From Causal Inference to Machine Learning

Four Essays in Empirical Economics

Michael Rainer Johann Kaiser

Dissertation

Ludwig Maximilian University Munich
Department of Economics
September 2020

From Causal Inference to Machine Learning

Four Essays in Empirical Economics

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

2020

vorgelegt von:
Michael Rainer Johann Kaiser

Referent: Prof. Dr. Florian Englmaier
Korreferent: Prof. Dr. Joachim Winter
Promotionsschlussberatung: 3. Februar 2021

Datum der mündlichen Prüfer: 21. Januar 2021
Berichterstatter: Prof. Dr. Florian Englmaier, Prof. Dr. Joachim Winter,
Prof. Rajshri Jayaraman, Ph.D.

Acknowledgments

> As we express our gratitude, we must never forget that the highest appreciation is not to utter words but to live by them.
>
> *John F. Kennedy*

First and foremost, I would like to thank my doctoral adviser, Florian Englmaier, who offered me the opportunity to be a doctoral student and complete this dissertation. His guidance throughout the last four years and the plentiful opportunities to present and share my work with others he provided helped me grow as a researcher and person. I would also like to thank my second adviser, Joachim Winter, whose feedback, especially on matters of empirical economics, helped me write better research papers. Finally, I am indebted to Rajshri Jayaraman, who has kindly accepted to serve on my doctoral committee, and from whom I have learned a lot about conducting research and writing papers.

I am indebted to the Deutsche Forschungsgemeinschaft (DFG)—German research foundation—which provided generous funding for my doctoral position under the GRK 1928 *Microeconomic Determinants of Individual Productivity* program.

This dissertation consists of four chapters, three of which are the result of productive collaboration. I have benefited incredibly from the expertise, support and guidance of my co-authors. I am proud of the work we have done as teams, and there can be no doubt that these have become better research papers because a group of people joined forces. I enjoyed working with Rajshri Jayaraman of University of Toronto (Canada) and Marrit Teirlinck of the European School of Management and Technology in Berlin on the project on which Chapter 2 is based. The experiment in Uganda, presented in Chapter 3, has been a massive but exciting undertaking, one which would have never been achievable alone. I am lucky to have been working with Kristina Czura of University of Groningen (The Netherlands), Vojtěch Bartoš of LMU, Timm Opitz of the Max Planck Institut in Munich, and Brendan Shanks of LMU on that project. Finally, the research presented in Chapter 4 is the result of productive collaboration with Florian Englmaier, Jose E. Galdon-Sanchez of Universidad Pública de Navarra (Spain), and Ricard Gil of Queen's University (Canada).

Throughout the last four years, I have been given the opportunity to present at various conferences and workshops. My research has become more insightful through the many discussions I had with and comments I received from participants. Kilian Heilmann, Andreas Steinmayr and Steve Stillman deserve a special mention. I also profited a great deal from frequent exchanges with my friends and colleagues.

Of course, the research projects that form this dissertation owe a lot to the support of research assistants. In particular, Martin Fankhauser, Madeleine Trattler, and Lisa Eitinger supported the work in Chapter 1. Albert Thieme was indispensable to the analysis of Chapter 2. Nathalie Römer, Alea Lasu and Valentina Stöhr have been offering invaluable support in the field and at home to the experiment in Uganda in Chapter 3. Nathalie Römer and Edgar Akopyan also worked on the research project in Chapter 4. I wish all of them success in their future careers.

I have greatly enjoyed my time as a doctoral student, and this would not have been possible without a tremendous group of friends and colleagues. I would like to extend my thanks to Franziska Hünekes, Corinna Hartung, Anne Niedermaier, Nathalie Römer, Marvin Deversi, Benjamin Häusinger, Brendan Shanks, Felix Hagemeister, Lion Henrich, Fabian Stürmer-Heiber, and Maximilian Breu. I wish them all the best, and hope to stay in touch for times to come. A special thank you goes to my great friend Oscar Rivera Arenas who shared the classroom with me during my first ever economics class at Tilburg University in 2010, and provided helpful comments on my dissertation in 2020.

