.0295
	(0.00987)	(0.0173)	(0.0279)	(0.0375)	(0.0107)	(0.0121)	(0.0592)
Constant	0.0199***	0.0473***	0.0750***	0.0824***	0.0462***	0.0510***	0.665***
	(0.00326)	(0.00570)	(0.00920)	(0.0124)	(0.00352)	(0.00399)	(0.0195)
R-squared	0.356	0.365	0.268	0.135	0.519	0.292	0.465
N	329	329	328	328	329	329	329

Notes: The table shows the relationship between initial industry shares in 2014 and percentage changes in average county characteristics for the period 2014-2019. Each Column represents a separate regression of the industry shares in a respective county and the controls presented in the table. The *Manufacturing (1)* sector refers to the manufacture of food products, manufacture of beverages, retail sale of food, beverages and tobacco in specialised stores, wood and wood products, as well as other manufacturing. The *Manufacturing (2)* sector refers to the manufacture of coke and refined petroleum products, chemical and pharmaceutical products, as well as rubber and plastics, and of basic metals and fabricated metal products. The *Manufacturing (3)* sector refers to the manufacture of computer, electronic and optical products, electrical equipment, mechanical engineering, and vehicle manufacturing. We weight each county by its population in 2014. Standard errors are in parentheses. Source: SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Exogeneity of Growth Rates

Borusyak, Hull, and Jaravel (2022) apply a different train of thoughts to the shift-share instrument. They approach the validity of the shift-share instrument via the shock component of the instrument. Their argument relies on a slightly different interpretation of shift-share instruments, namely that it is the shocks that instrument for the aggregate treatment while exposure shares are weights. The exclusion restriction then relies on shocks being exogenous to underlying unobservable county characteristics driving both AI adoption and migration dynamics. Put differently, instrumented AI adoption at the county level should be as good as random.

To investigate the exogeneity of our instrument, we apply a placebo regression. We regress AI adoption at the county level for the period 2014-2016 and 2017-2019 on outcomes for the period 1980-1990 and 1991-2000. Table 3.18 demonstrates that there is no significant effect of AI adoption on past cumulative net internal migration inflows. The same is true for wages (see Table 3.20) and unemployment rates (see Table 3.21). Still, there is some slight evidence in favor of a significant impact of AI adoption on immigrant inflows for the period 1980-1990 and 1991-2000. The coefficients presented in Table 3.19 are significant at the 10 percent significance level in case of the high- and low-skilled. Importantly, coefficients are only significant for OLS regressions. The IV coefficients, on the other hand, are insignificant. Based on these results, we conclude that IV coefficients are more robust to unobservable confounding factors.

In general, given that our placebo regressions reveal that the shock component of our shift-share instrument is likely exogenous, exogeneity in shares is not necessary (Borusyak, Hull, and Jaravel 2022). Consequently, we can rely on our shift-share instrument for the research question at hand.

Table 3.18: AI skill demands and cumulative net internal migration inflows by skill groups and citizenship at the county level (placebo regressions)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	-1.954	-120.4	-1.043	-47.10	-1.622*	-81.17	0.710	7.832
	(1.547)	(1054.3)	(0.633)	(409.4)	(0.954)	(707.9)	(0.628)	(69.37)
AI	-0.0430	-56.90	0.00646	-22.52	-0.0516**	-39.25	0.00214	3.413
	(0.0300)	(511.4)	(0.00691)	(198.4)	(0.0222)	(343.0)	(0.00912)	(33.62)
Foreign*AI	0.0519	116.5	-0.00675	45.26	0.0631**	78.25	-0.00446	-7.004
	(0.0348)	(1021.3)	(0.00905)	(396.8)	(0.0252)	(685.8)	(0.0114)	(67.29)
Constant	42.93	128.7	22.11	23.88	7.518	40.55	13.30	-5.203
	(46.07)	(542.6)	(14.38)	(204.7)	(35.70)	(354.0)	(11.88)	(34.68)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.226	.	0.181	.	0.220	.	0.107	.
N	610	610	610	610	610	610	610	610

Notes: The table presents stacked long-difference regressions for two periods of ten years, 1980-1990 and 1991-2000. The unit of analysis are the county-period-citizen cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. We measure AI skill demands via the share of AI-related skill demand in all skill demand in a local labor market. The immigrant inflow is the cumulative inflow into a specific county. We identify the immigrant inflow by aggregating the number of times a foreigner appears for the first time in the SIAB per year. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Clustered standard errors (at the NUTS-3 level) are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.19: AI skill demands and cumulative immigrant inflows from abroad by skill groups and at the county level (placebo regressions)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	1.335	-28.81	0.103*	-4.323	0.439	-11.49	0.773*	-11.51
	(0.809)	(55.48)	(0.0613)	(5.755)	(0.286)	(22.24)	(0.458)	(27.22)
Constant	9664.3***	4768.8	1076.4***	492.4	3247.7***	1199.6	5230.1***	3057.5
	(3234.9)	(3947.5)	(302.7)	(461.0)	(1159.9)	(1493.8)	(1762.5)	(1982.9)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.798	0.438	0.861	0.195	0.812	0.370	0.770	0.528
N	305	305	305	305	305	305	305	305

Notes: The table presents stacked long-difference regressions for two periods of ten years, 1980-1990 and 1991-2000. The units of analysis are county-period cells. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of job skills mentioning AI-related expressions in all skill demand. The outcome variable is the cumulative immigrant inflow into a specific county. We identify the immigrant inflow by aggregating the number of times a foreigner appears for the first time in the SIAB per stacked-long difference. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Clustered standard errors (at the NUTS-3 level) are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.20: AI skill demands and percentage change in daily wages by skill groups and citizenship at the county level (placebo regressions)

	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	0.167	6.770	-0.228	-71.84	7.937***	-78.07
	(2.778)	(30.88)	(0.661)	(610.0)	(0.932)	(765.5)
AI	0.0490	2.903	0.00329	-35.08	-0.0156	-41.82
	(0.0344)	(15.37)	(0.00962)	(295.8)	(0.0258)	(371.9)
Foreign*AI	0.0941	-6.376	0.00440	70.51	-0.0236	84.65
	(0.103)	(30.22)	(0.0939)	(591.4)	(0.0626)	(742.9)
Constant	109.3	-14.61	50.01	23.98	16.78	15.02
	(161.0)	(14.85)	(32.33)	(305.1)	(40.41)	(383.0)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0215	.	0.0924	.	0.219	.
N	535	535	609	609	608	608

Notes: The table presents stacked long-difference regressions for two periods of ten years, 1980-1990 and 1991-2000. The unit of analysis are the county-period-citizen cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of AI-related skills in the total number of skills. The outcome variable is the percentage change in daily wages. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to information and communication technology as well as trade, using data from Eurostat as well as Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Clustered standard errors (at the NUTS-3 level) are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.21: AI skill demands and percentage change unemployment rates by skill groups and citizenship at the county level (placebo regressions)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	5.665***	268.3	-3.554***	-5.300	-4.895***	-372.9	-3.041***	5.280
	(0.292)	(57823.4)	(0.500)	(6.718)	(0.360)	(31009.2)	(0.468)	(87.55)
AI	-0.0112	127.3	0.0679	-1.193	-0.00459	-116.2	0.00917	1.965
	(0.00684)	(27724.1)	(0.0491)	(5.366)	(0.00497)	(9820.8)	(0.00614)	(33.35)
Foreign*AI	0.0334**	-240.3	-0.0665	2.876	-0.0276	276.2	-0.0173**	-7.773
	(0.0150)	(52850.0)	(0.0489)	(13.96)	(0.0280)	(23284.5)	(0.00804)	(80.61)
Constant	-104.7***	1445.3	-76.92***	-167.9	-84.56***	147.5	-75.66***	-18.84
	(14.22)	(338487.4)	(11.02)	(474.8)	(5.690)	(18636.4)	(14.61)	(695.8)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.684	.	0.587	.	0.781	.	0.515	.
N	537	537	235	235	486	486	474	474

Notes: The table presents stacked long-difference regressions for two periods of ten years, 1980-1990 and 1991-2000. The unit of analysis are the county-period-citizen cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of job vacancies mentioning AI-related expressions in the total number of job vacancies. The outcome variable is the percentage change in unemployment rates. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Clustered standard errors (at the NUTS-3 level) are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.8.2 Robustness to Urban-Rural Migration Dynamics

As many countries, Germany has been subject to urbanization during the last decades. While back in 1990, 73 percent of the population lived in urban areas, this share has increased to 77 percent in 2019 (The World Bank 2022). In addition, AI adoption might mainly take place in cities and urban areas. Although we are less worried about urban-rural migration dynamics confounding our results, given that urbanization has stagnated at 77 percent since

2011 and therefore overall not changed during the period under consideration, we analyze the robustness of our findings to urbanization. We do so by investigating the impact of AI adoption on migration dynamics restricting our sample to urban counties.

We find that our results are robust to restricting the sample to urban counties. Table 3.22 reveals that there is a significant increase in the cumulative net internal migration inflow among urban areas, and that this increase is lower for foreign citizens than German citizens. There is an important difference between the results presented in Table 3.4 and Table 3.22 though. For the urban sample, the coefficients reported in Row 3 are not larger than the ones in Row 2. From this evidence we conclude that estimates on net internal migration inflows for foreigners remain positive (although lower than the ones for natives). The same does not hold for the results presented in Table 3.4. When considering all counties, the impact of AI adoption on foreigners' cumulative net internal migration inflows becomes negative. One possible interpretation of these differences is a decreased internal migration inflow of foreigners from rural to urban counties in Germany as a response to AI.

In the case of immigrant inflows, we find that our results from urban counties coincide with those from the full sample (see Table 3.23). All coefficients presented in the table are insignificant at the common significance levels. AI adoption does not lead to a significant increase in immigrant inflows from abroad. Similarly, there are no systematic differences between regression estimations on wages. Table 3.24 reveals that all but one point estimate are insignificant at the common significance levels. While the OLS coefficient on daily wages of the middle-skilled is significant at the 10 percent significance level, this is not the case for yearly earnings. Combining the evidence from daily wages and yearly earnings makes a case against significant effects on wages in urban counties, which is in line with observations from the full sample.

Moving attention to Table 3.25 reveals that there are important differences between the full and urban sample in the case of unemployment rates. While there were no significant displacement effects when investigating all German counties, there is evidence pointing towards an increase in unemployment rates among high- and middle-skilled German citizens as a response to AI adoption. The IV coefficients in Column 4 and 6 are significant at the 5 percent significance level. According to these results, a 0.01 percentage point increase in AI-related skill demands leads to an average increase in unemployment rates of 18.6 percent for high-skilled natives. Similarly, middle-skilled workers' unemployment rates raise by 10.9 percent in response to AI adoption. In addition, one of the OLS coefficients in Row 3 is significant at the 5 percent significance level. Middle-skilled foreigners also face displacement effects as a result of AI adoption. Consequently, high-skilled task automation might indeed result in displacement effects. One possible explanation for the absence of these kind

Table 3.22: AI skill demands and cumulative net internal migration inflows by skill groups and citizenship at the county level (urban sample)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	-16.22	120.2*	-23.25**	26.53	4.649*	67.44**	2.386	26.27**
	(10.42)	(66.25)	(9.011)	(23.97)	(2.377)	(31.41)	(2.659)	(11.66)
AI	11.73**	234.8**	4.113*	80.94**	4.833**	92.17***	2.788***	42.00***
	(4.920)	(91.55)	(2.134)	(32.86)	(1.896)	(31.89)	(1.027)	(15.13)
Foreign*AI	-9.257**	-182.1**	-3.759**	-66.83**	-3.475**	-83.02***	-2.022***	-32.28***
	(3.944)	(71.50)	(1.854)	(28.18)	(1.495)	(31.21)	(0.776)	(12.52)
Constant	-92.50	-174.1	52.24	-14.75	-119.7	-33.30	-25.05	-25.34**
	(137.6)	(2146.1)	(106.8)	(24.23)	(75.15)	(29.00)	(43.73)	(12.66)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0508	.	0.126	.	0.0367	.	0.0490	.
N	692	692	692	692	692	692	692	692

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. We restrict the sample to urban counties. The unit of analysis are the county-period-citizen cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. We measure AI skill demands via the share of AI-related skill demand in all skill demand in a local labor market. The immigrant inflow is the cumulative inflow into a specific county. We identify the immigrant inflow by aggregating the number of times a foreigner appears for the first time in the SIAB per year. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.23: AI skill demands and cumulative immigrant inflows from abroad by skill groups and at the county level (urban sample)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	-0.0177	49.38	-0.0983	17.33	0.0493	21.27	0.0246	10.40
	(3.331)	(38.28)	(0.989)	(13.54)	(1.586)	(16.97)	(0.798)	(7.526)
Constant	1922.1	-455.6	727.9	39.75	722.8	-335.7	463.2	-153.1
	(1542.7)	(1634.1)	(468.8)	(530.6)	(693.1)	(729.8)	(384.7)	(372.1)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.870	0.323	0.888	0.286	0.858	0.324	0.848	0.373
N	346	346	346	346	346	346	346	346

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. We restrict the sample to urban counties. The unit of analysis are county-citizen cells. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of job skills mentioning AI-related expressions in all skill demand. The outcome variable is the cumulative immigrant inflow into a specific county. We identify the immigrant inflow by aggregating the number of times a foreigner appears for the first time in the SIAB per stacked-long difference. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of displacement effects in the full sample might be due to urban-rural migration dynamics mitigating these displacement effects.

Our findings on unemployment rates should be taken with caution, given that there is no significant impact in the case of OLS coefficients for natives. Vice versa, the coefficient on middle-skilled foreign workers' unemployment rate becomes insignificant for IV regressions. Moreover, we observe a inconsistency in the sign of OLS and IV estimates for middle-skilled natives. These deviations might point towards an important inconsistency in the IV and OLS estimates.

Table 3.24: AI skill demands and percentage change in daily wages by skill groups and citizenship at the county level (urban sample)

	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	1.036	1.262	-0.216	0.292	0.434	1.381
	(1.124)	(1.959)	(0.663)	(1.221)	(0.900)	(1.901)
AI	0.154	1.032	-0.0210	0.231	0.0855	0.160
	(0.113)	(0.756)	(0.0390)	(0.378)	(0.118)	(0.650)
Foreign*AI	0.412	0.126	0.193*	-0.451	-0.263	-1.463
	(0.459)	(1.961)	(0.106)	(0.919)	(0.212)	(1.716)
Constant	35.25	4.302*	-2.524	5.028***	-3.587	3.029
	(71.99)	(2.458)	(14.41)	(1.360)	(37.00)	(3.491)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0256	0.00479	0.206	0.144	0.0810	0.0266
N	692	692	692	692	692	692

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. We restrict the sample to urban counties. The unit of analysis are county-period-citizen cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of AI-related skills in the total number of skills. The outcome variable is the percentage change in daily wages. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to information and communication technology as well as trade, using data from Eurostat as well as Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.25: AI skill demands and percentage change in unemployment rates by skill groups and citizenship at the county level (urban sample)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	13.42***	18.40**	-8.155	-13.78	32.34***	38.33***	22.25**	16.29
	(3.533)	(8.651)	(7.938)	(17.43)	(5.487)	(11.50)	(10.14)	(16.94)
AI	0.485	11.77**	-0.402	18.61**	0.726	10.92**	0.860	6.105
	(0.523)	(4.674)	(2.058)	(9.490)	(0.552)	(4.297)	(1.248)	(13.01)
Foreign*AI	1.450	-4.945	0.824	5.578	3.943**	-3.632	-2.068	6.117
	(1.262)	(10.15)	(3.122)	(20.44)	(1.698)	(12.17)	(2.103)	(21.48)
Constant	-112.1	-114.9	463.6	571.3	-222.2	-217.7	170.7	117.9
	(155.8)	(217.0)	(317.6)	(395.8)	(192.3)	(230.0)	(445.4)	(496.2)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0565	.	0.0318	.	0.135	0.0283	0.0628	.
N	671	671	509	509	643	643	603	603

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. We restrict the sample to urban counties. The unit of analysis are county-period-citizen cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of job vacancies mentioning AI-related expressions in the total number of job vacancies. The outcome variable is the percentage change in unemployment rates. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population and county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.8.3 Alternative Empirical Specifications

We next investigate the robustness of our findings to alternative empirical specifications. In particular, we vary the number of fixed effects included in our main regression specification and the level at which we cluster our standard errors. Column 1 and 2 in Table 3.26 show the results for a regression specification without fixed effects. Standard errors are clustered at the county (NUTS-3) level in this case. These clustered standard errors account for the fact that AI adoption and migration flows may be correlated across the stacked 3-year periods at the county level.

Column 3 and 4 present results from our baseline specification, which controls for federal state (NUTS-1) fixed effects and clusters standard errors at the NUTS-3 level. Controlling for federal state fixed effects allows us to account for observable and unobservable differences in federal state characteristics that may affect both AI adoption and migration flows. One example would be migration policies or innovation strategies at the federal state level. This specification only partly eliminates potential omitted variable bias as it only accounts for differences in federal state characteristics that are constant over time. If unobservable factors within federal states, such as economic conditions of local labor markets or more conservative attitudes towards immigration and innovations, are correlated with AI adoption and also affect migration flows, our OLS regressions might still suffer from omitted variable bias. Comparing our baseline specification to the specification without fixed effects demonstrates that the OLS coefficient is lower while the IV coefficient is larger. The opposite is true for standard errors. Controlling for federal state fixed effects decreases standard errors in the case of OLS coefficients and increases them for IV coefficients.

Column 5 and 6 in addition add time fixed effects to our regression. These time fixed effects control for observable and unobservable variables which are constant across counties but vary over the two stacked 3-year periods analyzed in this paper. Example are economic downturns affecting all counties (e.g. a recession), developments in inflation, or changes in the ruling party at the national level in Germany. Adding time fixed effects only slightly affects IV regressions and standard errors, while it further decreases the OLS estimate.

Lastly, Column 7 and 8 cluster standard errors at the regional (NUTS-2) level instead of the county level. Under this specification, AI adoption at the county level is not only correlated across stacked periods, but also across counties located in the same region. This assumption might be less realistic for the underlying research question, given that AI adoption might be a local phenomena, depending on the local presence of companies and the necessary infrastructure to develop and employ AI. Still, we include this option for the sake of completeness. Standard errors increase under this specification for IV estimations, but decrease for OLS estimations.

In summary, our main insights from Table 3.6 hold under the different model specification investigated. There is no significant increase in cumulative immigrant inflows from abroad as a response to an increase in AI-related skill demands.

Table 3.27 presents similar results for cumulative net internal migration inflows. The coefficients presented in Row 2 are sensitive to model specifications. The point estimate is insignificant when including time fixed effects or clustering standard errors at the regional

Table 3.26: AI skill demands and cumulative immigrant inflows from abroad by skill groups and at the county level - alternative empirical specifications

	No fixed effects	No fixed effects	Baseline	Baseline	Time fixed effects	Time fixed effects	Regional clusters	Regional clusters
AI	0.840	33.02	0.478	48.88	0.243	49.31	0.243	49.31
	(0.789)	(22.10)	(0.463)	(45.00)	(0.629)	(40.90)	(0.605)	(43.29)
Constant	555.8	-188.8	2147.6*	-565.7	2126.3*	-576.0	2126.3*	-576.0
	(1639.4)	(1248.6)	(1208.1)	(1347.9)	(1207.9)	(1345.8)	(1261.2)	(1435.8)
Federal States fixed effect	No	No	Yes	Yes	Yes	Yes	No	No
Period fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
Adj. R-squared	0.439	0.147	0.864	.	0.867	.	0.867	.
N	762	762	762	762	762	762	762	762

Notes: The table presents stacked long-difference regressions for two periods of three years, 2014-2016 and 2017-2019, for different empirical specifications. The unit of analysis are county-period-citizen cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of skills mentioning AI-related expressions in all skills. The outcome variable is the cumulative immigrant inflow into a specific county. We identify the immigrant inflow by aggregating the number of times a foreigner appears for the first time in the SIAB per stacked long-difference. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Clustered standard errors are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

level. We are confident that a correlation of AI adoption across counties within one region is unlikely and are therefore less worried about the insignificance of the last coefficient. Still, it is an important limitation that the inclusion of period fixed effects results in insignificant estimators. Moving to Row 3, which shows the impact on foreign citizens' immigrant inflows, illustrates that coefficients are robust to all model specifications in this case.

Table 3.27: AI skill demands and net internal migration inflows by skill groups and citizenship at the county level - alternative empirical specifications

	No fixed effects	No fixed effects	Baseline	Baseline	Time fixed effects	Time fixed effects	Regional clusters	Regional clusters
Foreign	-8.898	94.24	-8.898	94.24	-8.898	94.24	-8.898	94.24
	(7.948)	(61.10)	(7.987)	(61.10)	(7.990)	(61.10)	(8.005)	(61.60)
AI	1.974	299.8*	1.984	293.7**	0.841	536.8	0.841	536.8
	(1.587)	(154.0)	(1.614)	(146.5)	(0.534)	(524.1)	(0.542)	(496.2)
Foreign*AI	-1.658	-233.4**	-1.658	-233.4**	-1.658	-233.4**	-1.658	-233.4**
	(1.312)	(111.4)	(1.319)	(111.4)	(1.319)	(111.4)	(1.339)	(111.3)
Constant	67.22	-4490.9	-43.97	-4394.3	-147.0*	-10197.5	-147.0*	-10197.5
	(72.65)	(5618.6)	(82.76)	(4728.3)	(84.82)	(15000.3)	(88.24)	(14167.1)
Federal States fixed effect	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Period fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
Adj. R-squared	0.00601	.	0.00886	.	0.255	.	0.255	.
N	1524	1524	1524	1524	1524	1524	1524	1524

Notes: The table presents stacked long-difference regressions for two periods of three years, 2014-2016 and 2017-2019, for different empirical specifications. The unit of analysis are county-period-citizen cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of skills mentioning AI-related expressions in all skills. The immigrant inflow is the cumulative inflow into a specific county. We identify the immigrant inflow by aggregating the number of times a foreigner appears for the first time in the SIAB per year. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Clustered standard errors are in parentheses. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.9 Conclusion

This paper analyzes if automation affects migration flows. We consider both internal migration as well as immigrant inflows from abroad. In particular, we focus on automation of high-skilled tasks, which we measure by AI-related skill demand relying on online job vacancy data. This is a relevant question as AI adoption involves highly specialized tasks (Zanzotto 2019), which might lead to skill shortages in local labor markets. Skill shortages, on the other hand, might result in labor market frictions and productivity losses (Acemoglu et al. 2020). We bring a new aspect to the literature studying labor market effects of automation of high-skilled tasks by making a distinction between foreigners and natives. This distinction is interesting as it can shed light on important underlying drivers of labor market effects of technological change.

We develop a novel shift-share instrument using AI adoption in Switzerland. We employ online job vacancy data and matched employer-employee data from Germany. We apply our research question to the context of Germany since it is one of the leading automation economies and one of the main recipients of immigrants over recent decades. Our analyses reveal that an increase in AI-related skill demand increases the cumulative net internal migration inflow. This result could mean that firms cover skill shortages in certain areas by recruiting in other areas. At the same time, there is no significant increase in immigrant inflows from abroad. Interestingly, while natives are more likely to move between counties as a response to AI adoption, this is not the case for foreigners. These findings are in contrast to the standard economic theory (Borjas 2003) and could imply that foreigners have less access to relevant information about the need to adapt their skill-set, retrain, or move to other counties. Generally, our results speak against foreigners and natives being skill type perfect substitutes. Our findings are not confounded by urbanization. The shock component of our shift-share instrument seems to satisfy the exclusion restriction, speaking for the validity of our empirical strategy. Still, there are some important inconsistencies between OLS and IV coefficients, which could point towards important limitations in the analysis at hand, or are evidence of unobservable variables that bias OLS estimates.

Finally, we consider possible explanations for why internal migration responses differ by citizenship. Labor market effects of AI adoption are unlikely driving the different internal migration patterns of foreigners and natives. Estimating the impact of AI adoption on wages and unemployment rates reveals that there are no significant labor market effects of automation of high-skilled tasks. These results hold independently of citizenship and education levels. In addition, we demonstrate that spillover effects between most and least exposed sectors are unlikely to influence our main results on cumulative net internal migration inflows. Restricting the sample to the most and least exposed sectors does not result in systematic differences. There is some evidence showing that a diverging probability to switch sectors might drive the difference in migration dynamics by citizenship.

We also make a comparison to low-skilled task automation relying on data on robot exposure. These analyses illustrate that automation of low-skilled tasks does not lead to any significant effects on unemployment rates nor cumulative immigrant inflows from abroad. In contrast to automation of high-skilled tasks, there are wage losses for foreign citizens, but increases for natives. This pattern of results speaks for systematic differences in the impact of high- and low-skilled task automation on labor market outcomes. Internal migration dynamics also differ by the skills automation is likely to replace. Differently from high-skilled task automation, industrial robots results in a cumulative net internal migration inflow for foreigners, but decreases it for natives.

This paper illustrates that migration dynamics are an important mechanism in response to labor market shocks induced by automation of high-skilled tasks. In addition, our analyses demonstrate that the impact of automation differs by the skill group it mainly addresses. Future research should investigate if our findings also hold for longer time periods. Investigating the underlying research question through more robust estimation techniques could also bring valuable insights. Future research should also investigate potential drivers of diverging migration responses to automation by citizenship in more detail.

Bibliography

Abrams, David S (2021). "COVID and crime: An early empirical look". In: *Journal of public economics* 194, p. 104344.

Acemoglu, Daron and Pascual Restrepo (2018a). "Low-skill and high-skill automation". In: *Journal of Human Capital* 12.2, pp. 204–232.

— (2018b). "The race between man and machine: Implications of technology for growth, factor shares, and employment". In: *American Economic Review* 108.6, pp. 1488–1542.

Acemoglu, Daron et al. (2020). *AI and jobs: Evidence from online vacancies*. Tech. rep. National Bureau of Economic Research.

Agarwal, Saharsh and Ananya Sen (2022). "Antiracist Curriculum and Digital Platforms: Evidence from Black Lives Matter". In: *Management Science* 68.4, pp. 2932–2948.

Agencia de Calidad de la Educación (2021). *Base de Datos de la Agencia de Calidad de la Educación 2015-2018*. `https://www.agenciaeducacion.cl/`. Santiago de Chile.

Agranov, Marina, Matt Elliott, and Pietro Ortoleva (2021). "The importance of Social Norms against Strategic Effects: The case of COVID-19 vaccine uptake". In: *Economics Letters* 206, p. 109979.

Agüero, Jorge M (2021). "COVID-19 and the rise of intimate partner violence". In: *World development* 137, p. 105217.

Aizer, Anna (2010). "The gender wage gap and domestic violence". In: *American Economic Review* 100.4, pp. 1847–59.

Aizer, Anna et al. (2016). "The long-run impact of cash transfers to poor families". In: *American Economic Review* 106.4, pp. 935–71.

Akee, Randall, Maggie R Jones, Emilia Simeonova, et al. (2020). "The EITC and Linking Data for Examining Multi-Generational Effects". In: *NBER Chapters*.

Akee, Randall et al. (2018). "How does household income affect child personality traits and behaviors?" In: *American Economic Review* 108.3, pp. 775–827.

Alekseeva, Liudmila et al. (2021). "The demand for AI skills in the labor market". In: *Labour Economics*, p. 102002.

Almond, Douglas, Janet Currie, and Valentina Duque (2018). "Childhood circumstances and adult outcomes: Act II". In: *Journal of Economic Literature* 56.4, pp. 1360–1446.

Almond, Douglas, Hilary W Hoynes, and Diane Whitmore Schanzenbach (2011). "Inside the war on poverty: The impact of food stamps on birth outcomes". In: *The review of economics and statistics* 93.2, pp. 387–403.

Almond, Douglas, Bhashkar Mazumder, and Reyn Van Ewijk (2015). "In utero Ramadan exposure and children's academic performance". In: *The Economic Journal* 125.589, pp. 1501–1533.

Amaral, Sofia, Siddhartha Bandyopadhyay, and Rudra Sensarma (2015). "Employment programmes for the poor and female empowerment: The effect of NREGS on gender-based violence in India". In: *Journal of interdisciplinary economics* 27.2, pp. 199–218.

Amaral, Sofia, Sonia Bhalotra, and Nishith Prakash (2021). "Gender, crime and punishment: Evidence from women police stations in india". In: *CESifo Working Paper*.

Amarante, Verónica et al. (2016). "Do cash transfers improve birth outcomes? Evidence from matched vital statistics, program, and social security data". In: *American Economic Journal: Economic Policy* 8.2, pp. 1–43.

Amenta, Edwin and Drew Halfmann (2000). "Wage wars: Institutional politics, WPA wages, and the struggle for US social policy". In: *American Sociological Review*, pp. 506–528.

Anderson, Julia, Paco Viry, and Guntram B Wolff (2020). "Europe has an artificial-intelligence skills shortage". In: *Bruegel-Blogs*.

Angrist, Joshua D and Jörn-Steffen Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Antoni, Manfred et al. (2021). *Weakly anonymous Version of the Sample of Integrated Labour Market Biographies (SIAB) – Version 7519 v1*. Tech. rep. Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg.

Asesorías para el Desarrollo (2012). *Evaluación de Impacto del Sistema de Protección Integral a la Infancia (Chile Crece Contigo) – Informe Final Revisado 2012*. `http://www.crececontigo.gob.cl/wp-content/uploads/2012/09/Informe-Final-Evaluacio%CC%81n-de-Impacto-ChCC-2012.pdf`.

Attanasio, Orazio and Miguel Székely (1999). "An asset-based approach to the analysis of poverty in Latin America". In.

Attanasio, Orazio et al. (2020). "Estimating the production function for human capital: results from a randomized controlled trial in Colombia". In: *American Economic Review* 110.1, pp. 48–85.

Bailey, Martha J, Shuqiao Sun, and Brenden Timpe (2021). "Prep School for poor kids: The long-run impacts of Head Start on Human capital and economic self-sufficiency". In: *American Economic Review* 111.12, pp. 3963–4001.

Bailey, Martha J et al. (2020). *Is the social safety net a long-term investment? Large-scale evidence from the food stamps program.* Tech. rep. National Bureau of Economic Research.

Baker, Michael, Jonathan Gruber, and Kevin Milligan (2008). "Universal child care, maternal labor supply, and family well-being". In: *Journal of political Economy* 116.4, pp. 709–745.

Balestrino, Alessandro (2008). "It is a theft but not a crime". In: *European Journal of Political Economy* 24.2, pp. 455–469.

Bandyopadhyay, Debasis, James Allan Jones, and Asha Sundaram (2020). "Gender Bias and Male Backlash as Drivers of Crime Against Women: Evidence from India". In: *The University of Auckland Business School Research Paper Forthcoming.*

Barrera-Osorio, Felipe, Leigh L Linden, and Juan E Saavedra (2019). "Medium-and long-term educational consequences of alternative conditional cash transfer designs: Experimental evidence from Colombia". In: *American Economic Journal: Applied Economics* 11.3, pp. 54–91.

Bartik, Timothy J (1991). "Who benefits from state and local economic development policies?" In: *WE Upjohn Institute for Employment Research Kalamazoo, MI.*

Basso, Gaetano, Giovanni Peri, and Ahmed S Rahman (2020). "Computerization and immigration: Theory and evidence from the United States". In: *Canadian Journal of Economics/Revue canadienne d'économique* 53.4, pp. 1457–1494.

Battisti, Michele, Ilpo Kauppinen, and Britta Rude (2022). *Twitter and crime: The effect of social movements on gender-based violence.* Tech. rep. ifo Working Paper.

Battisti, Michele et al. (2018). "Immigration, search and redistribution: A quantitative assessment of native welfare". In: *Journal of the European Economic Association* 16.4, pp. 1137–1188.

Becker, Gary S et al. (1995). "The economics of crime". In: *Cross Sections* 12.Fall, pp. 8–15.

Beerli, Andreas, Ronald Indergand, and Johannes S Kunz (2021). *The supply of foreign talent: How skill-biased technology drives the location choice and skills of new immigrants.* Tech. rep. GLO Discussion Paper.

Bénabou, Roland and Jean Tirole (2006). "Incentives and prosocial behavior". In: *American economic review* 96.5, pp. 1652–1678.

Bender, Stefan et al. (1996). "Die IAB-Beschäftigtenstichprobe 1975–1990. Beiträge zur Arbeitsmarkt-und Berufsforschung 197". In: *Nürnberg: Institut für Arbeitsmarkt-und Berufsforschung.*

Berniell, Inés and Gabriel Facchini (2021). “COVID-19 lockdown and domestic violence: Evidence from internet-search behavior in 11 countries”. In: *European Economic Review* 136, p. 103775.

Besley, Timothy and Maitreesh Ghatak (2018). “Prosocial motivation and incentives”. In: *Annual Review of Economics* 10, pp. 411–438.

Bhalotra, Sonia et al. (2021a). “Intimate partner violence: The influence of job opportunities for men and women”. In: *The World Bank Economic Review* 35.2, pp. 461–479.

Bhalotra, Sonia et al. (2021b). “Job displacement, unemployment benefits and domestic violence”. In: *CEPR Discussion Paper No. DP16350.*

Bharadwaj, Prashant, Juan Pedro Eberhard, and Christopher A Neilson (2018a). “Health at birth, parental investments, and academic outcomes”. In: *Journal of Labor Economics* 36.2, pp. 349–394.

— (2018b). “Health at birth, parental investments, and academic outcomes”. In: *Journal of Labor Economics* 36.2, pp. 349–394.

Bharadwaj, Prashant, Katrine Vellesen Løken, and Christopher Neilson (2013). “Early life health interventions and academic achievement”. In: *American Economic Review* 103.5, pp. 1862–91.

Black, Maureen M et al. (2017). “Early childhood development coming of age: science through the life course”. In: *The Lancet* 389.10064, pp. 77–90. DOI: `10.1016/S0140-6736(16)31389-7`. URL: `https://doi.org/10.1016/S0140-6736(16)31389-7`.

Black, Sandra E et al. (2014). “Care or cash? The effect of child care subsidies on student performance”. In: *Review of Economics and Statistics* 96.5, pp. 824–837.

Blanchard, OJ and LF Katz (1992). “Regional Evolutions, in Brookings papers on Economic Activity, n. 1”. In.

Bobonis, Gustavo J, Melissa González-Brenes, and Roberto Castro (2013). “Public transfers and domestic violence: The roles of private information and spousal control”. In: *American Economic Journal: Economic Policy* 5.1, pp. 179–205.

Borjas, George J (2001). “Does immigration grease the wheels of the labor market?” In: *Brookings papers on economic activity* 2001.1, pp. 69–133.

— (2003). “The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market”. In: *The quarterly journal of economics* 118.4, pp. 1335–1374.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel (2022). “Quasi-experimental shift-share research designs”. In: *The Review of Economic Studies* 89.1, pp. 181–213.

Brassiolo, Pablo (2016). “Domestic violence and divorce law: When divorce threats become credible”. In: *Journal of Labor Economics* 34.2, pp. 443–477.

Bremer, Björn, Swen Hutter, and Hanspeter Kriesi (2020). "Dynamics of protest and electoral politics in the Great Recession". In: *European Journal of Political Research* 59.4, pp. 842–866.

Breton, Albert and Raymond Breton (1969). "An economic theory of social movements". In: *The American Economic Review* 59.2, pp. 198–205.

Brown, Stephen J and Jerold B Warner (1985). "Using daily stock returns: The case of event studies". In: *Journal of financial economics* 14.1, pp. 3–31.

Bullinger, Lindsey Rose, Jillian B Carr, and Analisa Packham (2021). "COVID-19 and crime: Effects of stay-at-home orders on domestic violence". In: *American Journal of Health Economics* 7.3, pp. 249–280.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin (2020). "From extreme to mainstream: The erosion of social norms". In: *American economic review* 110.11, pp. 3522–48.

Bursztyn, Leonardo et al. (2021). "Persistent political engagement: Social interactions and the dynamics of protest movements". In: *American Economic Review: Insights* 3.2, pp. 233–50.

Cadena, Brian C and Brian K Kovak (2016). "Immigrants equilibrate local labor markets: Evidence from the Great Recession". In: *American Economic Journal: Applied Economics* 8.1, pp. 257–90.

Cameron, Roslyn (2011). "Responding to Australia's regional skill shortages through regional skilled migration". In: *Journal of Economic & Social Policy* 14.3, pp. 46–80.

Campbell, Frances et al. (2014). "Early childhood investments substantially boost adult health". In: *Science* 343.6178, pp. 1478–1485.

Carneiro, Pedro and Rita Ginja (2014). "Long-term impacts of compensatory preschool on health and behavior: Evidence from Head Start". In: *American Economic Journal: Economic Policy* 6.4, pp. 135–73.

Carter, Michael R and Christopher B Barrett (2006). "The economics of poverty traps and persistent poverty: An asset-based approach". In: *The Journal of Development Studies* 42.2, pp. 178–199.

Cascio, Elizabeth U (2017). *Does universal preschool hit the target? Program access and preschool impacts.* Tech. rep. National Bureau of Economic Research.

Cattaneo, Matias D, Nicolás Idrobo, and Rocío Titiunik (2019). *A practical introduction to regression discontinuity designs: Foundations.* Cambridge University Press.

Cattaneo, Matias D, Rocio Titiunik, and Gonzalo Vazquez-Bare (2016). "Inference in regression discontinuity designs under local randomization". In: *The Stata Journal* 16.2, pp. 331–367.

Chakraborty, Tanika et al. (2018). “Stigma of sexual violence and women’s decision to work”. In: *World Development* 103, pp. 226–238.

Chernin, Yulia and Yaron Lahav (2014). ““The people demand social justice” a case study on the impact of protests on financial markets”. In: *Accounting, Economics and Law* 4.2, pp. 99–121.

Chetty, Raj et al. (2011). “How does your kindergarten classroom affect your earnings? Evidence from Project STAR”. In: *The Quarterly journal of economics* 126.4, pp. 1593–1660.

Chiarello, Filippo et al. (2021). “Towards ESCO 4.0–Is the European classification of skills in line with Industry 4.0? A text mining approach”. In: *Technological Forecasting and Social Change* 173, p. 121177.

Clarke, Damian, Gustavo Cortés Méndez, and Diego Vergara Sepúlveda (2020). “Growing together: assessing equity and efficiency in a prenatal health program”. In: *Journal of Population Economics*, pp. 1–74.

Clarke, Damian and Kathya Tapia-Schythe (2021). “Implementing the panel event study”. In: *The Stata Journal* 21.4, pp. 853–884.

Comito, Carmela (2021). “How covid-19 information spread in us the role of twitter as early indicator of epidemics”. In: *IEEE Transactions on Services Computing*.

Cools, Sara and Andreas Kotsadam (2017). “Resources and intimate partner violence in Sub-Saharan Africa”. In: *World Development* 95, pp. 211–230.

Cooper, Jasper, Donald P Green, and Anna M Wilke (2020). “Reducing Violence against Women in Uganda through Video Dramas: A Survey Experiment to Illuminate Causal Mechanisms”. In: *AEA Papers and Proceedings*. Vol. 110, pp. 615–19.

Cornelissen, Thomas et al. (2018). “Who benefits from universal child care? Estimating marginal returns to early child care attendance”. In: *Journal of Political Economy* 126.6, pp. 2356–2409.

Cortés, Claudio Lara (2016). “The Global Crisis and the Chilean Economy”. In: *Latin America after the Financial Crisis*. Springer, pp. 117–140.

Cramer, Ulrich (1985). “Probleme der genauigkeit der beschäftigtenstatistik”. In: *Allgemeines Statistisches Archiv* 69, pp. 56–68.

Cullen, Claire (2020). “Method matters: Underreporting of intimate partner violence in Nigeria and Rwanda”. In: *World Bank Policy Research Working Paper* 9274.

Cunha, Flavio and James Heckman (2007). “The technology of skill formation”. In: *American Economic Review* 97.2, pp. 31–47.

Cunha, Flavio, James J Heckman, and Susanne M Schennach (2010). "Estimating the technology of cognitive and noncognitive skill formation". In: *Econometrica* 78.3, pp. 883–931.

Currie, Janet and Douglas Almond (2011). "Human capital development before age five". In: *Handbook of labor economics*. Vol. 4. Elsevier, pp. 1315–1486.

Daelmans, Bernadette et al. (2017). "Early childhood development: the foundation of sustainable development". In: *The Lancet* 389.10064, pp. 9–11. DOI: 10.1016/S0140-6736(16)31659-2. URL: https://doi.org/10.1016/S0140-6736(16)31659-2.

Dahl, Gordon B and Lance Lochner (2012). "The impact of family income on child achievement: Evidence from the earned income tax credit". In: *American Economic Review* 102.5, pp. 1927–56.

Danzer, Alexander, Carsten Feuerbaum, and Fabian Gaessler (2020). "Labor supply and automation innovation". In: *Max Planck Institute for Innovation & Competition Research Paper* 20-09.

Dauth, Wolfgang and Johann Eppelsheimer (2020). "Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: a guide". In: *Journal for Labour Market Research* 54.1, pp. 1–14.

Dauth, Wolfgang et al. (2021). "The adjustment of labor markets to robots". In: *Journal of the European Economic Association* 19.6, pp. 3104–3153.

Dave, Dhaval M et al. (2020). *Black lives matter protests and risk avoidance: The case of civil unrest during a pandemic*. Tech. rep. National Bureau of Economic Research.

Delaporte, Magdalena and Francisco Pino (2022). "Female Political Representation and Violence Against Women: Evidence from Brazil". In: *IZA Discussion Paper*.

Delker, Brianna C et al. (2020). "Who has to tell their trauma story and how hard will it be? Influence of cultural stigma and narrative redemption on the storying of sexual violence". In: *Plos one* 15.6, e0234201.

Deming, David (2009). "Early childhood intervention and life-cycle skill development: Evidence from Head Start". In: *American Economic Journal: Applied Economics* 1.3, pp. 111–34.

Dolley, James Clay (1933). "Characteristics and procedure of common stock split-ups". In: *Harvard business review* 11.3, pp. 316–326.

Duflo, Esther (2001). "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment". In: *American economic review* 91.4, pp. 795–813.

Duvvury, Nata et al. (2013). "Intimate partner violence: Economic costs and implications for growth and development". In: *World Bank. License: CC BY 3.0 IGO.* URL: https://openknowledge.worldbank.org/handle/10986/16697.

ElSherief, Mai, Elizabeth Belding, and Dana Nguyen (2017). "# notokay: Understanding gender-based violence in social media". In: *Eleventh International AAAI Conference on Web and Social Media*.

Elsner, Benjamin and Klaus F Zimmermann (2016). "Migration 10 years after: EU enlargement, closed borders, and migration to Germany". In: *Labor migration, EU enlargement, and the great recession*. Springer, pp. 85–101.