A special mention must go out to my office mate Dominik Grothe! My time as a doctoral student would have been much less productive and fun without him. Going for swims in the *Eisbach*, regular efforts to keep Bavarian traditions alive and almost daily lunch breaks were a major reason I always enjoyed coming to the office. I sincerely hope that these days will continue!

My time as a doctoral student also gave me something that I cherish beyond any academic credentials. In April of 2019, a beautiful woman called Victoria walked into my office, and the rest is history. Words cannot describe what I feel for her.

Finally, everything I have ever achieved I owe to my parents, Rainer and Rosmarie. I have benefited in every possible way from their support and devotion. Thank you!

MICHAEL RAINER JOHANN KAISER
München, 14. September 2020

Table of Contents

Preface. 1

1. Minimum Wage Policy and Human Capital Investment 10
- 1.1 Introduction 11
- 1.2 Institutional background 18
- 1.3 Data sources 20
- 1.4 Estimating the causal effect of the minimum wage implementation . 26
- 1.5 Empirical results 30
- 1.6 Randomization inference 42
- 1.7 Conclusion 43

2. Charitable Donations to Natural Disasters – Evidence from an Online Platform 46
- 2.1 Introduction 47
- 2.2 Related literature 51
- 2.3 Data 55
- 2.4 Which natural disasters get funding? 64
- 2.5 Donor fatigue 70
- 2.6 Fundraising for disaster relief 78
- 2.7 Conclusion 83

3. Identifying and Teaching High-Growth Entrepreneurship in Uganda – Experimental Evidence from Academies for University Students in Uganda . . . 86
3.1 Introduction . . . 87
3.2 Research design . . . 91
3.3 Data . . . 104
3.4 Analysis . . . 112

4. Management Practices and Firm Performance – Evidence from Spanish Survey Data . . . 120
4.1 Introduction . . . 121
4.2 Data . . . 125
4.3 Estimating latent management styles . . . 128
4.4 Management style and firm performance *before* the Great Recession 136
4.5 Management Style and firm performance *during* the Great Recession 139
4.6 Conclusion . . . 145

Appendices . . . 147
Appendices to Chapter 1 . . . 148
Appendices to Chapter 2 . . . 169
Appendices to Chapter 3 . . . 189
Appendices to Chapter 4. . . . 194

Bibliography . . . 247

List of tables in Main Text

1.1 Implied hourly earnings by state. 24
1.2 Minimum wage bite and dropout rates. 32
1.3 Minimum wage bite and completion rates by school types. 36
1.4 Geographical variation in the minimum wage's effects. 39

2.1 Natural disaster severity: Casualties and number affected. 56
2.2 Donation demand & supply: Summary statistics. 60
2.3 Correlates of giving—summary statistics 63
2.4 Correlates of disaster relief donation demand and supply on Betterplace. 70
2.5 Effect of fundraising activity on disaster relief. 83

3.1 Overview of families of hypotheses. 97
3.2 Research project timeline. 99
3.3 Balance in entrepreneurship training sample. 111

4.1 Percentage of firms by size. 126
4.2 Five practices with highest loading in each style. 134
4.3 Correlates of Style 2 intensity. 135
4.4 Management style and firms' TFP before the crisis. 138
4.5 Style 2 intensity and firm performance during the crisis. 142
4.6 Management style and ease of adjustment. 144

List of Figures in Main Text

1.1 Geographical variation in wages and dropout rates. 16
1.2 Composition of school leavers across years. 21
1.3 Trends in dropout rates by minimum wage bite. 29
1.4 Quartile effects of minimum wage bite. 33
1.5 Year-by-year effects of the minimum wage on dropout rates. 35

2.1 Project fundraising & donation pages on Betterplace. 58
2.2 Disaster relief fundraising activity, donations, and casualties. 61
2.3 Pareto rule in disaster relief donations. 65
2.4 Disaster severity, fundraising, and donations. 66
2.5 Media coverage of disaster events. 67
2.6 Temporal fatigue. 73
2.7 Crowding out. 75
2.8 Disaster fatigue. 77