Erten, Bilge and Pinar Keskin (2018). "For better or for worse?: Education and the prevalence of domestic violence in turkey". In: *American Economic Journal: Applied Economics* 10.1, pp. 64–105.

Faber, Marius, Andrés P Sarto, and Marco Tabellini (2022). *Local Shocks and Internal Migration: The Disparate Effects of Robots and Chinese Imports in the US*. Tech. rep. National Bureau of Economic Research.

Falk, Armin and Urs Fischbacher (2002). ""Crime" in the lab-detecting social interaction". In: *European Economic Review* 46.4-5, pp. 859–869.

Felfe, Christina and Rafael Lalive (2018). "Does early child care affect children's development?" In: *Journal of Public Economics* 159, pp. 33–53.

Felten, Edward, Manav Raj, and Robert Channing Seamans (2019). "The effect of artificial intelligence on human labor: An ability-based approach". In: *Academy of Management Proceedings*. Vol. 2019. 1. Academy of Management Briarcliff Manor, NY 10510, p. 15784.

Fernández-Fontelo, Amanda et al. (2019). "Untangling serially dependent underreported count data for gender-based violence". In: *Statistics in medicine* 38.22, pp. 4404–4422.

Finn, Mary A and Pamela Bettis (2006). "Punitive action or gentle persuasion: Exploring police officers' justifications for using dual arrest in domestic violence cases". In: *Violence against women* 12.3, pp. 268–287.

Fitzgerald, Louise F and Lilia M Cortina (2018). "Sexual harassment in work organizations: A view from the 21st century." In.

Folke, Olle et al. (2020). "Sexual harassment of women leaders". In: *Daedalus* 149.1, pp. 180–197.

Francois, Patrick and Michael Vlassopoulos (2008). "Pro-social motivation and the delivery of social services". In: *CESifo Economic Studies* 54.1, pp. 22–54.

French, Michael T et al. (2015). "What you do in high school matters: High school GPA, educational attainment, and labor market earnings as a young adult". In: *Eastern Economic Journal* 41.3, pp. 370–386.

Fry, M Whitney, Asheley C Skinner, and Stephanie B Wheeler (2019). "Understanding the relationship between male gender socialization and gender-based violence among refugees in Sub-Saharan Africa". In: *Trauma, Violence, & Abuse* 20.5, pp. 638–652.

Gagliarducci, Stefano and M Daniele Paserman (2012). "Gender interactions within hierarchies: evidence from the political arena". In: *The Review of Economic Studies* 79.3, pp. 1021–1052.

Gangadharan, Lata et al. (2019). "Female leaders and their response to the social environment". In: *Journal of Economic Behavior & Organization* 164, pp. 256–272.

García, Jorge Luis, James J Heckman, and Anna L Ziff (2018). "Gender differences in the benefits of an influential early childhood program". In: *European economic review* 109, pp. 9–22.

Gertler, Paul et al. (2014). "Labor market returns to an early childhood stimulation intervention in Jamaica". In: *Science* 344.6187, pp. 998–1001.

Giesing, Yvonne and Britta Rude (2022). "Robots, AI, and Immigration–A Race for Talent or of Displaced Workers". In: *CESifo Forum*. Vol. 23. 5. Institut für Wirtschaftsforschung (Ifo), pp. 20–23.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift (2020). "Bartik instruments: What, when, why, and how". In: *American Economic Review* 110.8, pp. 2586–2624.

González, Libertad and Núria Rodríguez-Planas (2020). "Gender norms and intimate partner violence". In: *Journal of Economic Behavior & Organization* 178, pp. 223–248.

Goodman-Bacon, Andrew (2018). "Public insurance and mortality: evidence from Medicaid implementation". In: *Journal of Political Economy* 126.1, pp. 216–262.

Google Trends (2021). *National Domestic Violence Hotline*. `https://trends.google.com/trends/explore?date=all&geo=US&q=nationaldomesticviolencehotline`.

Graetz, Georg and Guy Michaels (2018). "Robots at work". In: *Review of Economics and Statistics* 100.5, pp. 753–768.

Greenland, Andrew, John Lopresti, and Peter McHenry (2019). "Import competition and internal migration". In: *Review of Economics and Statistics* 101.1, pp. 44–59.

Guarnieri, Eleonora and Helmut Rainer (2021). "Colonialism and female empowerment: A two-sided legacy". In: *Journal of Development Economics* 151, p. 102666.

Hanson, Gordon H (2021). *Immigration and Regional Specialization in AI*. Tech. rep. National Bureau of Economic Research.

Harris, Wesley Eugene (2017). "The Effect of Stigma on Intimate Partner Violence Reporting Among Men Who Have Sex with Men". In: *East Tennessee State University*.

Havnes, Tarjei and Magne Mogstad (2011). "No child left behind: Subsidized child care and children's long-run outcomes". In: *American Economic Journal: Economic Policy* 3.2, pp. 97–129.

— (2015). "Is universal child care leveling the playing field?" In: *Journal of public economics* 127, pp. 100–114.

Hawelka, Bartosz et al. (2014). "Geo-located Twitter as proxy for global mobility patterns". In: *Cartography and Geographic Information Science* 41.3, pp. 260–271.

Heckman, James, Rodrigo Pinto, and Peter Savelyev (2013). "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes". In: *American Economic Review* 103.6, pp. 2052–86.

Heckman, James J (2006). "Skill formation and the economics of investing in disadvantaged children". In: *Science* 312.5782, pp. 1900–1902.

Heckman, James J et al. (2010). "The rate of return to the HighScope Perry Preschool Program". In: *Journal of public Economics* 94.1-2, pp. 114–128.

Hendren, Nathaniel (2016). "The policy elasticity". In: *Tax Policy and the Economy* 30.1, pp. 51–89.

Hendren, Nathaniel and Ben Sprung-Keyser (2020). "A unified welfare analysis of government policies". In: *The Quarterly Journal of Economics* 135.3, pp. 1209–1318.

Hoynes, Hilary, Marianne Page, and Ann Huff Stevens (2011). "Can targeted transfers improve birth outcomes?: Evidence from the introduction of the WIC program". In: *Journal of Public Economics* 95.7-8, pp. 813–827.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond (2016a). "Long-run impacts of childhood access to the safety net". In: *American Economic Review* 106.4, pp. 903–34.

— (2016b). "Long-run impacts of childhood access to the safety net". In: *American Economic Review* 106.4, pp. 903–34.

Hunt, Jennifer and Marjolaine Gauthier-Loiselle (2010). "How Much Does Immigration Boost Innovation?" In: *American Economic Journal: Macroeconomics* 2.2, pp. 31–56. DOI: 10.1257/mac.2.2.31. URL: https://www.aeaweb.org/articles?id=10.1257/mac.2.2.31.

Hutto, Clayton and Eric Gilbert (2014). "Vader: A parsimonious rule-based model for sentiment analysis of social media text". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 8. 1.

Instituto Nacional de Estadísticas (2020). *Migración Interna en Chile. Censo 2017.* www.ine.cl.

Iyer, Lakshmi et al. (2012). “The power of political voice: women’s political representation and crime in India”. In: *American Economic Journal: Applied Economics* 4.4, pp. 165–93.

Joseph, George et al. (2017). “Underreporting of gender-based violence in Kerala, India: An application of the list randomization method”. In: *World Bank Policy Research Working Paper* 8044.

Khatua, Aparup, Erik Cambria, and Apalak Khatua (2018). “Sounds of silence breakers: Exploring sexual violence on twitter”. In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, pp. 397–400.

Ko, Hansoo, Renata Howland, and Sherry Glied (2020). “The Effects of Income on Children’s Health: Evidence from Supplemental Security Income Eligibility Under New York State Medicaid”. In: *NBER Working Paper* w26639.

LaFree, Gary and Kriss A Drass (1996). “The effect of changes in intraracial income inequality and educational attainment on changes in arrest rates for African Americans and whites, 1957 to 1990”. In: *American Sociological Review*, pp. 614–634.

Lee, Lung-fei and Jihai Yu (2010). “Estimation of spatial autoregressive panel data models with fixed effects”. In: *Journal of econometrics* 154.2, pp. 165–185.

Levitt, Steven D (1998). “Why do increased arrest rates appear to reduce crime: deterrence, incapacitation, or measurement error?” In: *Economic inquiry* 36.3, pp. 353–372.

Levy, Ro’ee (2021). “Social media, news consumption, and polarization: Evidence from a field experiment”. In: *American economic review* 111.3, pp. 831–70.

Levy, Roee and Martin Mattsson (2021). “The effects of social movements: Evidence from# MeToo”. In: *Available at SSRN 3496903*.

Lewis, Ethan (May 2011). “Immigration, Skill Mix, and Capital Skill Complementarity*”. In: *The Quarterly Journal of Economics* 126.2, pp. 1029–1069. ISSN: 0033-5533. DOI: 10.1093/qje/qjr011. eprint: https://academic.oup.com/qje/article-pdf/126/2/1029/5289078/qjr011.pdf. URL: https://doi.org/10.1093/qje/qjr011.

Li, Tianshu, Sonal Pandya, and Sheetal Sekhri (2019). *Repelling Rape: Foreign Direct Investment Empowers Women*. Tech. rep. Working Paper.

Li, Susan (2018). *Multi-Class Text Classification Model Comparison and Selection*. https://towardsdatascience.com/multi-class-text-classification-model-comparison-and-selection-5eb066197568. Accessed: 2018-25-09.

Linos, Natalia et al. (2013). “Influence of community social norms on spousal violence: a population-based multilevel study of Nigerian women”. In: *American journal of public health* 103.1, pp. 148–155.

Lochmann, Alexia, Hillel Rapoport, and Biagio Speciale (2019). "The effect of language training on immigrants' economic integration: Empirical evidence from France". In: *European Economic Review* 113, pp. 265–296.

Ludwig, Jens and Douglas L Miller (2007). "Does Head Start improve children's life chances? Evidence from a regression discontinuity design". In: *The Quarterly journal of economics* 122.1, pp. 159–208.

Manning, Maryann and Janice Patterson (2006). "Lifetime effects: The High/Scope Perry preschool study through age 40". In: *Childhood Education* 83.2, p. 121.

Martén, Linna, Jens Hainmueller, and Dominik Hangartner (2019). "Ethnic networks can foster the economic integration of refugees". In: *Proceedings of the National Academy of Sciences* 116.33, pp. 16280–16285.

Matta, Samer, Michael Bleaney, and Simon Appleton (2021). "The economic impact of political instability and mass civil protest". In: *Economics & Politics*.

McCart, Michael R, Daniel W Smith, and Genelle K Sawyer (2010). "Help seeking among victims of crime: A review of the empirical literature". In: *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies* 23.2, pp. 198–206.

McKague, Kevin, David Wheeler, and Aneel Karnani (2015). "An integrated approach to poverty alleviation: roles of the private sector, government and civil society". In: *The business of social and environmental innovation*. Springer, pp. 129–145.

Metz, Cade (2017). "Tech giants are paying huge salaries for scarce AI talent". In: *The New York Times* 22.

Millán, Teresa Molina et al. (2020). "Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program". In: *Journal of Development Economics* 143, p. 102385.

Miller, Amalia R and Carmit Segal (2019). "Do female officers improve law enforcement quality? Effects on crime reporting and domestic violence". In: *The Review of Economic Studies* 86.5, pp. 2220–2247.

Milligan, Kevin and Mark Stabile (2011). "Do child tax benefits affect the well-being of children? Evidence from Canadian child benefit expansions". In: *American Economic Journal: Economic Policy* 3.3, pp. 175–205.

Ministerio de Desarrollo Social y Familia (2017). *Encuesta de caracterización socioeconómica nacional*. `http://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen`.

Ministerio de Desarrollo Social y Familia (2021). *Encuesta Longitudinal de Primera Infancia 2010-2017*. http://observatorio.ministeriodesarrollosocial.gob.cl/elpi-tercera-ronda.

Ministerio de Educación (2021a). *Datos Abiertos*. https://datosabiertos.mineduc.cl/.

— (2021b). *Requisitos de edades para ingresar al sistema escolar*. https://www.ayudamineduc.cl/ficha/requisitos-de-edades-para-ingresar-al-sistema-escolar.

Ministry of Health (2017). *Chile Crece Contigo cumplió 10 años*. https://www.minsal.cl/presidenta-bachelet-encabezo-el-10-aniversario-del-programa-chile-crece-contigo/f.

Mishra, Ankita, Vinod Mishra, and Jaai Parasnis (2021). "The asymmetric role of crime in women's and men's labour force participation: Evidence from India". In: *Journal of Economic Behavior & Organization* 188, pp. 933–961.

Morrison, Andrew, Mary Ellsberg, and Sarah Bott (2007). "Addressing gender-based violence: a critical review of interventions". In: *The World Bank Research Observer* 22.1, pp. 25–51.

Müller, Karsten and Carlo Schwarz (2020). "From hashtag to hate crime: Twitter and anti-minority sentiment". In: *Available at SSRN 3149103*.

Muralidharan, Karthik and Nishith Prakash (2017). "Cycling to school: Increasing secondary school enrollment for girls in India". In: *American Economic Journal: Applied Economics* 9.3, pp. 321–50.

Nature Index (2021). *Top 25 countries/territories in artificial intelligence*. https://www.natureindex.com/supplements/nature-index-2020-ai/tables/countries. Accessed: 2021-12-28.

OECD (2021). *OECD.stats.Patents by technology.* https://stats.oecd.org/Index.aspx?DataSetCode=PATS_IPC. Accessed: 2021-12-23.

OECD (2021). *Quarterly GDP (indicator).* https://data.oecd.org/gdp/quarterly-gdp.htm.

Ouedraogo, Rasmane and David Stenzel (2021). "The Heavy Economic Toll of Gender-based Violence: Evidence from Sub-Saharan Africa". In: *IMF Working Papers* 2021.277.

Palermo, Tia, Jennifer Bleck, and Amber Peterman (2014). "Tip of the iceberg: reporting and gender-based violence in developing countries". In: *American journal of epidemiology* 179.5, pp. 602–612.

Parkhi, Omkar M, Andrea Vedaldi, and Andrew Zisserman (2015). "Deep face recognition". In: *British Machine Vision Association*.

Pei, Zhuan et al. (2022). "Local polynomial order in regression discontinuity designs". In: *Journal of Business & Economic Statistics* 40.3, pp. 1259–1267.

Peri, Giovanni and Chad Sparber (2011). “Highly educated immigrants and native occupational choice”. In: *Industrial Relations: a journal of economy and society* 50.3, pp. 385–411.

Piyapromdee, Suphanit (2021). “The impact of immigration on wages, internal migration, and welfare”. In: *The Review of Economic Studies* 88.1, pp. 406–453.

Richter, Linda M et al. (2017). “Investing in the foundation of sustainable development: pathways to scale up for early childhood development”. In: *The lancet* 389.10064, pp. 103–118. DOI: http://dx.doi.org/10.1016/S0140-6736(16)31698-1.

Rude, Britta (2022a). “Can we grow with our children? The effects of a comprehensive early childhood development program”. In: *ifo Working Paper*.

— (2022b). “Middle-run impacts of comprehensive early childhood interventions: Evidence from a pioneer program in Chile”. In: *ifo Working Paper*.

Rude, Britta and Yvonne Giesing (2022). “Technological Change and Immigration - A Race for Talent or of Displaced Workers”. In: *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2022: Big Data in Economics*.

Rudé, George E (1954). “Prices, wages and popular movements in Paris during the French Revolution”. In: *The Economic History Review* 6.3, pp. 246–267.

Rumens, Nick and John Broomfield (2012). “Gay men in the police: Identity disclosure and management issues”. In: *Human Resource Management Journal* 22.3, pp. 283–298.

Schmidheiny, Kurt and Sebastian Siegloch (2019). “On event study designs and distributed-lag models: Equivalence, generalization and practical implications”. In.

Serengil, Sefik Ilkin and Alper Ozpinar (2020). “LightFace: A Hybrid Deep Face Recognition Framework”. In: *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, pp. 23–27. DOI: 10.1109/ASYU50717.2020.9259802. URL: https://doi.org/10.1109/ASYU50717.2020.9259802.

Siddique, Zahra (2022). “Media-Reported Violence and Female Labor Supply”. In: *Economic Development and Cultural Change* 70.4, pp. 000–000.

SimpleMaps (2012). *United States Cities Database*. data retrieved from SimpleMaps, https://simplemaps.com/data/us-cities.

Standish, Katerina (2014). “Understanding cultural violence and gender: honour killings; dowry murder; the zina ordinance and blood-feuds”. In: *Journal of Gender Studies* 23.2, pp. 111–124.

Statista (2022). *Leading countries based on number of Twitter users as of October 2021*. data retrieved from Statista, https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/.

Swamy, Vighneswara (2014). “Financial inclusion, gender dimension, and economic impact on poor households”. In: *World development* 56, pp. 1–15.

Temple, Judy A and Arthur J Reynolds (2007). “Benefits and costs of investments in preschool education: Evidence from the Child–Parent Centers and related programs”. In: *Economics of Education Review* 26.1, pp. 126–144.

The World Bank (2018). *10 years of Chile Grows with You.* http://documents1.worldbank.org/curated/en/992351537159031673/pdf/129940-WP-PUBLIC-Chile-Crece-Contigo-10-a%C3%B1os-FINAL-July-2018.pdf.

— (2021). *World Bank Data.* https://data.worldbank.org/. Accessed: 2021-12-23.

— (2022). *Urban population (% of total population).* data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?locations=DE.

Tur-Prats, Ana (2019). “Family types and intimate partner violence: A historical perspective”. In: *Review of Economics and Statistics* 101.5, pp. 878–891.

UN Women (June 21, 2021). “Facts and figures: Ending violence against women”. In: *UN Women.* URL: https://www.unwomen.org/en/what-we-do/ending-violence-against-women/facts-and-figures (visited on 08/11/2021).

UNHCR (Jan. 29, 2022). “Gender-based Violence”. In: *UNHCR.* URL: https://www.unhcr.org/gender-based-violence.html (visited on 01/29/2022).

UNICEF (2018). *Reporte Metodológico. Encuesta Longitudinal de Primera Infancia (III Ronda).* http://observatorio.ministeriodesarrollosocial.gob.cl/storage/docs/elpi/2017/Reporte_Metodologico_ELPI_III.pdf.

United Nations (2022). *Country Classification.* https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESP2022_ANNEX.pdf.

United States Bureau of Justice Statistics (2023). *National Incident-Based Reporting System.* Inter-university Consortium for Political and Social Research.https://doi.org/10.3886/ICPSR37650.v2.

Universidad de Chile (2015). *Informe Resultados Evaluaciones. Segunda Ronda Encuesta Longitudinal de la Primer Infancia.* http://www.crececontigo.gob.cl/wp-content/uploads/2015/12/Resultados_Finales_Test2012-1.pdf.

US Census Bureau (2021). *American Community Survey.* https://www.census.gov/programs-surveys/acs.

Villalobos, Veronica Silva (2011). *Memoria de la Instalación del Sistema de Protección Integral a la Infancia Chile Crece Contigo 2006-2010.* http://www.crececontigo.gob.cl/wp-content/uploads/2015/08/ChCC_MEMORIA.pdf.

Viscusi, W Kip, Joel Huber, and Jason Bell (2011). "Promoting recycling: private values, social norms, and economic incentives". In: *American Economic Review* 101.3, pp. 65–70.

Walby, Sylvia and Philippa Olive (2014). "Estimating the costs of gender-based violence in the European Union". In: *European Institute for Gender Equality*.

Walia, Harsha (2010). "Transient servitude: Migrant labour in Canada and the apartheid of citizenship". In: *Race & Class* 52.1, pp. 71–84.

Webb, Michael (2019). "The impact of artificial intelligence on the labor market". In: *Available at SSRN 3482150*.

Weichselbaumer, Doris (2016). "Discrimination against female migrants wearing headscarves". In: *IZA Discussion Paper*.

Weikart, David P et al. (1970). "Longitudinal Results of the Ypsilanti Perry Preschool Project. Final Report. Volume II of 2 Volumes." In: *ERIC*.

Welsh, Sandy (1999). "Gender and sexual harassment". In: *Annual review of sociology* 25.1, pp. 169–190.

Wen, Jinglin (2021). "Female Mayors and Violence Against Women: Evidence from the US". In.

Williams, Breyon J (2019). "The spillover benefits of expanding access to preschool". In: *Economics of Education Review* 70, pp. 127–143.

World Bank Group (2022). *World Development Indicators*. `https://data.worldbank.org/`.

World Economic Forum (2020). *The Global Social Mobility Report 2020*. `https://www3.weforum.org/docs/Global_Social_Mobility_Report.pdf`.

World Health Organization (2020). *Early Childhood Development*. `https://www.who.int/topics/early-child-development/en/`.

World Inequality Database (2022). *World Inequality Database*. `https://wid.world/`. Accessed: 2022-09-11.

Yilmaz, Okan (2018). "Female autonomy, social norms and intimate partner violence against women in Turkey". In: *The Journal of Development Studies* 54.8, pp. 1321–1337.

Zanzotto, Fabio Massimo (2019). "Human-in-the-loop artificial intelligence". In: *Journal of Artificial Intelligence Research* 64, pp. 243–252.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov (2020). "Political effects of the internet and social media". In: *Annual Review of Economics* 12, pp. 415–438.

Appendix A

Appendix to Chapter 1

A.1 Program Roll-out

The table below details the inclusion of beneficiaries into the program during the early years of ChCC.

Table A.1: Beneficiaries of ChCC (roll-out)

Coverage	2007	2008	2009
Municipalities	159	345 (all)	345 (all)
Pregnant women	47,683	202,729	205,935
Births	40,119	160,643	171,373
Children under 1		168,823	173,733
Children aged 1 to 2		174,286	176,854
Children aged 2 to 4			324,338

Notes: The table presents the roll-out of ChCC over time and by municipality. Source: Adapted from The World Bank (2018).

A.2 Detailed Program Description

A.2.1 Summary of the Program

The social subsystem ChCC is a decentralized program that operates locally through municipal networks (called *Red Comunal*). Children and their mothers start to form part of the social subsystem ChCC during the first prenatal control check-up. From that moment onwards, children are part of the program, with special services offered to them and their families. The services offered start during gestation. They consist of regular health check-ups, parental education programs, the hand-over of materials for stimulation, as well as the assessment of risk factors and the development of personalized health plans. Additionally, pregnant women who are part of the socioeconomically vulnerable population have access to a family subsidy, and home visits. The program also includes a personalized birth-giving process, which is facilitated through a number of actionable items.

ChCC offers a variety of services and benefits to children under five and to their parents. These services range from the handover of didactic materials on how to stimulate children to the introduction of educational group workshops, personalized hospitalization, the development of individual health plans and special services offered to children with disabilities or development lags. It also gives children who are part of the 40 percent most vulnerable population free access to early childhood education.

In the following, I will describe the different components of ChCC in more detail.

A.2.2 Pregnancy and Childbirth

ChCC offers special services to pregnant women. It also significantly improved the birth-giving experience. The program increased the time of prenatal checkups from 20 to 40 minutes.[1] The program also introduced psycho-social risk factors into the risk screening of pregnant women. ChCC introduced the development of a personalized health plan and personalized home visits. These plans are applied to all women who suffer from any kind of risk factor.[2]

Another entry point of ChCC is the guarantee of equal access to information about pregnancy and birth-giving. Families receive a so-called Gestation Guide during their first prenatal check-up.[3] Moreover, ChCC provides the possibility to participate in prenatal workshops targeted at pregnant women and their partners. The workshops consist of six sessions and provide information about birth-giving and child-care, as well as cognitive and emotional support. Also, ChCC introduced the transfer of educational materials about pregnancy and birth-giving to expectant parents. Additionally, ChCC personalized the birth-giving process and launched a campaign with the goal to raise awareness about the importance of being accompanied while giving birth. It introduced a variety of actionable items aiming at the

[1] The so called EPsA (Psycho-social evaluation) is an evaluation of risk factors, such as depression or gender-based violence. ChCC increased the duration of the pregnancy control from 20 minutes to 40 minutes. Pregnant women are subject to the EPsA twice, once during their first pregnancy control and then again during the third gestation trimester.

[2] These women get access to personalized social services through the municipality network ChCC.

[3] The Gestation Guide contains information about the pregnancy, birth-giving, labor rights, and other practical advice.

facilitation of birth-giving. Additionally, ChCC introduced an additional education session with information about the child-bed. In 2008, a nutritional component was developed, called *Purita Mamá*.

A.2.3 Newborns

In 2009, ChCC introduced a component specifically addressing the needs of newborns, called PARN (*Programa de apoyo al recien nacido*). The program consists of in-kind transfers of materials that are useful for the care-taking of a newborn (as oils, creams, a towel, clothes, and a blanket, among others). It also includes educational materials for parents with information on how to take care of newborns.

A.2.4 Health

In 2007, the government of Chile introduced evaluation tools that aim to detect risk factors for the development of children under four.[4] Similarly, ChCC introduced the evaluation of psycho-motor deficits.[5] As part of ChCC the attention of children in hospitals was revisited. ChCC introduced a concept that aimed at minimizing the stress experienced by children hospitalized during early childhood. This involved, among others, the introduction of a special technical orientation of medical staff.

A.2.5 Parental Education

ChCC offers several other group education programs targeted at caregivers, addressing topics such as a child's socio-psychological stimulation, educational child-rearing guidelines, and more. It introduced a variety of workshops targeting socioeconomically vulnerable children. Moreover, ChCC diffuses information as well as materials on child-care for free. These are available through the web portal of ChCC [6], a telephone line, through which it is possible to clarify doubts, a radio program, a TV program, educational booklets, TV campaigns, and a manual of best practices. The goal of these components is to create easy access to experts and informational materials about early child development. In 2008, ChCC launched a special musical program directed at children between zero and five years old.

In 2009, the program *Nadie es perfecto* (Nobody is perfect) was introduced. *Nadie es perfecto* is a workshop series, which consists of six to eight sessions directed at all caretakers.

[4] The risk assessment includes the detection of neurological problems and maternal depression.

[5] The evaluation is conducted through the EEDP (Scale of Psycho-motor Development Evaluation), as well as the TEPSI (Test of Psycho-motor Development).

[6] www.crececontigo.cl

The program was inspired by a similar program in place in Canada. It also involved in-kind transfers directed at the cognitive stimulation of children between zero and five years old.

A.2.6 Early Childhood Education

Another set of actions forming part of ChCC are the ones addressing early childhood education. These policies aim to achieve equal access to early childhood education through increasing its coverage and quality. The goal was to create 70,000 new places in nurseries and 43,000 new places in kindergardens between 2006 and 2010 through the network of preschools and kindergardens *INTEGRA* (in Spanish: Red de Salas Cuna y Jardines Infantiles) as well as Chile's national daycare association *JUNJI* (in Spanish: Junta Nacional de Jardines Infantiles). Moreover, there was an increase in opening hours. Early childhood education facilities also increasingly open during holidays. They also offer parental education, and a special educational program for rural children.

The number of early childhood education facilities with increased opening hours increased from 484 in 2006 to 655 in 2009. Moreover, the number of facilities opening exclusively during the summer holidays augmented from 82 in 2006 to 102 in 2009. Also, ChCC introduced a mobile kindergarden, which reached 187 children in 2009. From 2005 to 2010 *JUNJI* increased its number of daycare centers by 505 percent (from 539 to 3,259) and its number of kindergartens by 92 percent (from 46,990 spots to 85,690 spots). *JUNJI* also introduced a new educational program in its facilities.

A.2.7 Special Services for the Most Vulnerable Children

ChCC comprises special services and benefits offered to children forming part of the vulnerable population in Chile. Health officials develop a personalized action plan addressing deficits and risks detected through thorough professional evaluations. These health action plans consist of a set of psycho-social actionable items targeting both children and their families.[7] In addition to that, ChCC grants special social protection services to families characterized by some form of socioeconomic vulnerability.[8] These special services are the inclusion of socioeconomically vulnerable pregnant women into the the *Unified Family Subsidy* (in Spanish: Subsidio Único Familiar). The program also offers these families prioritized access to social services offered by the public sector.

[7]These actionable items consist of home visits, group educational programs, local stimulation centers, playrooms, among others.

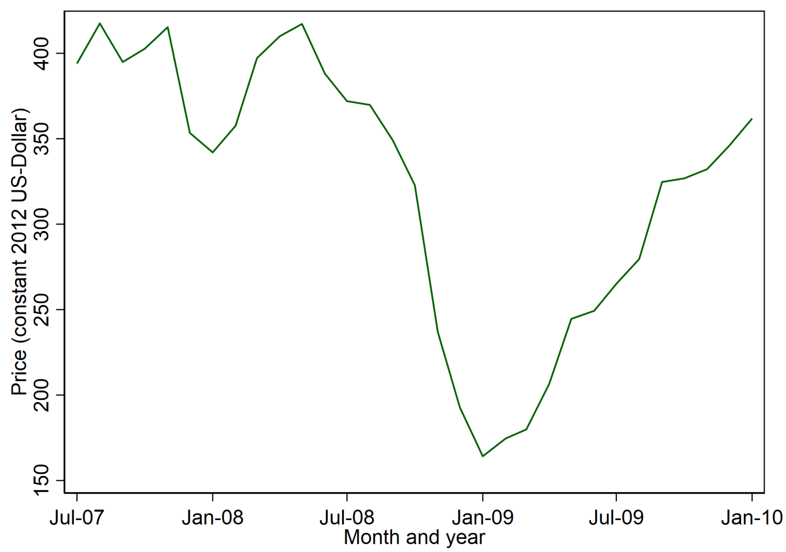
[8]From 2007 to 2009 these targeted the 40 percent most vulnerable, in 2010 to the 50 percent, and in 2011 to the 60 percent most vulnerable.

A.2.8 Services Targeting the General Citizenship

One of ChCC's main goals was to raise the general awareness about the importance of investments in early childhood development. For this purpose, it introduced a website, a free hotline and a radio program targeting the overall Chilean population.

A.3 Copper Prices

Figure A.1: Evolution of monthly copper prices (2007-2009)



Notes: The figure plots the evolution of copper prices on a monthly basis for the period between June 2007 to June 2009.

A.4 Additional Tables

A.4.1 Summary Statistics of ELPI Outcomes

Table A.2: Summary statistic of ELPI Sample (2010-2017)

	Control group	Treatment group
Age in months	92.77	59.43
	(42.01)	(33.17)
Male	0.489	0.488
	(0.500)	(0.500)
Vulnerable	0.425	0.398
	(0.494)	(0.489)
Indigenous	0.111	0.126
	(0.314)	(0.331)
Household members	3.621	3.942
	(2.085)	(2.014)
Share of adults with low education	0.121	0.0998
	(0.173)	(0.158)
No. of employed household members	1.658	1.696
	(0.930)	(0.951)
First survey-round	0.247	0.0742
	(0.431)	(0.262)
Second survey-round	0.278	0.185
	(0.448)	(0.389)
Third survey-round	0.475	0.740
	(0.499)	(0.438)
Observations	12404	19291

Notes: The table shows the descriptive statistics of children included in the ELPI survey by treatment status. Treated children are children born after the implementation of ChCC. *Male*, *Vulnerability*, and *Indigenous* are indicator variables. The information on the respective survey round represents the share of people included in the survey round under consideration. I weight each variable by the survey weights. Source: ELPI 2010, 2012 and 2017.

Table A.3: Summary Statistics of intermediate outcomes

Variable	Mean	Std. Dev.	Min	Max	N
TEPSI	53.387	12.080	19	80	321,110
TVIP	102.772	17.098	55	145	770,323
TADI	51.337	9.390	23	81	521,280
BDS	45.301	7.772	33	90	503,259
HTKS	49.469	10.592	20	91	507,912
CBCL1	57.376	11.159	28	96	888,009
Abnormal height (ECD)	0.210	0.408	0	1	752,548
PSI (Int.)	41.998	35.325	1	99	734,602
PSCS	66.628	10.360	20	85	734,602
CESD	7.219	5.389	0	30	734,662
HOME	11.785	4.375	0	27	1,745,490
Gender-neutral parenting	0.840	0.367	0	1	3,652,612
Space for toys	0.893	0.309	0	1	3,265,159
Learning equipment	0.704	0.456	0	1	3,264,847
Books	0.872	0.334	0	1	3,263,646
Reading (Mom)	0.457	0.498	0	1	727,518
Sharing meals	0.857	0.350	0	1	730,955
Lacking dental care	0.367	0.482	0	1	3,304,557
ECE	0.497	0.500	0	1	3,652,612

Notes: The table shows the summary statistics of the intermediate outcomes investigated in this paper. I restrict the sample to all children born 18 months before and after the roll-out of ChCC as this allows a large enough sample size on either side of the cutoff. I leverage data from the ELPI survey. For this purpose, I pool the survey waves from 2010, 2012, and 2017. *TEPSI* is a score that measures children's psycho-motor development. *TVIP* is a norm-referenced measure of Spanish hearing vocabulary analyzing verbal reasoning, as well as language skills. *TADI* evaluates children ages three months to six years and measures four dimensions of child development: cognition, motor skills, language and socio-emotional development. *BDS* and *HTKS* measure children's executive functioning and *CBCL1* behavioral abnormalities. *PSI* is an index measuring parental stress, *PSCS* is a perceived self-confidence scale, *CESD* is the Center for Epidemiologic Studies Depression Scale, and *HOME* is the Home Observation Measurement of the Environment Index. All other variables are dummy variables. *ECE* stands for early childhood education and is the share of children attending an ECE facility. I weight each variable by survey weights. Source: ELPI (2010-2017).

A.4.2 Alternative Cutoff Windows

To analyze if the number of windows around the cutoff drive my results, I repeat my analysis using an alternative cutoff window. While the assumption on a similarity of unobserved characteristics is most plausible in smaller cutoff windows, there are downsides to restricting the sample to few windows. I might lose important information on the variation or trends in the data when relying on a window length of less than five. For this reason, I validate my findings considering a larger cutoff window. I choose a cutoff window of 20 months for my robustness check. This means that I consider all students born ten months before and after ChCC's implementation.

Table A.4 shows that students systematically differ from each other on observable characteristics when considering the larger cutoff window. The p-values are zero in the case of gender and socioeconomic vulnerability. Consequently, it is not possible to reject the null hypotheses of treatment effects on observable covariates in the larger cutoff window. I confirm this by a t-test on baseline characteristics. Table A.5 shows that the related coefficient p-values on gender and socioeconomic vulnerability are zero. The systematic differences in individual controls are an important caveat and might confound my results in the larger cutoff window. The possibility of significant unobservable confounding factors might be more plausible under this model specification.

Turning attention to results reported on the three schooling outcomes, the local RD approach in the larger window confirms my findings from the optimal cutoff window. Table A.5 provides evidence that exposure to ChCC significantly improves schooling outcomes. Column 1 illustrates that the program leads to increases in standardized math scores of 0.347 points, in standardized reading scores of 2.987 points, and in grade point averages of 0.03 points. When compared to the local RD estimators in the optimal window length, the point estimates are similar in terms of magnitude in the case of grade point averages, but smaller in the case of standardized test scores. Especially the coefficient on standardized math scores more than halves when compared to the baseline estimator. Furthermore, Column 2 shows that the p-value on standardized math scores increases in the larger window. The point estimate associated with standardized math scores is only significant at the 2.5 percent significance level. In contrast, the coefficient p-values on standardized reading scores and grade point averages remain at zero and are highly significant.

Table A.4: Baseline characteristics (20 periods around cutoff)

	Control mean	Treatment mean	T-test p
Female	0.50	0.51	0.00
Vulnerable student	0.73	0.74	0.00
Rural	0.10	0.10	0.60
Observations	172,695	186,707	359,402

Notes: The table shows the baseline characteristics of the control and treatment group in the 20 periods around the cutoff. Column 1 reports the mean of the control group (those children born before the roll-out of ChCC). Column 2 reports the mean of the treatmetn group (those children born after the roll-out of ChCC). Column 3 reports the coefficient p-values of a two-sided t-test. Source: SIMCE (2015-2018), MINEDUC (2015-2018).

Table A.5: Local RD effect of ChCC on schooling outcomes in the 20 periods around the cutoff

		RD Estimate	P-Value	N (left)	N (right)
1	Standardized math score	0.357	0.025	172,695	186,707
2	Standardized reading score	2.987	0.000	172,695	186,707
3	Grade point averages	0.030	0.000	172,695	186,707
4	Gender	0.008	0.000	172,695	186,707
5	Vulnerability	0.020	0.000	172,695	186,707
6	Rural	0.000	0.829	172,695	186,707

Notes: The table shows the local RD effect of ChCC in the 20 months around the threshold. This means that the estimation considers all students born ten months before and after the roll-out of ChCC as well as those born during the roll-out. The first column shows the point estimates of exposure to ChCC on the three schooling outcomes and observable covariates. Column 2 presents the related coefficient p-values. Column 3 and 4 show the number of observations N on each side of the cutoff. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

A.4.3 Additional Heterogeneity Analysis

The below table shows the local randomization approach of estimating the impact of ChCC on schooling outcomes by subgroups in the 20 periods around the cutoff. The findings confirm my results in the optimal window around the cutoff.

Table A.6: Local RD effect of ChCC in the 20 periods around the cutoff on schooling outcomes by groups

	Group	Boys	Girls	Vulnerability	Non-vulnerability
	Panel 1: Standardized math scores				
1	RD Estimate	1.056	-0.274	-0.023	1.435
2	P-Value	0.000	0.218	0.902	0.000
3	Panel 2: Standardized reading scores				
4	RD Estimate	3.597	2.237	2.437	4.545
5	P-Value	0.000	0.000	0.000	0.000
6	Panel 3: Grade point averages				
7	RD Estimate	0.040	0.020	0.030	0.040
8	P-Value	0.000	0.000	0.000	0.000
9	N (left)	86,562	86,133	125,975	46,720
10	N (right)	92,026	94,681	137,991	48,716

Notes: The table shows the local RD effect of ChCC in the 20 months around the cutoff window. This means that the estimation considers all students born ten months before and after the roll-out of ChCC as well as those born during the roll-out. I first report the RD estimator and then the coefficient p-value. Panel 1 shows results for standardized math scores, Panel 2 for standardized reading scores, and Panel 3 for grade point averages. Column 1 reports results for boys, Column 2 for girls, Column 3 for socioeconomic vulnerable children and Column 4 for socioeconomic non-vulnerable children. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

The below tables show the parametric estimation of exposure to ChCC in the larger window, using 20 periods around the cutoff. The results confirm the findings from the non-parametric estimation.

Table A.7: Local RD effect of ChCC in 20 periods around cutoff window on standardized math scores by groups (p=1)

		RD Estimate	P-Value	N (left)	N (right)	Covariates
1	Boys	3.087	0.000	86,562	92,026	Yes
2	Girls	3.242	0.000	86,133	94,681	Yes
3	Vulnerability	2.899	0.000	12,5975	137,991	Yes
4	Non-vulnerability	3.915	0.000	46,720	48,716	Yes

Notes: The table shows the local RD effect of ChCC on standardized math scores in the 20 months around the cutoff window. This means that the estimation considers all students born ten months before and after the roll-out of ChCC as well as those born during the roll-out. I assume a local polynomial order of degree one. Column 1 reports the RD coefficient, Column 2 the coefficient p-values. N is the number of observations on each side of the eligibility cutoff. I first report results for boys in Row 1, for girls in Row 2, for vulnerable children in Row 3 and for non-vulnerable children in Row 4. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

Table A.8: Local RD effect of ChCC in 20 periods around cutoff window on standardized reading scores by groups (p=1)

		RD Estimate	P-Value	N (left)	N (right)	Covariates
1	Boys	3.927	0.000	86,562	92,026	Yes
2	Girls	3.798	0.000	86,133	94,681	Yes
3	Vulnerability	3.498	0.000	125,975	137,991	Yes
4	Non-vulnerability	5.023	0.000	46,720	48,716	Yes

Notes: The table shows the local RD effect of ChCC in the 20 months around the cutoff window on standardized reading scores. This means that the estimation considers all students born ten months before and after the roll-out of ChCC as well as those born during the roll-out. I assume a local polynomial order of degree one. Column 1 reports the RD coefficient, Column 2 the coefficient p-values. N is the number of observations on each side of the eligibility cutoff. I first report results for boys in Row 1, for girls in Row 2, for vulnerable children in Row 3 and for non-vulnerable children in Row 4. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

Table A.9: Local RD effect of ChCC in 20 periods around cutoff window on grade point averages by groups (p=1)

		RD Estimate	P-Value	N (left)	N (right)	Covariates
1	Boys	0.042	0.000	86,562	92,026	Yes
2	Girls	0.043	0.000	86,133	94,681	Yes
3	Vulnerability	0.039	0.000	125,975	137,991	Yes
4	Non-vulnerability	0.054	0.000	46,720	48,716	Yes

Notes: The table shows the local RD effect of ChCC on grade point averages in the 20 months around the cutoff window. This means that the estimation considers all students born ten months before and after the roll-out of ChCC as well as those born during the roll-out. I assume a local polynomial order of degree one. Column 1 reports the RD coefficient, Column 2 the coefficient p-values. N is the number of observations on each side of the eligibility cutoff. I first report results for boys in Row 1, for girls in Row 2, for vulnerable children in Row 3 and for non-vulnerable children in Row 4. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

A.4.4 Analysis of intermediate outcomes by subgroups

The below table shows the Intention-To-Treat (ITT) effect of ChCC on intermediate outcomes by subgroups.

Table A.10: The impact of ChCC on intermediate outcomes (boys)

Variable	Bandwidth=4 RD estimator	P-value	Bandwidth=20 RD estimator	P-value
TEPSI	-2.072	0.044	-2.532	0.014
TVIP	-0.046	0.971	1.027	0.389
TADI	1.133	0.198	1.346	0.109
BDS	-0.137	0.834	-0.086	0.893
HTKS	0.297	0.761	0.362	0.702
CBCL1	-1.386	0.047	-0.820	0.234
Abnormal height (ECD)	-0.014	0.616	0.011	0.685
PSI (Int.)	2.851	0.371	2.456	0.423
PSCS	-0.867	0.377	-0.627	0.495
CESD	-0.249	0.616	-0.048	0.920
HOME	0.119	0.603	0.012	0.958
Gender-neutral parenting	-0.003	0.846	-0.008	0.548
Space for toys	0.011	0.372	-0.003	0.830
Learning equipment	0.040	0.021	0.055	0.001
Books	0.025	0.061	0.038	0.003
Reading (Mom)	-0.009	0.793	0.076	0.018
Sharing meals	0.002	0.933	-0.009	0.668
Lacking dental care	0.011	0.533	0.040	0.023
ECE	0.018	0.292	0.049	0.004

Notes: The table shows intention-to-treat (ITT) effects of ChCC on intermediate outcomes for boys. I restrict the sample to the optimal window. The first two columns report results in a window length of four while the last two columns report them in a window length of 20. I leverage data from the ELPI survey. For this purpose, I pool survey waves from 2010, 2012, and 2017. *TEPSI* is a score that measures children's psycho-motor development. *TVIP* scores are a norm-referenced measure of Spanish hearing vocabulary analyzing verbal reasoning, as well as language skills. *TADI* scores evaluates children ages three months to six years and measures four dimensions of child development: cognition, motor skills, language and socio-emotional development. *BDS* and *HTKS* measure children's executive functioning and *CBCL1* behavioral abnormalities. *PSI* is an index measuring parental stress, *PSCS* is a perceived self-confidence scale, *CESD* is the Center for Epidemiologic Studies Depression Scale, and *HOME* is the Home Observation Measurement of the Environment Index. All other variables are dummy variables. *ECE* stands for early childhood education. Source: ELPI (2010-2017).