3.1 Experimental design and data collection. 93
3.2 Simulated statistical power. 103

4.1 Understanding styles by comparing single- and multi-plant firms. . . 133

Preface

> Without data, you're just another person with an opinion.
>
> *W. Edwards Deming*

The overarching theme and objective of my dissertation is to furnish empirical evidence through the careful and sophisticated analysis of data. The research described in the following four chapters addresses and answers questions in the fields of development, labor and organizational economics by developing insights grounded in rigorous empirical research. Especially in a time of frequent challenges of empirical evidence by opinions reflecting ideology, partisanship or populism, researchers must remain steadfast in relying on empirical evidence to inform arguments. I strive to derive recommendations for economic policy from the research papers presented in this dissertation that reflect arguments built on observed data rather than opinions about facts.[1]

The research laid out in this dissertation reflects on the one hand a personal and academic interest in the economics of low-income and developing countries, and, on the other hand, a keen interest in the methods of applied economics themselves. Low-income countries face myriad of challenges and obstacles, and my research works towards informing economic policy, in particular in the domains of youth labor and human capital policy. My work cautions policymakers to ponder the effects minimum wage policy may have on human capital investment, and offers insights on how entrepreneurship can empower young generations in the Global South. I also offer a perspective on how the Global North channels donations to the Global South in order to help it cope with the human toll and economic repercussions of increasingly

[1] The introductory quote's author, W. Edwards Deming (1900-1993), was an American scholar trained as an electrical engineer, mathematician and physicist. His academic contributions also included writings in statistics, psychology and management science. His consulting role in rebuilding the Japanese economy after World War II is widely acclaimed and earned him national honors in Japan. In the U.S. his contribution to the sampling techniques of the census are widely recognized.

frequent natural disasters. These insights are based on classic tools of empirical economics, and yet machine learning has recently been enriching our discipline with exciting and powerful methods that profoundly expand the realm of possibilities for research. In the last chapter, I present an application of unsupervised learning in organizational economics, and discuss research results that show how firms adapting a more flexible management style can better weather times of economic hardship.

The work and results in the present text build on more than a century of Mathematicians, Statisticians and Economists developing ever-more potent analysis techniques. These powerful tools and methods, in combination with widespread availability of data and growing computing power, now enable researchers to provide detailed recommendations to policymakers across a wide range of settings and applications. My dissertation is testament to this versatility, and illustrates four distinct combinations of data source and technique in different fields of economics. In doing so, this work also bespeaks the protean nature of economics as a discipline entertaining multifarious scientific inquiry.

The overwhelming share of empirical work in Economics is concerned with establishing and quantifying a *causal relationship* between two quantities (Imai, Keele, Tingley, and Yamamoto, 2011, p.1).[2] Practitioners of empirical research routinely distinguish between two paths to drawing causal conclusions, an endeavor always complicated by the fact that one is unable to witness an individual make the exact same decisions more than once. On the one hand, there are experimental methods which allow researchers to shape specific parameters of individuals' decision environments. On the other hand, *quasi-experimental* methods exploit naturally occurring variation in individuals' decision spaces. More recently, data-driven empirical methods whose end is not necessarily causal inference have expanded researchers' toolkits, and bestowed upon them the opportunity to delve into new questions. In economics,

[2] At this point I would like to point the reader to a book which has been indispensable to this dissertation. *Mostly Harmless Econometrics: An Empiricist's Companion* by Joshua Angrist and Jörn-Steffen Pischke has exactly been to me what its title claims, a companion. It provides a rigorous and comprehensive yet intuitive treatment of the major topics in applied econometrics and causal inference.

these methods are commonly referred to as *machine learning*.[3] In contrast to causal inference, machine learning helps us to make accurate predictions from or find natural groupings in observed data. My thesis draws on and combines tools from all these spectra, thereby highlighting their adaptability, complementarity and potential, in four distinct settings.