Table A.11: The impact of ChCC on intermediate outcomes (girls)

Variable	Bandwidth=4 RD Estimate	P-Value	Bandwidth=20 RD Estimate	P-Value
TEPSI	-1.549	0.156	-1.041	0.344
TVIP	-0.188	0.877	0.943	0.429
TADI	0.101	0.906	0.264	0.747
BDS	1.032	0.144	0.959	0.160
HTKS	2.008	0.051	2.303	0.020
CBCL1	0.003	0.997	0.653	0.341
Abnormal height (ECD)	-0.035	0.213	-0.042	0.127
PSI (Int.)	-7.289	0.016	-6.167	0.038
PSCS	2.145	0.013	2.281	0.008
CESD	-0.936	0.041	-1.094	0.016
HOME	0.047	0.841	0.045	0.842
Gender-neutral parenting	0.014	0.302	0.001	0.910
Space for toys	0.002	0.894	-0.016	0.146
Learning equipment	0.026	0.115	0.026	0.109
Books	0.007	0.521	0.018	0.129
Reading (Mom)	-0.037	0.273	-0.005	0.889
Sharing meals	-0.018	0.453	0.001	0.979
Lacking dental care	0.011	0.513	0.031	0.060
ECE	-0.006	0.718	0.009	0.595

Notes: The table shows ITT effects of ChCC on intermediate outcomes for girls. I restrict the sample to the optimal window. The first two columns report results in a window length of four while the last two columns report them in a window length of 20. I leverage data from the ELPI survey. For this purpose, I pool survey waves from 2010, 2012, and 2017. *TEPSI* is a score that measures children's psycho-motor development. *TVIP* scores are a norm-referenced measure of Spanish hearing vocabulary analyzing verbal reasoning, as well as language skills. *TADI* scores evaluates children ages three months to six years and measures four dimensions of child development: cognition, motor skills, language and socio-emotional development. *BDS* and *HTKS* measure children's executive functioning and *CBCL1* behavioral abnormalities. *PSI* is an index measuring parental stress, *PSCS* is a perceived self-confidence scale, *CESD* is the Center for Epidemiologic Studies Depression Scale, and *HOME* is the Home Observation Measurement of the Environment Index. All other variables are dummy variables. *ECE* stands for early childhood education. Source: ELPI (2010-2017).

Table A.12: The impact of ChCC on intermediate outcomes (socioeconomic non-vulnerable children)

Variable	Bandwidth=4 RD Estimate	P-Value	Bandwidth=20 RD Estimate	P-Value
TEPSI	-2.434	0.026	-2.456	0.025
TVIP	0.294	0.824	1.913	0.127
TADI	0.841	0.338	1.432	0.093
BDS	-0.250	0.732	0.523	0.466
HTKS	-0.075	0.944	1.001	0.332
CBCL1	-0.513	0.481	0.550	0.440
Abnormal height (ECD)	-0.015	0.620	-0.023	0.433
PSI (Int.)	-5.893	0.038	-3.602	0.191
PSCS	1.314	0.121	1.526	0.060
CESD	-0.928	0.037	-0.699	0.109
HOME	0.185	0.430	0.152	0.502
Gender-neutral parenting	0.007	0.572	-0.008	0.488
Space for toys	0.035	0.001	0.026	0.008
Learning equipment	0.035	0.018	0.052	0.000
Books	0.005	0.632	0.016	0.142
Reading (Mom)	-0.020	0.579	0.032	0.353
Sharing meals	-0.006	0.802	-0.006	0.786
Lacking dental care	-0.006	0.724	0.027	0.101
ECE	0.017	0.312	0.034	0.035

Notes: The table shows ITT effects of ChCC on intermediate outcomes for socioeconomic non-vulnerable children. I restrict the sample to the optimal window. The first two columns report results in a window length of four while the last two columns report them in a window length of 20. I leverage data from the ELPI survey. For this purpose, I pool survey waves from 2010, 2012, and 2017. *TEPSI* is a score that measures children's psycho-motor development. *TVIP* scores are a norm-referenced measure of Spanish hearing vocabulary analyzing verbal reasoning, as well as language skills. *TADI* scores evaluates children ages three months to six years and measures four dimensions of child development: cognition, motor skills, language and socio-emotional development. *BDS* and *HTKS* measure children's executive functioning and *CBCL1* behavioral abnormalities. *PSI* is an index measuring parental stress, *PSCS* is a perceived self-confidence scale, *CESD* is the Center for Epidemiologic Studies Depression Scale, and *HOME* is the Home Observation Measurement of the Environment Index. All other variables are dummy variables. *ECE* stands for early childhood education. Source: ELPI (2010-2017).

Table A.13: The impact of ChCC on intermediate outcomes (socioeconomic vulnerable children)

Variable	Bandwidth=4 RD Estimate	P-Value	Bandwidth=20 RD Estimate	P-Value
TEPSI	-1.476	0.231	-2.010	0.110
TVIP	1.032	0.461	1.492	0.270
TADI	1.382	0.194	1.440	0.152
BDS	1.035	0.112	0.150	0.823
HTKS	2.541	0.026	1.956	0.080
CBCL1	-0.313	0.724	-0.727	0.392
Abnormal height (ECD)	-0.011	0.745	-0.009	0.791
PSI (Int.)	2.608	0.452	0.108	0.974
PSCS	-0.237	0.819	-0.033	0.973
CESD	-0.147	0.775	-0.448	0.379
HOME	-0.072	0.787	-0.131	0.601
Gender-neutral parenting	0.007	0.629	0.008	0.602
Space for toys	-0.030	0.036	-0.052	0.000
Learning equipment	0.029	0.139	0.025	0.187
Books	0.030	0.039	0.042	0.003
Reading (Mom)	-0.026	0.526	0.036	0.351
Sharing meals	-0.009	0.743	-0.010	0.712
Lacking dental care	0.043	0.029	0.054	0.004
ECE	-0.002	0.896	0.022	0.234

Notes: The table shows ITT effects of ChCC on intermediate outcomes for socioeconomic vulnerable children. I restrict the sample to the optimal window. The first two columns report results in a window length of four while the last two columns report them in a window length of 20. I leverage data from the ELPI survey. For this purpose, I pool survey waves from 2010, 2012, and 2017. *TEPSI* is a score that measures children's psycho-motor development. *TVIP* scores are a norm-referenced measure of Spanish hearing vocabulary analyzing verbal reasoning, as well as language skills. *TADI* scores evaluates children ages three months to six years and measures four dimensions of child development: cognition, motor skills, language and socio-emotional development. *BDS* and *HTKS* measure children's executive functioning and *CBCL1* behavioral abnormalities. *PSI* is an index measuring parental stress, *PSCS* is a perceived self-confidence scale, *CESD* is the Center for Epidemiologic Studies Depression Scale, and *HOME* is the Home Observation Measurement of the Environment Index. All other variables are dummy variables. *ECE* stands for early childhood education. Source: ELPI (2010-2017).

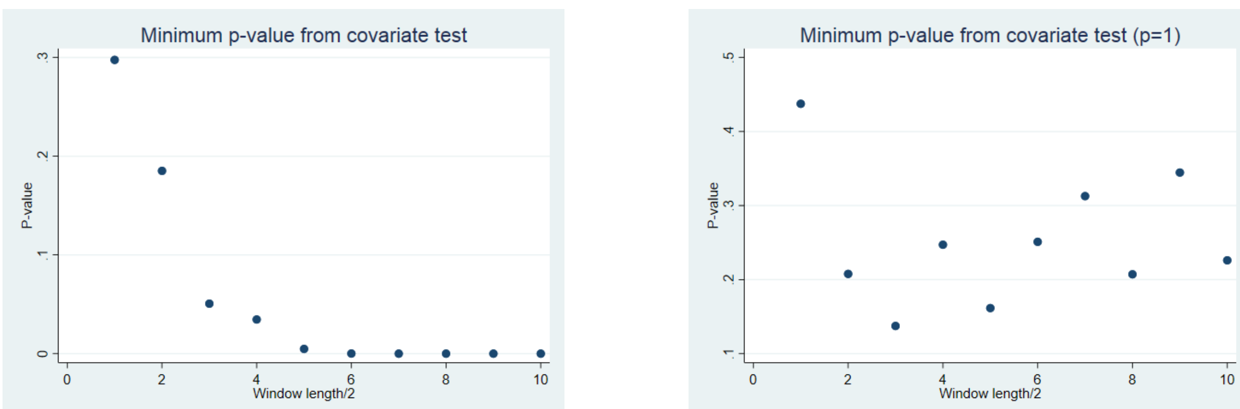
A.5 Additional Figures

A.5.1 Optimal Window Selection

The left graph of Figure A.2 shows the optimal window selection when abstracting from polynomial orders. The optimal bandwidth is four. This result suggests that the optimal local randomization estimation should include the two birth cohorts before and after ChCC's roll-out.

The right graph of Figure A.2 repeats the optimal window selection when including a linear slope. Although the optimal bandwidth resulting from the graph is the minimum window, I consider a window length of 20 for the linear local randomization approach, because it might be challenging to estimate the slope coefficient with only one data point on each side of the cutoff.

Figure A.2: Optimal window selection under a polynomial order of 0 and 1



Notes: The graph shows the optimal window selection for the local randomization approach. For a detailed overview of the methodology see Cattaneo, Titiunik, and Vazquez-Bare (2016). I include the following three variables for the covariance balance tests: a child's gender, a dummy variable for socioeconomic vulnerability as well as if the child assists a rural or urban school. The covariate balance test uses a large-sample approximation. The left graph shows the results under an estimation assuming a polynomial order of zero. The right graph shows the results for an estimation assuming a polynomial order of one.

A.5.2 Regression Discontinuity Plots

The figures below show the RD plots of a local randomization approach without slopes in the optimal window of bandwidth four. The cutoff is zero and refers to the date on which ChCC was rolled out.

Figure A.3: RD Plot (Standardized math scores)

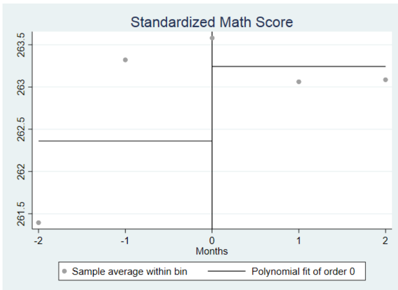


Figure A.4: Rd Plot (Standardized reading scores)

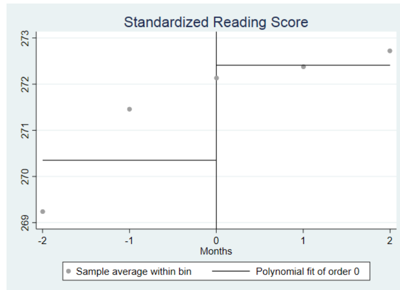
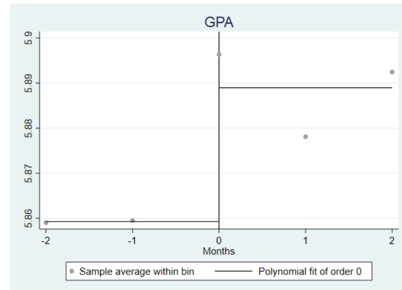


Figure A.5: RD Plot (Grade point averages)



Notes: The figures above show the local randomization design plots for schooling outcomes. The left panel shows the plot for standardized math scores, the middle panel the one for standardized reading scores, and the right panel the one for grade point averages. I restrict the periods shown to the optimal window length, namely four periods. This means that the figures show the average values of schooling outcomes for all children born two months previous to the roll-out of ChCC to two months after its roll-out. The black horizontal line features the threshold of the local RD approach, namely zero. Source: SIMCE (2015-2018) and MINEDUC (2015-2018). For details on the estimation procedure see Cattaneo, Titiunik, and Vazquez-Bare (2016).

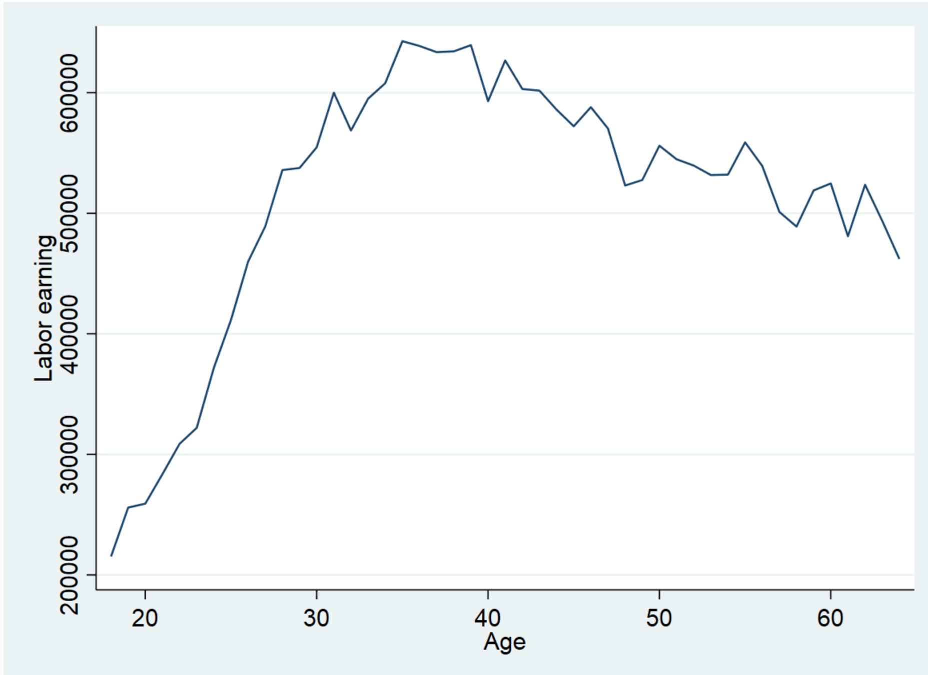
A.6 Appendix G - Detailed Cost-Benefit Analysis

To calculate the program's costs per participant I take advantage of data provided by the government of Chile. I have data on the program's costs per component per year as well as on the number of units benefiting from each component. Depending on the program component these units are children, pregnant women, municipalities or newborns. I restrict my period to the years 2007 to 2017, as the program's target group was expanded in 2017. The total costs of the program for all units amount to 236,472.2 US-Dollar for the period 2007 to 2017. I then calculate the average unit cost per year and convert these values to US-Dollars using data on exchange rates published by the OECD for each year. I then sum up the costs for each respective year from 2007 to 2017. Next, I divide the sum by the number of years. This gives me the average unit cost per year for the period 2007 to 2017. The average unit cost per year is 23,647.2 US-Dollar.

To calculate the marginal willingness to pay for ChCC by program participant I calculate the present value of lifetime earnings in Chile. I take advantage of data published by the Ministry of Social Protection, namely the socioeconomic survey (CASEN) from 2017.

I first calculate the mean labor income of all individuals between 18 and 65 years old in 2017. The results are shown in figure A.6. I assume that this distribution is representative for the average lifetime earning distribution in Chile. I then calculate the present value of this earning stream. I assume a discount rate of 3 percent. I then convert the values to 2017-US-Dollar, taking the average exchange rate for 2017 from data on exchange rates

Figure A.6: Average lifetime earnings in Chile (2017 in CL)



Notes: The graph shows the average lifetime earnings in Chile in Chilean Pesos. The x-axis represents the age in years and the y-axis the labor earning in Chilean Pesos. Source: Ministerio de Desarrollo Social y Familia (2017)

published by the OECD. This results in an average present value of lifetime earnings of 220,312.4 US-Dollar.

Next, I equally distribute the average per year program unit cost of 23,647.2 US-Dollar across a typical lifetime of an individual. I then calculate the present value of this cost stream, which is 12,864.5 US-Dollar. This is 6 percent of the average present value of lifetime earning in Chile. Consequently, participants would need to increase their earnings by 6 percent over the course of the life to meet the program costs of ChCC.

Work by French et al. (2015) shows that a 1 percent increase in the GPA leads to an average increase of around 12 to 14 percent in earnings. The average GPA in my sample is 5.8 (see table 1.1). Based on the different model specifications investigated in this paper, the average impact of ChCC on grade point averages is approximately 0.3. This is an increase of 0.5 percent over the mean grade point average. The equivalent increase in income would therefore be approximately 7.5 percent. From this information I create the post-program average income flow, adding 7.5 percent to the average income per age year. I then take the net present value of this post-program income flow. A 7.5 percent increase in the lifetime earning leads to a difference in the present value of lifetime earnings between the pre- and post-program world of 16,523.4 US-Dollar per participant and an additional present value of tax revenues of 1,156.6 US-Dollar per participant. The average income tax in Chile was 7.0 percent in 2019[9].

The marginal value of public funds is equal to the ratio of participants' marginal willingness to pay for the program and the initial program costs (costs minus fiscal externalities). I therefore divide the difference in the pre- and post-program NPLE by the difference between the costs per participant and the fiscal revenue generated through the program. The MVPF is then 1.41 per participant.

[9] `https://www.oecd.org/tax/tax-policy/taxing-wages-chile.pdf`

Appendix B

Appendix to Chapter 2

B.1 Twitter Dataset Creation

We access our Twitter tweets by creating an Academic Developer Account and accessing tweets via the Twitter Full Archive Search API V2. The API has a rate limit of 10 million tweet per month and 150,000 tweets per 15 minutes. We take advantage of the Twarc2 command line tool and Python library. We define a customized search query, restricting our time frame to the year 2017 and certain keywords. The resulting data is organized in Json (JavaScript Object Notation) Objects, such as a *User* object or a *Tweet* object. Each object comes along with *attributes* describing the Json Object, such as the author, the actual message, a unique ID, a timestamp of when it was posted, and sometimes geo metadata about the location of the user or tweet. There are also *entity* objects associated with some tweets, such as hashtags, mentions, media, and links. A single tweet can have up to 150 attributes coming along with the actual text. There are four overall JSON Keys: Data, Includes, Error, Meta. Each comes with several nested JSON Objects.

In order to narrow down our keywords on which we filter our API query, we ask ourselves the question of how to best proxy the conversation on GBV on Twitter. For this purpose, we identify 10 of the biggest movements related to GBV on Twitter. We then extract all tweets on the first 4 weeks of each of the 10 movements. We select the following movements:

- *#aufschrei*: 24th of Jan. 2013, 39,130 tweets in 1 month
- *#yesallwomen*: 24th of May 2014, 330,428 tweets in 1 month
- *#rapecultureiswhen*: 25th of May 2014, 6,698 tweets in 1 month
- *#WhyIStayed*: 8th of Sept. 2014, 37,915 tweets in 1 month

Figure B.1: Top 500 most mentioned hashtags in the first month of 10 Twitter movements

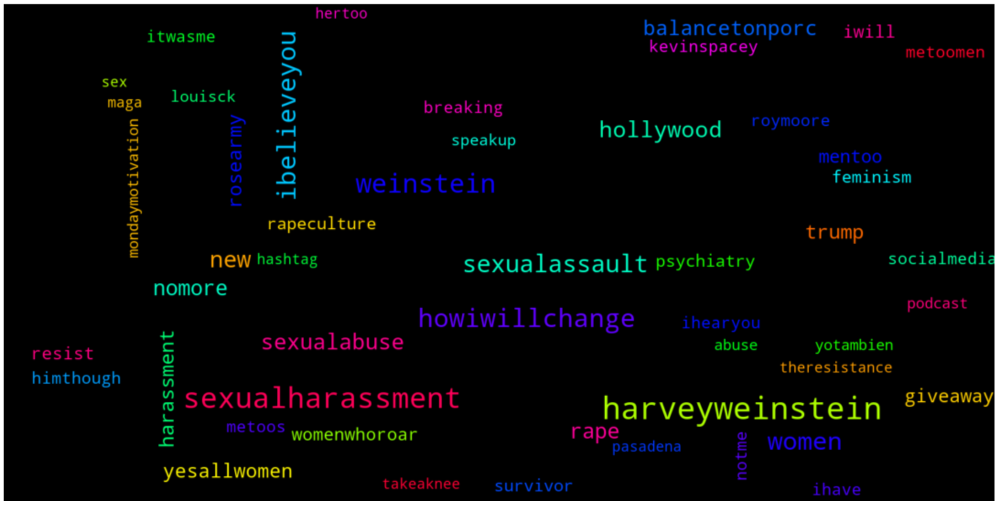
Notes: The graph plots a word cloud of the Top 500 most mentioned hashtags in the first month of 10 Twitter social movements. The size of each word represents its relative occurrence when compared to all other words in the sample. Source: Twitter data.

- *#everydaysexism*: 16th of April 2015, 9,175 tweets in 1 month
- *#NoWomanEver*: 18th of June 2016, 15,643 tweets in 1 month
- *#notokay*: 7th of October 2016, 18,885 tweets in 1 month
- *#metoo*: 14th of October 2017, 278,769 tweets in 1 month
- *#MeAt14*: 10th of November 2017, 7,653 tweets in 1 month
- *#whyididntreport*: 21st of September 2018, 68,707 tweets in 1 month

We end up with a total of 813,003 tweets. From this sample, we extract a list of all hashtags mentioned in these close to one million tweets and end up with a list of 73,430 unique hashtags. Figure B.1 shows the 500 most used hashtags during the first month of each of these 10 debates. Figure B.2 further restrict this list to the 50 most used hashtags.

The figures illustrates that not all of these hashtags uniquely identify topics related to GBV. One example is the hashtag *#hollywood*. If we would include this hashtag in our query as a keyword, we would get a large variety of tweets not related to GBV but other topics. We therefore decide to only include hashtags in our query, which answer the following question: "*looking at this hashtag standing on its own, does it uniquely identify as a hashtag related to GBV?*" with yes. This question is therefore our first selection criteria. Going through all 73,430 unique hashtags and answering the above question for each one of them would take

Figure B.2: Top 50 most mentioned hashtags in the first month of 10 Twitter movements



Notes: The graph plots a word cloud of the Top 50 most mentioned hashtags in the first month of 10 Twitter social movements. The size of each words represents its relative occurrence when compared to all other words in the sample. Source: Twitter data.

a very long time. This is why we decide to train an algorithm, which conducts this task for us. We train our algorithm based on all hashtags mentioned in the first month of the *metoo* movement, a total of 32,487 different hashtags. This means that we go through each one of the 32,487 different hashtags (and 278,769 tweets) and manually code the data. When doing this, we notice that we can come up with 10 overall topics categorizing these hashtags, which are:

- **GBV/Sexism:** #sexualharassment, #YesAllWomen, #SexualAssault, #rapeculture, #sexism, #nomeansno, #vaw, #domesticviolence
- **Misogynie:** #misogyny, #misogynist, #MisogynyIsReal, #antimisogynoir, #endmisogyny
- **Support:** #IBelieveYou, #IHearYou, #BelieveWomen, #believesurvivors, #IStandWithYou
- **Gender Equality:** #GenderEquality, #GenerationEquality, #inequality, #Womensrights, #FuckBoysWillBeBoys
- **Feminism:** #Feminism, #feminist, #howtospotafeminist, #ThisIsWhyINeedFeminism, #FeministFriday
- **Non-White Feminism:** #solidarityisforwhitewomen, #whitefeminism, #Womanofcolor, #BelieveBlackWomen

- **Anti-Feminism:** #womenagainstfeminism, #antifeminismus, #NotMyFeminism, #FeminismIsCancer
- **Masculinity:** #teachourboys, #menneed2dobetter, #mencanendrape, #iwillteachmysonbetter, #fragilemasculinity
- **Misogynist, silencing:** #GoGetRaped, #ShutUp, #AttentionWhores, #GetOverIt, #Boyswillbeboys
- **Others:** #Hollywood, #news, #assault, #women, #psychiatry, #giveaway

We also apply an unsupervised machine learning algorithm, the *Latent Dirichlet Allocation* (LDA), in order to check if manual coding is needed or can be done by an algorithm in the first place. We conclude that the unsupervised machine learning algorithm does a fairly poor job, also when varying the number of topics. Therefore, we rely on our supervised machine learning algorithm.

As a next step we use the manually coded *#metoo* hashtags to train the remaining 534,234 tweets (40,943 hashtags) on the remaining 9 Twitter movements. We split our Training data (the 32,487 manually coded *metoo* hashtags) into a training and testing dataset by a 50-50-ratio. We feed the training data (16,243 hashtags) into the model for training. We test the performance of the model through running the model on the test data (the remaining 16,243 hashtags). We then measure the performance of the model through comparing the *predicted* categories of the *testing* data with the manually coded categories (Model Accuracy), but also based on alternative performance measures, such as the Precision Score, Recall Score, F1-Score, and Confusion Matrix. We then run the model on the new, unseen, not hand-coded data (the 40,943 remaining hashtags). We also conduct a cross validation (10-Fold Cross Validation) and tune our model parameters through a grid search. We spot-check a variety of different algorithms and compare their performance to each other:

- Multinomial Naive Bayes classifier
- K-Nearest Neighbor Classification
- Regularized Logistic Models
- Support Vector Machine
- Stochastic Gradient Descent
- Decision Tree Classifier

- Ensemble Methods:
  - Random Forest Classifier
  - AdaBoost Classifier

We encounter an Imbalanced Classification Problem as more than 90 percent of our hashtags belong to one category, our Garbage Category (see Figure B.3). In order to address this problem, we oversample the minority classes (all but "Other"). This means that we draw random samples (Fraction > 1) from each of the minority classes with equal probability weighting (see Figure B.4).

Figure B.3: Handcoded tweets by category

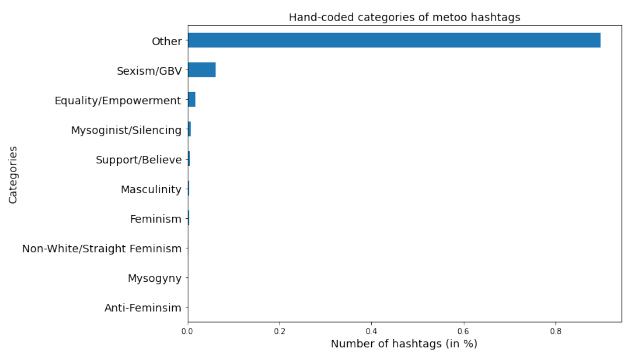
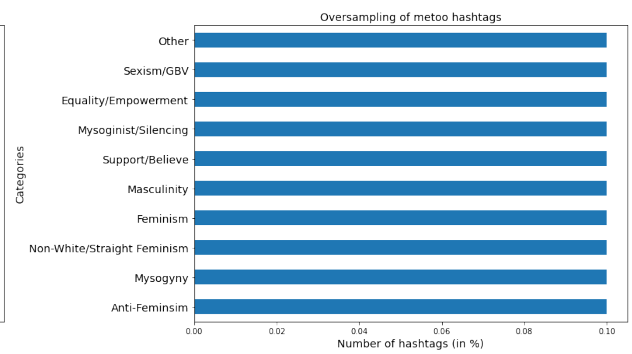


Figure B.4: Oversampled tweets by category

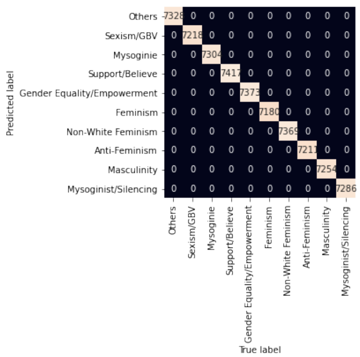


Notes: The left panel shows the share of hashtags used within the first month of the *#metoo* movement in October 2017 in 10 manually defined categories. The right panel shows the same share after oversampling the minority categories. Source: Twitter data.

Figure B.5 to B.12 present the confusion matrices and accuracy scores. The fastest algorithms are (in descending order) the Stochastic Gradient Descent (SGD), Decision Tree, Regularized logistic regression and Naive Bayes Classifier. Based on our performance measures, we choose two different classifier for our prediction: SGD alias linear SV and Logistic Regression. The SGD Classifier is characterized by high accuracy, high speed and is widely accepted as one of the best text data classifiers (Li, Susan 2018). The Logistic Regression Classifier is characterized by high accuracy, high speed, easy to interpret and use, and is widely used in economics.

Based on our different performance measures, the Linear Support Vector Machine is the best performing classifier for our underlying classification problem. We then run the linear SVM algorithm on the remaining 40,943 hashtags that are not hand-coded. We evaluate the performance of the prediction made by the algorithm on the unlabeled data by comparing a ten percent sample of the predicted classes to manually coded classifications of this same sample done by an independent research assistant.

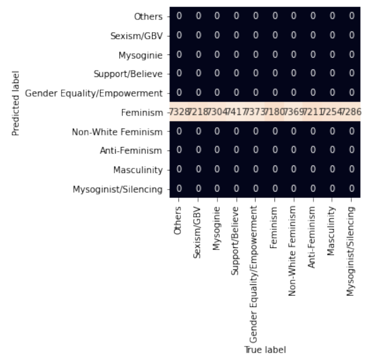
Figure B.5: Regularized logistic regression



[a]

[a]Penalty is elastic net with L-ratio of 0.5. The solver is the saga optimizer. Accuracy is 100.

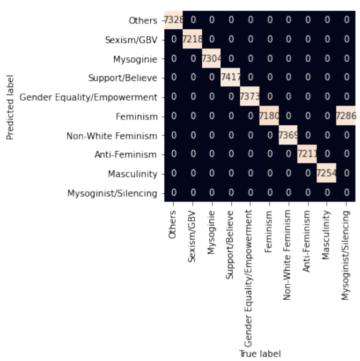
Figure B.6: Support Vector Machine



[a]

[a]Kernel is Rbf, gamma is auto, L-2 panelty. Accuracy is 9.8.

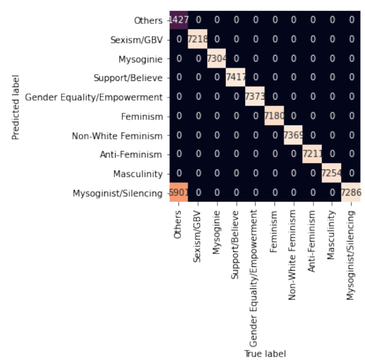
Figure B.7: Naive Bayes



[a]

[a]All default values. Accuracy is 90.

Figure B.8: KNN Classifier

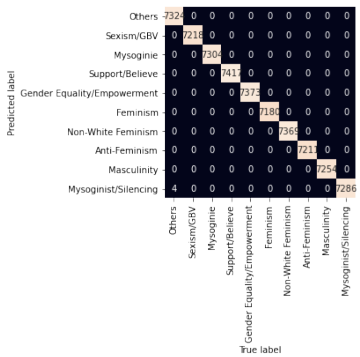


[a]

[a]N=27 (Square root of test data lenght), Metric is the Euclidean distance. Rest is default values. Accuracy is 91.9

After applying our supervised learning algorithm, we restrict our list of hashtags to category 1 (GBV). Lastly, we combine the hand-coded hashtags on GBV with the machine-coded hashtags on GBV, resulting in a total of 2,009 hashtags complying with our first selection criteria. Due to the character limit of 1,024 characters per request of the Twitter API query, we develop a second selection criteria. We decide to use relevance as our second selection criteria. Therefore, we order our sample of 2,009 hashtags by relevance (number of occurrences) and only include the top 62 most used hashtags in our final query. These hashtags are as follows: #yesallwomen, #metoo, #whyididntreport, #aufschrei, #whyistayed, #notokay, #everydaysexism, #meat14, #rapecultureiswhen, #yesallmen, #rapeculture, #sexismus, #sexualharassment, #sexualassault, #nomoore, #domesticviolence, #rape, #vaw, #bringback-

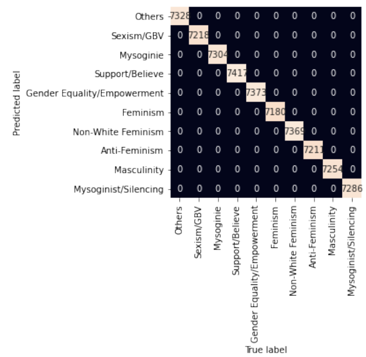
Figure B.9: Stochastic Gradient Classifier



a

[a] Linear SVM with L2-penalty and default settings. Accuracy is 100.

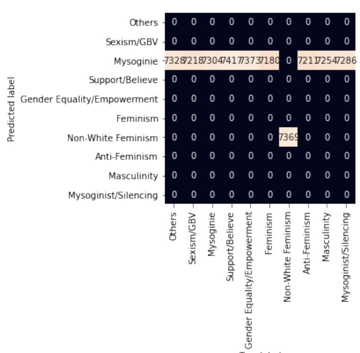
Figure B.10: Decision Tree Classifier



a

[a] Quality of split measured by entropy. Rest is default. Accuracy is 100.

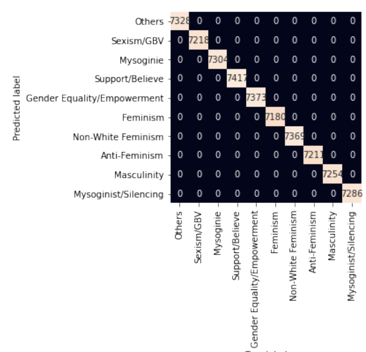
Figure B.11: AdaBoost Classifier



a

[a] Accuracy is 20.11.

Figure B.12: Random Forest Classifier



a

[a] All specifications are default (100 parallel trees, criteria for split is Gini impurity, no pruning and 2 parallel jobs). Accuracy is 100.

ourgirls, #sexualabuse, #harassment, #sexism, #balancetonporc, #abuse, #whywomendontreport, #nomeansno, #metoos, #yotambien, #roymoorechildmolester, #notok, #hertoo, #metoodebatte, #metoomovement, #consent, #violenceagainstwomen, #weinsteinscandal, #domesticabuse, #victimblaming, #moiaussi, #sexualviolence, #enoughisenough, #endrapeculture, #streetharassment, #abusefreeindia, #csa, #endvaw, #youtoo, #ustoo, #sexualharrassment, #wetoo, #metoomarch, #vawg, #sexualabise, #childabuse, #domesticviolenceawareness, #lockerroomtalk, #freethenipple, #stopvictimblaming, #gbv, #outrage, #dvam, #stopabuse', #sexist, #raped, #metoocampaign.

We extract all tweets (including retweets, quotes and replies) for the period 2014-2016, resulting in a total of 11,335,429 tweets measuring the conversation around GBV on Twitter during that year.

B.2 Assigning Location Information to Twitter Data

To take advantage of regional differences in social movements on Twitter as well as crime reports, we make use of the fact that 76.5 percent of tweets have a Twitter user location. Only 1.5 percent of tweets have a tweet location. We, therefore, decide to rely on the user location and not tweet location to extract geographic information. The Twitter user location is not available in a preprocessed format. This means that some users mention their country of residence, while others indicate their federal state, county, city, or even zip code. In order to match the Twitter data to crime data at the federal state level, we need to unify the data. We do so through using information on all census-recognized cities/towns provided by SimpleMaps (2012). This dataset contains information of the federal state, the federal state ID, the county and county ID, the city ID and zip codes. We then proceed as follows.

We first split the Twitter user location into different columns, based on commas. We then merge the city data to the Twitter data based on the first two columns identifying the user location and the city name and state ID from the city dataset. We next merge both datasets on the city name and state name. Next, we use the county name and state name. We then use the county zip code and state name. As a next step, we merge both datasets on the county name and state ID, and later on the state ID only (using first the first column of the user location and then the second column of the user's location). Similarly, we use the state name only (using first the first column of the user location and then the second column of the user's location only), and then the city name only (using first the first column of the user location and then the second column of the user location only). We repeat this for the city name, county name, city zip and county zip respectively.

Through this procedure, we can assign 74.8 percent (8,566,786 out of 11,449,223) tweets a location (4,244,582 out of 6,236,539 tweets in 2014, 2,112,004 out of 2,685,019 tweets in 2015, 2,210,200 out of 2,527,665 tweets in 2016). In a second step, we address duplicated values. Duplicated values occur, as many cities in the US have the same name. There are 2,744,093 duplicated values in 2014, 1,284,143 in 2015, and 1,394,794 in 2016. If a city is duplicated, we keep the value with the largest population. This leaves us with 1,500,489 tweets in 2014, 827,861 in 2015, and 815,406 in 2016. Consequently, we are able to associate 24.1 percent of tweets in 2014 with a federal state in the US, 30.8 percent of tweets in 2015, and 32.3 percent of tweets in 2016. Through this procedure we can assign 27.5 percent of tweets (3,143,756 out of 11,449,223) a federal state in the US. We believe that this captures a large enough share of all Twitter users in 2021, as 37.7 percent were from the United States (77.75 out of 206 million users worldwide) (Statista 2022).

B.3 The VADER Sentiment Analysis - Methodological Background

To analyze the content of what is written on Twitter within the conversation on GBV, we employ the VADER Sentiment Analysis tool (Hutto and Gilbert 2014). The VADER Sentiment Analysis tool is a lexicon and rule-based sentiment analysis tool, which was trained on social media data. The lexicon has been validated by 10 independent human raters. It builds upon pre-existing, well-established sentiment word-banks (LIWC, ANEW, and GI) and adds common lexical features used on social media to these word-banks. Examples are emoticons (such as ":-)"), acronyms (such as "LOL"), and slang (such as "nah" or "giggly").

The VADER Sentiment Analysis tool deduces both the intensity and polarity of sentiments. The polarity refers to a binary classification into positive, neutral, or negative text. The tool reports the fraction of text, which is positive, neutral, and negative. Adding all three columns results in a value of 1. Importantly, the three columns do not account for contextual interplays of words. The contextual interplay is reflected in the compound score. The compound score is a single uni-dimensional measure of a text's sentiment. It accounts for the contextual connection of independent words through a variety of different methodologies, such as taking into account word-order sensitive relationships, or degree modifiers. The score ranges from -1 to 1. -1 is the most negative and 1 the most positive classification possible.[1]

We showcase the resulting compound score by giving some artificial examples. The term "#metoo is great :-)" has a compound score of "0.7506" and is therefore overall positive. The sentence "#metoo is great." has a compound score of "0.6249". It is less positive as the previous example, as it lacks the smiley. In a similar fashion, "Gender-Based Violence is horrible." has a compound score of -0.8225, "Gender-Based Violence is HORRIBLE." a compound score of -0.8531, and "Gender-Based Violence is really horrible." a compound score of -0.8357.

[1]For examples on the Compound score see the VADER Github Repository. Link: https://github.com/cjhutto/vaderSentiment

B.4 Retrieving Socioeconomic Characteristics from Twitter Data

We apply two different tools to retrieve socioeconomic information from Twitter. Firstly, we make use of the *DeepFace* framework developed by Serengil and Ozpinar (2020). This framework is a lightweight face recognition and facial attribute analysis package in Python[2]. It allows to retrieve users' age, gender, emotion, and race from profile pictures. We make use of the default model, which is the VGG-Face model. The VGG-Face model was developed by Parkhi, Vedaldi, and Zisserman (2015) and is a convolutional neural network (CNN) model. This model was trained using photos of two million faces and a "very deep" network.

To shed light on the gender of those tweeting within our dataset, we employ the *GenderGuesser*.[3] This package allows for the detection of authors' gender based on their first names. The resulting sexes are male, female, mostly male, mostly female, andy (having an equal probability to be male and female) as well as unknown.

B.5 Additional Tables

B.5.1 Analysing the Twitter Tweets' Text - Sentiment Scores by Sub-type of Crime

[2]Its accuracy is above 97.53 percent (Serengil and Ozpinar 2020).

[3]For the details and license information on the *GenderGuesser* package see `https://pypi.org/project/gender-guesser/`.

Table B.1: The effect of the polarity of tweets on GBV on crime rates per 100,000 inhabitants (Sexual violence)

	(1) Crime rate	(2) Crime rate	(3) Crime rate
Compound Score	-0.00000246	-0.00841	-0.00916
	(0.000283)	(0.00812)	(0.00887)
L.Compound Score		0.00845	0.0114
		(0.00820)	(0.00811)
L2.Compound Score			-0.00218
			(0.00754)
Constant	0.765***	0.763***	0.764***
	(0.00408)	(0.00396)	(0.00401)
Mean (Dep. Var)	0.765	0.763	0.764
St. Dv. (Dep. Var.)	0.720	0.717	0.717
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: This table shows the results of a linear regression of the compound score on GBV-related crime rates. The unit of analysis is the state by week level. I restrict crimes to sexual violence. For the details behind the compound score see Appendix B.3. Source: Twitter and NIBRS (2014-2016).

Table B.2: The effect of the polarity of tweets on GBV on crime rates per 100,000 inhabitants (Physical violence)

	(1) Crime rate	(2) Crime rate	(3) Crime rate
Compound Score	-0.000167	-0.00578	-0.00634
	(0.000701)	(0.0215)	(0.0243)
L.Compound Score		0.00566	0.0126
		(0.0217)	(0.0193)
L2.Compound Score			-0.00640
			(0.0240)
Constant	4.000***	4.002***	4.005***
	(0.0175)	(0.0175)	(0.0174)
Mean (Dep. Var)	4.000	4.002	4.005
St. Dv. (Dep. Var.)	3.715	3.716	3.718
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: This table shows the results of a linear regression of the compound score on crime rates. I restrict crimes to physical violence. The unit of analysis is the state by week level. For the details behind the compound score see Appendix B.3. Source: Twitter and NIBRS (2014-2016).