Oftentimes there appears to be a subtle dissonance in the perception of machine learning by practitioners of causal inference. By relying on and combining methods from both paradigms, I endeavor to demonstrate that this tension is illusory. I begin from an apprehension that machine learning has little to add to the identification of causal effects; its fundamental goal is different. Rather than pinning down a *causal* effect of one variable on another, machine learning is most commonly concerned with finding the best possible prediction of a variable given a set of predictors. With this in mind, machine learning would not offer answers to those questions I pose in Chapters 1-3. Instead, I rely on *identification strategies* which exploit exogenous variation to quantify causal effects.[4] However, some questions in economics, and in the social sciences more generally, are not fundamentally questions of causal inference; rather, they can be characterized as *prediction problems*, and, as such, lend themselves to be answered with the help of machine learning (Kleinberg, Ludwig, Mullainathan, and Obermeyer, 2015). Chapter 2, for instance, provides a brief perspective on how machine learning can complement causal inference by comparing the predictive performance of various models. In Chapter 4, my co-authors and I leverage a powerful machine learning algorithm to develop a new source of data; something methods of causal inference would not have allowed us to accomplish. It is by the realization that traditional methods of causal inference and machine learning address fundamentally different objectives, but can complement one another occasionally, that empirical research produces the most reliable and robust answers to questions of our time.

[3] Today, the term machine learning is almost omnipresent. Therefore, I feel compelled to provide the reader with a general but powerful definition, courtesy of Mitchell (1997). He postulates that in order to be understood as machine learning, a program (algorithm) learns from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience in E. In a typical machine learning application, an algorithm's prediction of an object based on data improves if it is provided with more instances from which it can *learn* about the mapping of data to the object.

[4] The term *identification strategy* was coined by two pioneers of empirical economics, Joshua Angrist, of MIT, and the late Alan Krueger, formerly of Princeton University, who describe a "*clearly labeled source of identifying variation in a causal variable and the use of a particular econometric technique to exploit this information*" (Angrist and Krueger, 1999, p.8). They stress the distinction between a *causing* and a *control* variable to highlight the fact that the former needs to fulfill particular requirements.

The sequence of chapters in this dissertation chronologically follows the creation of the linchpins of the empirical researcher's toolkit. While some of these methods date back centuries, they remain the most popular and powerful tools of empirical economics to this day (The Economist, 2016). The first chapter uses a *difference-in-difference* design—a concept dating back to the 1850s—to establish a causal link between minimum wage legislation and individuals' human capital investment. The second chapter draws on *instrumental variable* techniques—a notion conceived in the 1920—to estimate how much more natural disaster relief could be raised if online platforms were to wholly embrace their potential and uprooted barriers to fundraising and giving. The third chapter describes a currently ongoing *Randomized Control Trial* (RCT)—pioneered in Medicine in the 1940s, and in Economics in the 1960s—in Uganda which seeks to provide robust causal evidence for whether high-skilled youths can act as job creators through innovative entrepreneurship training. The fourth and final chapter blends *unsupervised learning*—a family of tools that have entered applied economics no more than 20 years ago—with an extensive firm survey and describes how firms' management correlates with their ability to weather times of economic upheaval.

Chapter 1 implements a difference-in-difference estimation framework. Intuitively, one compares the difference in an outcome of interest over time for different subgroups, and an early written account of such a framework can be found in Snow (1855). He compared Cholera-related deaths in different boroughs of London, and ascribed changes in death rates to different suppliers of water. The technique has since been applied to myriad of questions with the goal of establishing and quantifying a causal effect. A leading application has been the study of the effects of minimum wage legislation, dating back to Obenauer and von der Nienburg (1915). Arguably the most famous application is David Card and Alan Krueger's (1994) study of the effects of minimum wages on employment by observing changes in fast food employment in New Jersey and Pennsylvania around 1992, where the former raised the minimum wage but the latter did not. Today, the study is widely credited with changing economists' perceptions about consequences of minimum wage legislation.

My study of the German federal minimum wage of 2015 is comparable in spirit but veers in its objective. I exploit regional variation in the gap between the prevailing local wage in 2014 and the uniform minimum wage—the *bite*—to estimate the causal effect of the bite on high school dropout rates. Since the minimum wage raises the wage individuals, especially low-skilled ones, can expect to earn in the labor market, it may incentivize them to forego further investments in education. My findings are consistent with this channel, for I estimate that dropout rates increased by five percent comparing counties at the inter-quartile range of minimum wage bite. These results point to a possibly unintended side effect of minimum wage legislation. As an increasing number of countries implement or start to enforce such legislation, my study provides an additional lens through which policymakers ought to ponder its possible effects.