Table B.3: The effect of the polarity of tweets on GBV on crime rates per 100,000 inhabitants (Emotional violence)

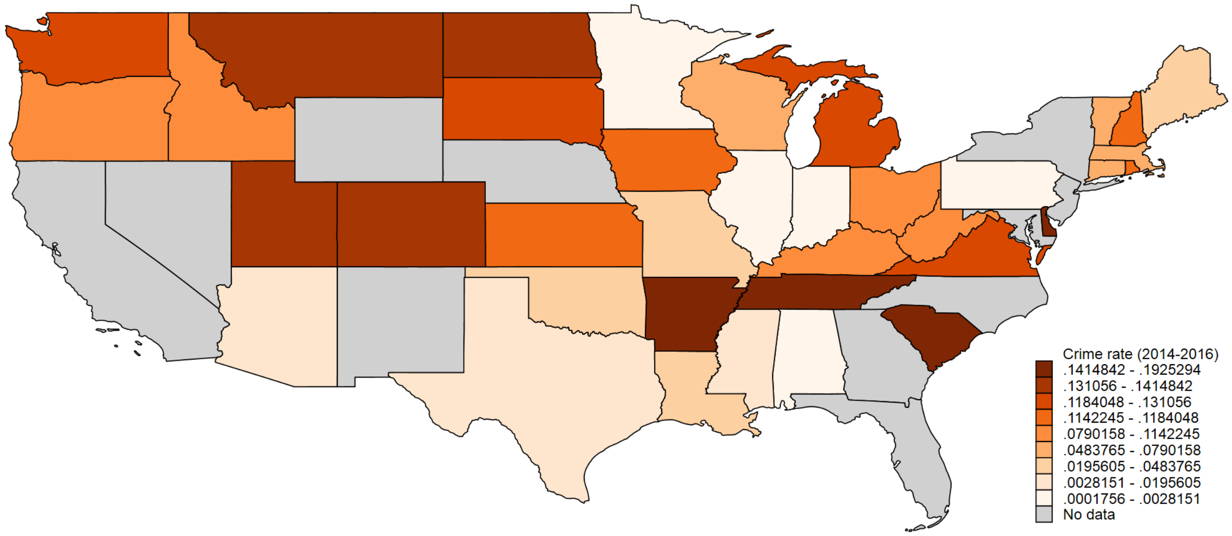
	(1) Crime rate	(2) Crime rate	(3) Crime rate
Compound Score	-0.000479	-0.00722	-0.00554
	(0.000353)	(0.00763)	(0.00874)
L.Compound Score		0.00686	0.00865
		(0.00768)	(0.00886)
L2.Compound Score			-0.00349
			(0.0105)
Constant	1.197***	1.198***	1.199***
	(0.00767)	(0.00762)	(0.00757)
Mean (Dep. Var)	1.197	1.198	1.199
St. Dv. (Dep. Var.)	1.418	1.419	1.420
State-Month fixed-effects	Yes	Yes	Yes
N	5751	5712	5673

Notes: This table shows the results of a linear regression of the compound score on crime rates. I restrict crimes to emotional violence. The unit of analysis is the state by week level. For the details behind the compound score see Appendix B.3. Source: Twitter and NIBRS (2014-2016).

B.6 Additional Figures

B.6.1 Maps

Figure B.13: Aggregated non-GBV-related crime rate (2014-2016) over the population in 2014



Notes: The map depicts the aggregate number of non-GBV-related crimes reported to the police for the years 2014-2016 divided by population estimates from 2014 at the federal state level in the US. The graph excludes Alaska, Hawaii, and Puerto Rico. Darker colors indicate higher aggregated GBV-related crime rates. Source: NIBRS and US Census Bureau.

Appendix C

Appendix to Chapter 3

C.1 Data Appendix on SIAB

The following section outlines the variable construction included in the SIAB data in more detail.

Employment. We define the following people as employed: Employees subject to social security, trainees, marginal part-time workers, employees in partial retirement, interns and student trainees, as well as those who have another employment status in the SIAB population. This information is based on the Employee History (Beschäftigtenhistorik - BeH) of the IAB. This dataset is based on the notification procedure for health, pension and unemployment insurance. Under this procedure, employers have to give notification of all their employees subject to social security at least once a year. This dataset covers all white- and blue-collar workers as well as apprentices subject to social security contributions. This means that the dataset lacks information on civil servants, self-employed persons and regular students without marginal part-time jobs. Since 1999 the dataset also covers marginal part-time employment and unpaid family workers.

Unemployment. We identify the unemployed through information gathered in the Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung - LHG). Information on unemployed people is available since 2005 (although incomplete until 2007). We also rely on information from the Jobseeker Histories (Arbeitsuchendenhistoriken – ASU/XASU). The ASU contains information on jobseekers registered with employment agencies while the XASU provides data on jobseekers who receive Unemployment Benefit II (ALG-II). The latter is available since 2005, while the former is available since 1997.

Migration Status. We identify a person's migration status by information gathered on citizenship. As soon as a person has a foreign citizenship, we define them as foreigners. There are several conditions for German citizenship. First, children born in Germany with at least one German parent are German citizens. Second, children born abroad with at least one German parent can acquire German citizenship within the first year of their lives. Marriage allows for German citizenship after two years of marriage and a legal residence of three years. Children born in Germany with foreign parents can acquire German citizenship if one parent has been in Germany for at least eight years and has a permanent right of residence. This means that the information provided reflects the foreign status of a person in the labor force instead of a migration background.

Wages. Wages are daily wages in Euros and equal to the salary for a given time period divided by the number of days of this time period. Until 1999, this variable only reported wages of employment subject to social security. In 1999, the inclusion of marginal part-time workers led to an amplification of this variable. Still, the variable is subject to the upper earnings limit. For a detailed overview of the data limitations see Bender et al. (1996). We follow Dauth and Eppelsheimer (2020) and prepare the wage variable by taking into account the contribution assessment ceiling, the marginal part-time income threshold, deflating wages and ceilings, and by imputing them. The yearly earnings are daily wages multiplied by the days of employment per year.

C.2 Additional Tables

Table C.1: AI skill demands and cumulative internal migration inflows (German) at the county level

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	2.300*	138.0**	0.737**	43.31*	0.787*	51.52**	0.400*	22.81**
	(1.184)	(65.87)	(0.364)	(23.72)	(0.442)	(22.52)	(0.208)	(10.99)
Constant	2477.5*	-2042.1	1046.1*	-667.9	581.5	-830.8	241.7	-243.2
	(1304.9)	(3395.2)	(561.9)	(1091.8)	(376.0)	(1269.7)	(149.6)	(557.4)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.729	.	0.706	.	0.776	.	0.757	.
N	762	762	762	762	762	762	762	762

Notes: The table reports the results for a stacked long-difference estimation on the internal migration inflow of German citizens. We consider two three year periods, namely 2014-2016 and 2017-2019. The units of analysis are county-period cells. We measure the internal migration inflow by creating a dummy variable which is equal to one if the current county of residence is not equal to the county of residence in the previous year. We then sum all observations at the county level for the periods investigated. The explanatory variable is the 0.01 percentage point change in AI-related skill demands. AI skill demands are measured via the share of AI-related skill demand in all skill demand. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population of each county, and the share of young- and old-aged workers. We also create a shift-share instrument for trade exposure, using data from COMTRADE. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective SIAB population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.2: AI skill demands and cumulative internal migration outflow (German) at the county level

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	-0.0699	18.33	0.0887	5.363	-0.109	11.26	-0.0498	1.701
	(0.790)	(27.92)	(0.252)	(8.784)	(0.384)	(15.49)	(0.167)	(4.064)
Constant	1888.8**	824.3	888.1**	287.3	755.2**	297.1	245.5**	239.9*
	(906.9)	(862.7)	(414.7)	(277.8)	(381.9)	(482.0)	(124.1)	(124.8)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.838	0.381	0.819	0.503	0.854	0.264	0.786	0.403
N	762	762	762	762	762	762	762	762

Notes: The table reports the results for a stacked long-difference estimation on the internal migration outflow of German citizens. We consider two three year periods, namely 2014-2016 and 2017-2019. The units of analysis are county-period cells. We measure the internal migration outflow by creating a dummy variable which is equal to one if the current county of residence is not equal to the county of residence in the previous year. We then sum all observations at the county level for the periods investigated. The explanatory variable is the 0.01 percentage point change in AI-related skill demands. AI skill demands are measured via the share of AI-related skill demand in all skill demand. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population of each county, and the share of young- and old-aged workers. We also create a shift-share instrument for trade exposure, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective SIAB population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.3: AI skill demands and the probability to migrate internally by skill groups and citizenship at the county-year-level

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Migrant	-0.010**	0.091***	-0.012***	0.069	-0.001	0.064***	-0.041***	0.079*
	(0.00238)	(0.0183)	(0.00209)	(0.0426)	(0.00212)	(0.00996)	(0.00324)	(0.0211)
AI	0.003	0.447***	-0.004	0.172**	-0.002	0.324***	0.013***	0.579***
	(0.0000146)	(0.000414)	(0.0000316)	(0.000513)	(0.00000769)	(0.000669)	(0.0000355)	(0.000618)
Migrant*AI	0.002**	-0.310**	0.003	-0.223	-0.000	-0.201	0.000	-0.482*
	(0.0000240)	(0.00302)	(0.0000637)	(0.00287)	(0.0000312)	(0.00341)	(0.0000161)	(0.00602)
Constant	***	***	**	***	**	**	***	***
	(0.0185)	(0.00280)	(0.0237)	(0.00838)	(0.0136)	(0.00491)	(0.0278)	(0.00380)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.002	.	0.002	.	0.002	.	0.006	.
N	9865642.000	4177551.000	1557376.000	757587.000	6958648.000	2885919.000	1349617.000	534044.000

Notes: The table reports the results for a linear probability model on the probability to migrate inside of Germany. The regression is at the individual and yearly level. We measure the probability to migrate via a dummy variable which is equal to one for all individuals which report a change in their county of residence from one year to the other. We choose the county of residence and not the county of individuals' workplace to also capture a movement between counties for the unemployed. The explanatory variable is the 0.01 percentage point change in AI-related skill demands. AI skill demands are measured as a share in all skill demand in local labor markets. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population of each county, and the share of young- and old-aged workers. We also create a shift-share instrument for trade exposure, using data from COMTRADE. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective SIAB population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.4: AI skill demands and cumulative immigrant inflow (employed) by skill groups at the county level

	All (OLS)	All (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	0.478	48.88	0.478	48.88	0.139	10.19
	(0.463)	(45.00)	(0.463)	(45.00)	(0.115)	(8.887)
Constant	2147.6*	-565.7	2147.6*	-565.7	538.0*	-96.30
	(1208.1)	(1347.9)	(1208.1)	(1347.9)	(290.4)	(288.4)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.864	.	0.864	.	0.844	.
N	762	762	762	762	762	762

Notes: The table reports the results for a stacked long-difference estimation on the cumulative immigrant inflow of employed people. The units of analysis are county-period cells. We measure the immigrant inflow via the first time a person with a foreign citizenship is observed in the SIAB in a given year and county. In this case, I only consider immigrants who are employed immediately. The explanatory variable is the 0.01 percentage point change in AI-related skill demands. AI skill demands are measured via the share of AI-related skill demand in all skill demand. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population of each county, and the share of young- and old-aged workers. We also create a shift-share instrument for trade exposure, using data from COMTRADE. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective SIAB population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.5: AI skill demands and percentage change in daily wages by skill groups and citizenship at the county level

	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	5.564***	9.188**	0.825	1.315	2.338**	4.760**
	(1.896)	(4.181)	(0.536)	(0.903)	(1.001)	(2.174)
AI	-0.00607	3.733	0.00203	-0.184	0.0280	1.258
	(0.0518)	(2.658)	(0.00778)	(0.588)	(0.0257)	(1.601)
Foreign*AI	0.0195	-7.903	0.0200	-1.081	-0.152	-5.554*
	(0.167)	(6.019)	(0.0379)	(1.101)	(0.116)	(3.247)
Constant	136.9	10.31	-30.56**	4.563***	11.64	5.295*
	(132.3)	(7.788)	(15.04)	(1.338)	(57.08)	(2.746)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0273	.	0.0774	.	0.0416	.
N	1513	1513	1524	1524	1513	1513

Notes: The table reports the results for a stacked long-difference estimation on the percentage change in daily wages. The units of analysis are county-period-citizenship cells. The explanatory variable is the 0.01 percentage point change in AI-related skill demands. AI skill demands are measured via the share of AI-related skills in the total number of skills. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The outcome variable is the percentage change in daily wages. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to information and communication technology as well as trade, using data from EUROSTAT as well as COMTRADE. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.3 Spillover Effects

Table C.6: AI skill demands and cumulative net internal migration inflows by skill groups and citizenship at the county level (most exposed sectors)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	-2.887*	8.949	-3.023*	2.176	0.626	5.583*	-0.491	1.190
	(1.677)	(6.662)	(1.717)	(2.771)	(0.521)	(3.270)	(0.387)	(0.886)
AI	0.261	26.98*	0.114	8.967**	0.0976*	7.064**	0.0499	2.700**
	(0.186)	(13.93)	(0.0985)	(4.245)	(0.0565)	(3.026)	(0.0357)	(1.188)
Foreign*AI	-0.234	-23.41*	-0.0994	-10.28*	-0.0823*	-9.789*	-0.0518	-3.343*
	(0.178)	(12.61)	(0.0982)	(5.875)	(0.0472)	(5.005)	(0.0381)	(1.814)
Constant	2.710	-129.0	4.814	-0.621	-2.638	-2.187	0.533	-1.345**
	(5.452)	(144.7)	(5.531)	(2.022)	(1.695)	(1.817)	(0.950)	(0.610)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0600	.	0.112	.	0.0172	.	0.102	.
N	1524	1524	1524	1524	1524	1524	1524	1524

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. We restrict the sample to those economic sectors with the highest AI adoption. The units of analysis are county-period-citizenship cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. We measure AI skill demands via the share of AI-related skill demand in all skill demand in a local labor market. The net internal migration inflow is the internal migration inflow into a county minus the internal migration outflow during the period under consideration. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.7: AI skill demands and cumulative immigrant inflows from abroad by skill groups and at the county level (most exposed)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	0.0641	1.812	0.0383	0.993	0.0173	0.608	0.00976	0.185
	(0.0582)	(4.688)	(0.0323)	(2.675)	(0.0183)	(1.468)	(0.00989)	(0.559)
Constant	78.27*	72.18	48.79**	43.15	20.48*	21.49	9.068*	7.446
	(39.95)	(49.04)	(24.44)	(28.63)	(10.59)	(14.23)	(5.272)	(6.426)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.900	0.199	0.889	0.198	0.899	0.177	0.867	0.251
N	762	762	762	762	762	762	762	762

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period cells. We restrict the sample to those economic sectors with the highest AI adoption. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of job skills mentioning AI-related expressions in all skill demand. The outcome variable is the cumulative immigrant inflow into a specific county. We identify the immigrant inflow by aggregating the number of times a migrant appears for the first time in the SIAB per stacked-long difference. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.8: AI skill demands and percentage change in daily wages by skill groups and citizenship at the county level (most exposed)

	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	44.50**	42.29	21.66***	17.90**	7.743	13.12
	(17.84)	(35.57)	(4.550)	(8.568)	(10.94)	(17.65)
AI	-0.0162	-9.770*	0.0518	-2.481	0.226	7.218
	(0.118)	(5.437)	(0.0540)	(3.235)	(0.219)	(7.046)
Foreign*AI	-0.813	-5.252	-0.120	5.437	-4.391	-15.85
	(1.176)	(35.87)	(0.265)	(10.95)	(4.865)	(15.09)
Constant	-37.19	3.943	-68.28*	-7.881*	51.37	29.90*
	(48.42)	(8.754)	(36.08)	(4.688)	(66.11)	(15.31)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0249	.	0.0631	.	0.0453	.
N	1017	1017	1110	1110	898	898

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period-citizenship cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. We restrict the sample to those economic sectors with the highest AI adoption. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of AI-related skills in the total number of skills. The outcome variable is the percentage change in daily wages. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to information and communication technology as well as trade, using data from Eurostat as well as Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.9: AI skill demands and percentage change unemployment rates by skill groups and citizenship at the county level (most exposed)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	-16.49	63.73	22.82	213.9	-7.606	-194.5	36.63	46.50***
	(22.55)	(135.3)	(22.80)	(355.4)	(56.53)	(342.8)	(27.11)	(12.85)
AI	-0.139	11.05	0.625	-6.967	0.189	9.892	-0.539	14.08
	(0.496)	(9.104)	(0.938)	(24.86)	(0.440)	(13.55)	(1.828)	(16.21)
Foreign*AI	7.581	-75.52	-5.290	-204.5	3.776	169.3	15.36	3.051
	(4.859)	(120.4)	(5.520)	(304.5)	(9.534)	(325.1)	(15.08)	(46.72)
Constant	85.40	23.08	2.147	290.3	132.6	-95.88	-36.96	-170.6
	(94.23)	(235.4)	(152.5)	(537.7)	(141.9)	(357.4)	(204.0)	(299.7)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.128	.	0.251	.	0.138	.	0.428	0.210
N	586	586	289	289	450	450	118	118

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period-citizenship cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. We restrict the sample to those economic sectors with the highest AI adoption. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured as a share in all skill demand in a local labor market. The outcome variable is the percentage change in unemployment rates. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.10: AI skill demands and cumulative net internal migration inflows by skill groups and citizenship at the county level (least exposed sectors)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	-2.887*	8.949	-3.023*	2.176	0.626	5.583*	-0.491	1.190
	(1.677)	(6.662)	(1.717)	(2.771)	(0.521)	(3.270)	(0.387)	(0.886)
AI	0.261	26.98*	0.114	8.967**	0.0976*	7.064**	0.0499	2.700**
	(0.186)	(13.93)	(0.0985)	(4.245)	(0.0565)	(3.026)	(0.0357)	(1.188)
Foreign*AI	-0.234	-23.41*	-0.0994	-10.28*	-0.0823*	-9.789*	-0.0518	-3.343*
	(0.178)	(12.61)	(0.0982)	(5.875)	(0.0472)	(5.005)	(0.0381)	(1.814)
Constant	2.710	-129.0	4.814	-0.621	-2.638	-2.187	0.533	-1.345**
	(5.452)	(144.7)	(5.531)	(2.022)	(1.695)	(1.817)	(0.950)	(0.610)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0600	.	0.112	.	0.0172	.	0.102	.
N	1524	1524	1524	1524	1524	1524	1524	1524

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period-citizenship cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. We restrict the sample to those economic sectors with the lowest AI adoption. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. We measure AI skill demands via the share of AI-related skill demand in all skill demand in a local labor market. The net internal migration inflow is the internal migration inflow into a county minus the internal migration outflow during the period under consideration. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.11: AI skill demands and cumulative immigrant inflows from abroad by skill groups and at the county level (least exposed)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
AI	0.0641	1.812	0.0383	0.993	0.0173	0.608	0.00976	0.185
	(0.0582)	(4.688)	(0.0323)	(2.675)	(0.0183)	(1.468)	(0.00989)	(0.559)
Constant	78.27*	72.18	48.79**	43.15	20.48*	21.49	9.068*	7.446
	(39.95)	(49.04)	(24.44)	(28.63)	(10.59)	(14.23)	(5.272)	(6.426)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.900	0.199	0.889	0.198	0.899	0.177	0.867	0.251
N	762	762	762	762	762	762	762	762

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period-citizenship cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. We restrict the sample to those economic sectors with the lowest AI adoption. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of job skills mentioning AI-related expressions in all skill demand. The outcome variable is the cumulative immigrant inflow into a specific county. We identify the immigrant inflow by aggregating the number of times a migrant appears for the first time in the SIAB per stacked-long difference. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, and the share of workers in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.12: AI skill demands and percentage change in daily wages by skill groups and citizenship at the county level (least exposed)

	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	44.50**	42.29	21.66***	17.90**	7.743	13.12
	(17.84)	(35.57)	(4.550)	(8.568)	(10.94)	(17.65)
AI	-0.0162	-9.770*	0.0518	-2.481	0.226	7.218
	(0.118)	(5.437)	(0.0540)	(3.235)	(0.219)	(7.046)
Foreign*AI	-0.813	-5.252	-0.120	5.437	-4.391	-15.85
	(1.176)	(35.87)	(0.265)	(10.95)	(4.865)	(15.09)
Constant	-37.19	3.943	-68.28*	-7.881*	51.37	29.90*
	(48.42)	(8.754)	(36.08)	(4.688)	(66.11)	(15.31)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.0249	.	0.0631	.	0.0453	.
N	1017	1017	1110	1110	898	898

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period-citizenship cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. We restrict the sample to those economic sectors with the lowest AI adoption. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured via the share of AI-related skills in the total number of skills. The outcome variable is the percentage change in daily wages. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the nuts-3 level. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to information and communication technology as well as trade, using data from Eurostat as well as Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.13: AI skill demands and percentage change unemployment rates by skill groups and citizenship at the county level (least exposed)

	All (OLS)	All (IV)	High-skilled(OLS)	High-skilled (IV)	Medium-skilled (OLS)	Medium-skilled (IV)	Low-skilled (OLS)	Low-skilled (IV)
Foreign	-16.49	63.73	22.82	213.9	-7.606	-194.5	36.63	46.50***
	(22.55)	(135.3)	(22.80)	(355.4)	(56.53)	(342.8)	(27.11)	(12.85)
AI	-0.139	11.05	0.625	-6.967	0.189	9.892	-0.539	14.08
	(0.496)	(9.104)	(0.938)	(24.86)	(0.440)	(13.55)	(1.828)	(16.21)
Foreign*AI	7.581	-75.52	-5.290	-204.5	3.776	169.3	15.36	3.051
	(4.859)	(120.4)	(5.520)	(304.5)	(9.534)	(325.1)	(15.08)	(46.72)
Constant	85.40	23.08	2.147	290.3	132.6	-95.88	-36.96	-170.6
	(94.23)	(235.4)	(152.5)	(537.7)	(141.9)	(357.4)	(204.0)	(299.7)
Federal States fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.128	.	0.251	.	0.138	.	0.428	0.210
N	586	586	289	289	450	450	118	118

Notes: The table presents stacked-long difference regressions for two periods of three years, 2014-2016 and 2017-2019. The units of analysis are county-period-citizenship cells. *Foreign* is a dummy variable equal to one for foreign citizens and zero otherwise. We restrict the sample to those economic sectors with the lowest AI adoption. The explanatory variable is the 0.01 percentage point change in AI skill demands over time. AI skill demands are measured as a share in all skill demand in a local labor market. The outcome variable is the percentage change in unemployment rates. The skill level is based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. Standard errors are in parentheses and clustered at the NUTS-3 level. We control for the following variables: The share of women in the SIAB population of each county, the share of young- and middle-aged workers, the share of part-time workers, the share of middle- and high-skilled workers, the share of workers in the manufacturing sector, in the information and communication sector. We also create a shift-share instrument for a county's exposure to trade, using data from Comtrade. We control for federal states (NUTS-1) fixed effects. We weight each cell by the respective population in a county. The period under consideration is 2014 to 2019, as the Burning Glass data is only available from 2014 onwards. Source: Burning Glass data and SIAB data. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.4 Additional Graphs

Figure C.1: Immigrant inflow to Germany over time

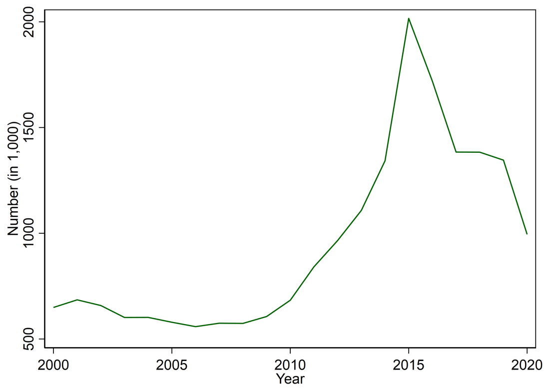
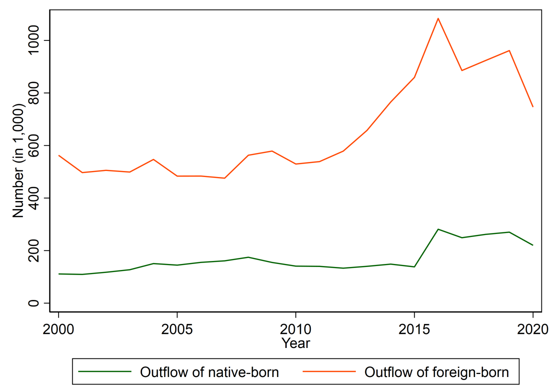


Figure C.2: Outflow of native- and foreign-born residents over time



Notes: The left figure shows the yearly immigrant inflow to Germany over time for the period 2000 to 2020 (per 1,000 people). Immigrants are foreign-born citizens. The right figure shows the yearly outflow of native-born and foreign-born citizens over time for the period 2000 to 2020 (per 1,000 people). Source: GENESIS (Federal Statistical Office).