The main finding of Chapter 2 is the product of instrumental variable (IV) estimation. While its origins can clearly be traced back to a book on agriculture, the question of who deserves intellectual credit is debated among economists and historians (Wright, 1928; Stock and Trebbi, 2003).[5] In the book, IV is used to solve the problem that demand and supply are jointly determined in equilibrium and exploits *external factors* (read: exogenous variation) to estimate coefficients on an endogenous variable. The name *instrumental variable* only came about in the 1940s, and has become a mainstay of econometric analysis ever since (Aldrich, 1993). Today, IV methods are still used to solve simultaneous equation models, yet a thriving literature applies the logic to address bias from endogenous regressors more generally, or evaluate local average treatment effects (Acemoglu, Johnson, and Robinson, 2001; Imbens and Angrist, 1994).

The analysis of Chapter 2 is comparable in character to the early rationale for using IV techniques. My co-authors and I study whether an online platform for charitable giving can help channel funds to the countless natural disasters that tend to be neglected by charities and donors. We match 680,000 time-stamped individual charitable donations made through a German online platform to 1720 natural disasters that occurred from 2013-2017, which cost over 86,000 lives and affected nearly a billion people around the globe. We utilize an IV approach that relies on exogenous variation in charities' cost of providing disaster relief to overcome the fact that the demand for and supply of disaster relief are jointly determined in equilibrium. We estimate that charities could raise twice the median current relief volume on the platform if they were to solicit funds for currently neglected marginal natural

[5] The treatment of what later became known as instrumental variable estimation is "buried" in Appendix B of said book. The fact that the appendix is a mathematical derivation, while the rest of the book is an extensive treatment of animal and vegetable oils, has lead academics to doubt that Philip G. Wright penned Appendix B himself. Some historians credit his son, Sewall, accidentally a statistician, for writing the seminal Appendix. I refer the interested reader to Stock and Trebbi (2003).

disasters. In additional results, we show that giving to disaster relief does not crowd out giving to other charitable causes, and provide systematic evidence that charities and donors tire of fundraising for and donating to natural disasters in the wake of large disasters.

In Chapter 3, I present work that takes another approach to credible causal identification. Rather than relying on naturally occurring variation, my co-authors and I randomly vary exposure to a *treatment* and estimate its causal effect. Commonly termed a RCT, this has become the gold standard of causal inference and policy evaluation. While the use of experiments in other disciplines dates back centuries—clinical trials can be traced back to at least Lind (1772)—the science of economics is a much younger consumer of credible causal evidence by virtue of RCTs. Jamison, Searle, Galda, and Heyneman (1981) are generally credited with pioneering them in development economics. To this date, RCTs are most widespread in the economics of development and education, a fact bespoken by 2019 Nobel Memorial Prize in Economic Sciences for Esther Duflo, Abhijit Banerjee and Michael Kremer.

The randomized experiment I present in Chapter 3 traverses development and education economics.[6] We randomly offer admission to an entrepreneurship training to university students in Uganda, and track employment outcomes—focusing on self-employment—for up to three years after the training. The project contributes evidence to a nascent research paradigm focusing on the psychology of entrepreneurship to equip individuals in the Global South with the tools to earn a livelihood and create businesses that provide employment opportunities. At the time of writing, we have implemented two of three waves of the training program—including baseline and implementation check surveys of the first two waves—and expect to continue collecting data in 2021 and 2022. Due to the ongoing Covid-19 pandemic, implementation of the third wave has been postponed to at least the spring of 2021.