Figure C.3: Immigrant inflow to main OECD countries

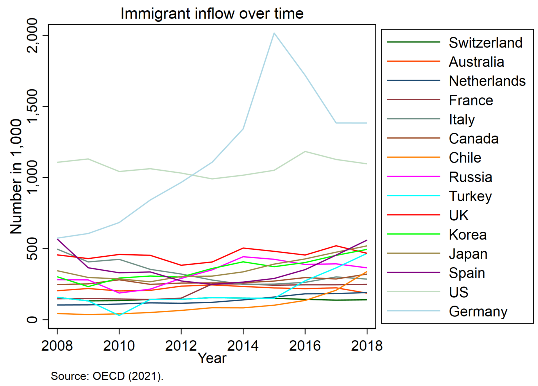
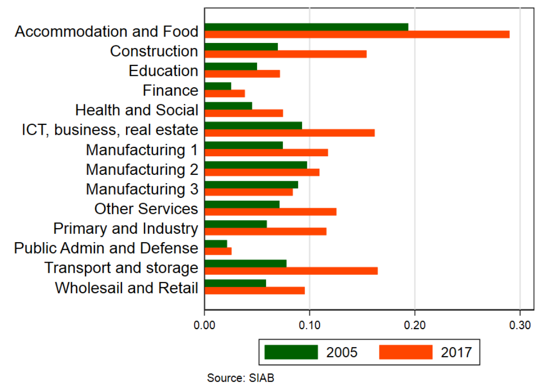
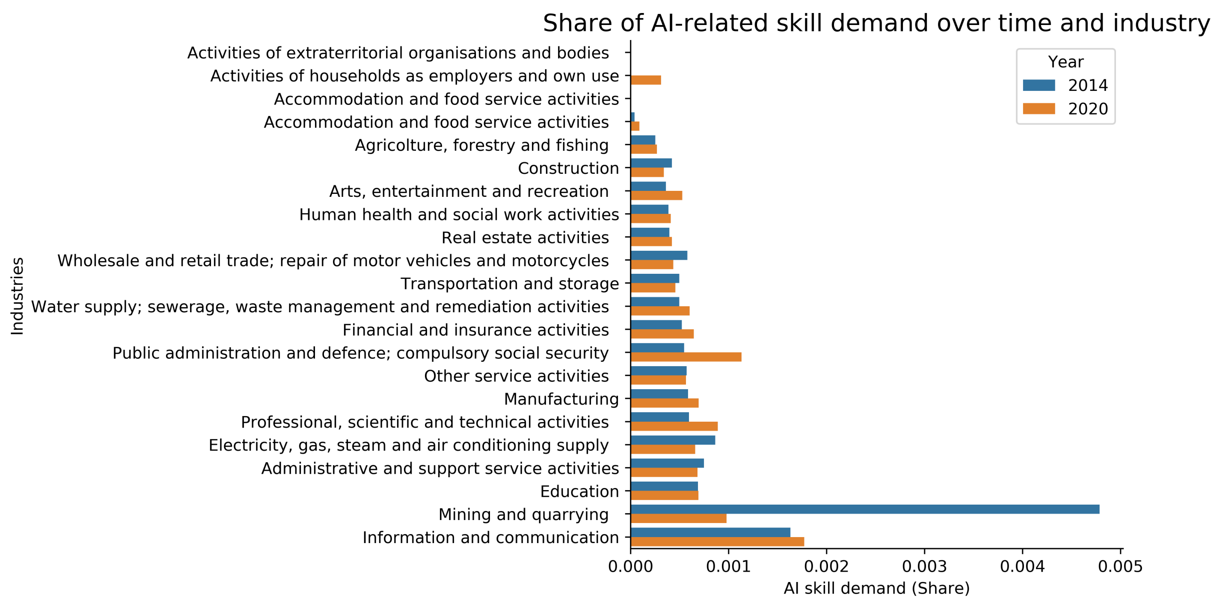


Figure C.4: Migrant Share by economic sector in 2005 and 2017



Notes: The left figure shows the yearly immigrant inflow to Germany to the main-receiving OECD countries for the period 2008 to 2018 (per 1,000 people). Immigrants are foreign-born citizens. The right figure shows the migrant share by economic sectors in 2005 and 2017. Source: OECD (2021) and SIAB (2021).

Figure C.5: Share of AI skill demand in overall skill demand by sector



Notes: The figure plots the share of AI-related skill demand by economic sector in 2014 (in blue) and 2020 (in orange). Source: BGD (2021).

C.5 Comparison to Low-Skilled Task Automation - Methodological Details

To measure low-skilled automation we make use of data provided by the Industrial Federation of Robotics on the installation and operational stock of industrial robots.[1] The data is available at the country-industry level and for the period 1994 to 2020. It reveals the number of newly installed industrial robots as well as the operational stock of already installed robots per year, country and industry. The data is available at one-digit industry codes for the non-manufacturing sectors and at the two or partly three-digit level for the manufacturing sector.

We analyze the effect of robot exposure at the level of local labor markets. We therefore aggregate our SIAB data at the county level. We consider the period of 2005 to 2018 due to data availability. To estimate the effect of robot adoption on immigration demand as well as labor market outcomes of migrants versus natives, we follow the approach by Acemoglu and Restrepo (2018b) and apply a shift-share instrument. Simply analyzing the impact of an increased exposure to robots through linear regressions could bias our estimates, as locations with more exposure to robots might systematically differ on unobservable characteristics from those with less exposure to robots.

An additional challenge is that the data provided by the IFR on robot adoption is only reported at the national level. As a result, we apply a shift-share instrument to proxy robot exposure at the local level r through industry shares in each county, similar to Dauth et al. (2021). This means that we construct our main explanatory variable as follows:

$$\Delta \mathrm{rob\hat{o}ts}_r = \sum_{i \in I} \frac{\mathrm{emp}_{ir}}{\mathrm{emp}_r} \times \frac{\Delta \mathrm{robots}_i}{\mathrm{emp}_i} \text{ , with I=28} \tag{C.1}$$

The term $\frac{\Delta \mathrm{robots}_i}{\mathrm{emp}_i}$ is the difference in robot counts in a respective industry between 2018 and 2005 over employment in the respective industry in 2004. This means that we first calculate the difference in the robotic operational stock between 2018 and 2005 for each industry. We then divide this number by the number of employed people in each industry in 2004. As the robotics data is only available at the national level, we proxy the county level exposure to robotics via the employment share of each respective industry in each region in 2004 ($\frac{\mathrm{emp}_{ir}}{\mathrm{emp}_r}$). emp_{ir} is the number of employed people in region r in industry i in our base year in 2004. emp_r is the number of employed people in region r in our base year in 2004. We then multiply the resulting scaled difference in robot counts by the share of people

[1]The IFR collects this data for a large number of countries using a survey of robot suppliers, covering more than 90 per cent of the world's robot market. The definition of a robot in this dataset is "automatically controlled, reprogrammable, and multipurpose machines" (IFR, 2016). This means that robots are machines that do not require a human operator and can be programmed to perform a variety of manual tasks.

employed in a particular industry in a certain county in the base year 2004. We fix industry shares to the pre-treatment period by convention (Goldsmith-Pinkham, Sorkin, and Swift 2020).

We follow Dauth et al. (2021) and conduct the following regression:

$$\Delta Y_r = \alpha X_r' + \beta_1 \times \Delta \hat{\text{robots}}_r + \beta_2 \times \Delta \hat{\text{trade}}_r + \beta_3 \times \Delta \hat{\text{ICT}}_r + \varphi_{REGr} + \epsilon_r \tag{C.2}$$

We regress our outcome variable of interest on the change of robot exposure. This means that we follow a differential exposure design. We control for demographic characteristics at the county-level in 2004 (the female share, the overall share of different skill-groups and the share of workers belonging to different age groups). We also control for regional dummies at the Federal State (NUTS-1) level and cluster our standard errors at geographic level of our analysis (the NUTS-3 level). We additionally control for the difference in ICT equipment as well as trade exposure at the local labor market. We weight our regression by the number of people observed in each local labor market.

Our identification strategy relies on the assumption that robot exposure at the industry level is exogenous and not correlated with labor demand. However, the adoption of robotics could be subject to domestic industry-specific demand shocks. To address this endogeneity concern we conduct an instrumental variable strategy; this closely follows the methodology proposed by Acemoglu and Restrepo (2018b). We use robot installations from Japan, South Korea and Taiwan as our instrumental variables. We select these countries as they are non-European and therefore not subject to the same unobservable shocks to migration as European counterparts would be. Additionally, they are major players in robotics worldwide. All three countries were among the ten countries with the largest number of robot installations in 2018. Figure C.7 demonstrates the robot exposure per 1,000 employees over time in all three countries compared to Germany. South Korea has been outperforming Germany in its robot adoption since 2009, while Taiwan outperformed it in 2013 and Japan in 2015. All countries are therefore a good option as they lead in robot adoption. Additionally, by combining three different countries, the empirical strategy becomes more robust to individual country-level shocks.

Table C.15 shows the first-stage results at the industry level. For the first stage, we simply regress robot adoption, meaning the difference in the operational stock of robots during the period under consideration, at the industry-level in Germany on robot adoption at the industry-level in our three instrumental countries. The coefficient is positive and significant and the F-statistic is well above ten.

Table C.14: Descriptive table of main variables of interest (2005-2018)

	N	Mean	Standard Dev.	Min	Max
Cum. immigrant inflow	402	290.709	550.7139	22	7968
Cum. immigrant inflow (High-skilled)	402	51.24129	129.806	3	1972
Cum. immigrant inflow (Medium-skilled)	402	151.0224	272.7884	9	4079
Cum. immigrant inflow (Low-skilled)	402	84.68408	142.9582	4	1691
Cum. internal migration inflow (Net)	402	36.89055	232.2065	-351	2211
Cum. internal migration inflow (Net, high-skilled German)	402	34.199	107.5841	-61	1130
Cum. internal migration inflow (Net, middle-skilled German)	402	32.41045	79.0961	-185	688
Cum. internal migration inflow (Net, low-skilled German)	402	-29.71891	78.35279	-190	750
Cum. internal migration inflow (Net, Foreign)	402	-6.052239	19.92992	-63	194
Cum. internal migration inflow (Net, high-skilled, Foreign)	402	-.0970149	9.65505	-30	100
Cum. internal migration inflow (Net, middle-skilled, Foreign)	402	-2.910448	11.8676	-57	78
Cum. internal migration inflow (Net, low-skilled, Foreign)	402	-3.044776	7.352065	-32	31
Pct. change in migrant share (all)	403	.0488026	.0355556	-.5	.1508509
Pct. change in migrant share (low-skilled)	402	.0874355	.0583018	-.0531018	.3263894
Pct. change in migrant share (medium-skilled)	403	.0444965	.035968	-.5	.161651
Pct. change in migrant share (high-skilled)	402	.067517	.0409622	-.069488	.2182018
Pct. change in unemployment rate (natives)	402	-55.92103	14.41693	-100	4.782657
Pct. change in unemployment rate (low-skilled natives)	397	-45.38345	58.07743	-100	275
Pct. change in unemployment rate (medium-skilled natives)	402	-55.5634	16.94183	-100	28.78048
Pct. change in unemployment rate (high-skilled natives)	351	-39.68069	57.86393	-100	277.027
Pct. change in unemployment rate (foreign)	326	-56.37943	37.7327	-100	186.9565
Pct. change in unemployment rate (low-skilled foreign)	245	-40.2825	75.87135	-100	335.7724
Pct. change in unemployment rate (medium-skilled foreign)	287	-9.819822	99.29795	-100	664.7152
Pct. change in unemployment rate (high-skilled foreign)	80	-38.70815	85.55682	-100	404.0932
Pct. change in daily wage (Foreign)	402	8.047142	42.20394	-81.21874	355.6371
Pct. change in daily wage (German)	403	11.94946	6.314409	-9.346504	37.33421
Pct. change in yearly labor earnings (German)	403	14.33619	10.61179	-23.03249	182.5034
Pct. change in yearly labor earnings (Low-skilled foreign)	382	25.71007	100.3223	-84.53728	1167.658
Pct. change in yearly labor earnings (Medium-skilled foreign)	396	24.2519	312.5652	-66.54087	6088.281
Pct. change in yearly labor earnings (High-skilled foreign)	341	26.298	163.2595	-87.15948	2003.878
Pct. change in yearly labor earnings (Low-skilled natives)	402	19.43951	26.63032	-32.07589	265.8091
Pct. change in yearly labor earnings (Medium-skilled natives)	403	9.218668	9.774381	-63.04432	154.1482
Pct. change in yearly labor earnings (High-skilled natives)	402	2.746858	16.93573	-31.21652	197.6927
Share of women 2004	402	.489217	.0394597	.3231241	.5844898
Share of medium-skilled 2004	402	.7496725	.0464956	.5925203	.8546042
Share of high-skilled 2004	402	.0953953	.0421964	.0263158	.2666236
Share of <35 in 2004	402	.3203929	.0307194	.21625	.4124424
Share of 35-54 in 2004	402	.5377275	.0303162	.4237918	.65875
Share of part-time 2004	402	.3055813	.0438054	.1544594	.479564
Share in manufacturing 2004	402	.2445977	.1030258	.0246305	.6248705
Share in ICT 2004	402	.0200683	.0184462	0	.1317073
ICT exposure	403	.0190358	.0025622	0	.0351929
Trade exposure	403	1862496	664482.9	0	5193728
No. of people	403	66891.47	154184.4	333	2688145
Employed (weighted)	402	1294.26	1640.083	226.5	19628.5
Robot exposure (Op. Stock)	403	.320462	2.849223	.0025944	57.1441
Robot exposure IV (Op. Stock)	403	.4650611	1.444664	-25.47161	8.558339
Observations	403				

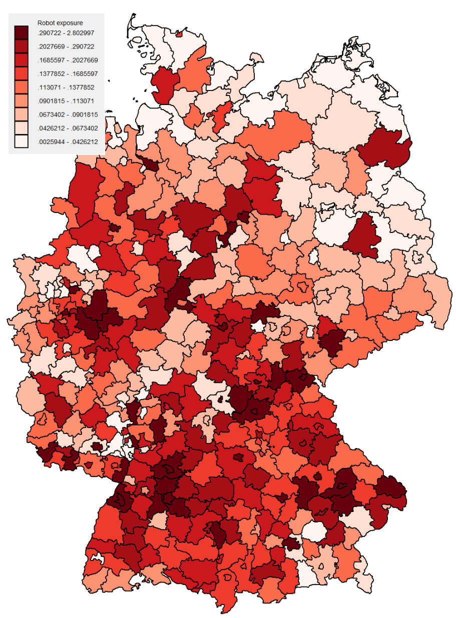
Notes: The table shows summary statistics for one long difference (2005-2018) at the county (NUTS-3) level. We follow the IAB's definition of immigrants, which is based on citizenship: a person is a foreigner as soon as they do not hold German citizenship. Immigrant inflows are identified via the sum of foreigners' first observations in the SIAB data for a given year by county cell. Skill levels are based on the imputed educational variable included in the SIAB. The low-skilled have neither vocational training nor a university degree. The middle-skilled have vocational training and the high-skilled hold a degree from a university or university of applied science. The sector variable captures the economic activity in accordance with the WZ08 classification, grouped into 14 broad categories. Source: SIAB, IFR, and Comtrade.

Table C.15: First-stage: Difference in robot counts by industry

	Robot exposure (DE)
Difference in robot count (KR, JP, TW)	0.223***
	(0.0335)
Constant	1303.9
	(1725.4)
Adj. R-squared	0.548
F-statistic	44.43
N	34

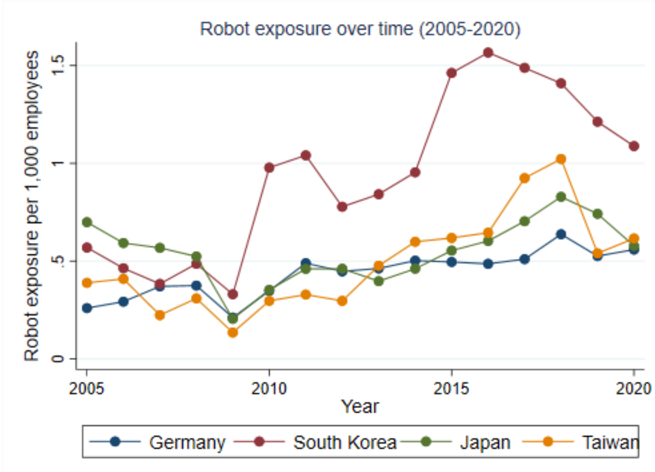
Notes: The table shows the first stage results using robot adoption in Korea (KR), Japan (JP), and Taiwan (TW) as an instrument for robot adoption in Germany (DE). We define robot adoption as the difference in the number of installed robots between 2005 and 2018. The unit of observation is the sector level. We use the IFR classification of sectors, which is close to the ISIC Rev4 sector classification. These classifications can be matched to the SIAB sector classification via crosswalks. This match results in a number of observations of 34. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: BGD (2014-2020).

Figure C.6: Robot exposure by county (2005 to 2018)



Notes: The map shows the counties' exposure to industrial robots. We proxy robot exposure by employing the Bartik instrument. This means that we first take the difference of the operational stock of industrial robots in a respective industry between 2005 and 2018. We then multiply this difference by the sectoral employment shares in each county in 2004. Darker shaded areas indicate a greater exposure to robot adoption, while lighter areas indicate a lower exposure. Source: SIAB data.

Figure C.7: Robot exposure in Germany and instrumental countries



Notes: The graph plots the yearly robot exposure in Germany (in blue), South Korea (in red), Japan (in green), and Taiwan (in yellow) for the period 2005-2018. Robot exposure is the number of installed industrial robots per 1,000 employees.

We consider several outcome variables of interest: the cumulative immigrant inflow, the percentage change in migrant share for this same period, the percentage change in unemployment rates as well as the percentage change in migrants' and natives' daily wages. In the case of daily wages and unemployment, we restrict our data to the migrant and native level and include an interaction term in order to analyze if the effect of robotics differs by nationality. Figure C.6 maps the robot exposure for the period 2005 to 2018 at county level. While certain counties report high exposure to robots, others have implemented very little robots over time in relation to their employed population.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Sofern ein Teil der Arbeit aus bereits veröffentlichten Papers besteht, habe ich dies ausdrücklich angegeben.

München, 14.04.2023, Britta Rude