The fourth and final chapter of this dissertation offers an application of machine learning in the realm of organizational economics. In contrast to the previously discussed methods, economists have only started to explore and benefit from machine

[6] The RCT is currently ongoing, and we are expecting to collect the first wave of outcome data in the fall of 2020. My co-authors and I have produced a detailed pre-analysis plan published on the American Economic Association RCT registry #4502 (DOI: `https://doi.org/10.1257/rct.4502-2.0`).

learning in the last 20 years (Athey, 2019).[7] Advances in statistics and computational power have spurred the development of increasingly powerful algorithms. Practitioners typically distinguish between *supervised*—using a set of features to predict an outcome—and *unsupervised*—finding clusters of similarities in the data—algorithms (Athey and Imbens, 2019). Especially the latter presents researchers with the opportunity to develop inputs for econometric analysis from abstract high-dimensional and unstructured objects, such as text, speech or images (Mullainathan and Spiess, 2017).

The analysis of Chapter 4 leverages unsupervised learning to build a scalar-valued representation of firms' management practices from an extensive survey conducted in Spain in 2006. My co-authors and I apply a probabilistic clustering algorithm, *Latent Dirichlet Allocation*, to more than 270 indicators which describe a firm's human resource policies. Specifically, the algorithm projects management styles onto a space enclosed by two extremes which we interpret as informal vs structured management. We proceed by showing management geared more intensely towards the latter allowed firms to prosper during the economic expansion from 2001-2006. Interestingly however, this correlation reverses its sign in the economic downturn of 2007-2010. This suggests that by remaining flexible and adaptable firms sacrificed performance during the expansion but were able to better weather the Great Recession.[8]

The chapters of this dissertation carry implications for policymakers in labor and development economics, as well as for practitioners of organizational design. The final chapter also illustrates how unsupervised learning can be used to develop new data sources—something that holds great promise in economic research.

Chapters 1 and 3 seek to inform youth labor market policy, especially in the Global South, where youth unemployment is a pervasive and imperative issue. "Generation jobless" is how The Economist (2013, 2020a) termed the 300m 15- to 24-years-olds around the world—and especially in Africa—who are unemployed. Chapter 3 evaluates labor market outcomes of high-skilled youths in Uganda after pro-

[7] In this paper, Susan Athey—one of the researchers at the forefront of applying machine learning in Economics—puts forward a narrower and arguably less abstract (than the one given above) definition of machine learning. She defines machine learning as the "*field that develops algorithms designed to be applied to datasets, with the main areas of focus being prediction (regression), classification, and clustering or grouping tasks* (Athey, 2019, p.3).

[8] The use of the world *correlation* is intentional and seeks to convey that we abstain from drawing causal conclusions in this setting. We are unable to exploit exogenous variation in firms' choice of management style; rather, management style, as we observe it, is the product of past firm performance which in turns affects current and future firm performance.

viding them with entrepreneurial capital. The study assesses a targeted intervention that could easily be replicated and, more importantly, scaled up and deployed nationwide. With a majority in the Global South working in subsistence self-employment, this study aims to provide novel insights by explicitly targeting high-skilled youths prior to entering the labor market.[9]

The findings in Chapter 1 provide another perspective on labor market policy and its potential effects on the youth. Minimum wage policy is a prominent and wide-spread policy, with 90 percent of countries having some form of statutory wage floor in place (The Economist, 2020b). Enforcement varies widely, and is typically low in countries of the Global South. My study addresses possibly unintended side effects of minimum wage legislation, and uncovers a causal relationship between minimum wage policy and individuals' education outcomes. Policymakers in the Global South should thus ponder effects on youths' education when they choose to impose or enforce minimum wages more stringently.

More generally, economic growth in the Global South is frequently impeded by natural disasters destroying physical and social infrastructure.[10] Through humanitarian aid, the Global North can alleviate the consequences and foster reconstruction. However, the donor community typically only reacts a few major events; yet, a large number of smaller-scale disasters, that attract little attention and no donations, wreck the Lion's share of havoc. The findings in Chapter 2 provide a perspective on a novel tool, online platforms, for soliciting donations to disaster relief. I explore reasons for why disaster relief is typically deficient after smaller-scale disasters, even in relative terms. The study also points out that online platforms could solicit greater disaster relief, and thus play a more prominent role in smoothing effects of natural hazards.

While the last chapter does not dovetail thematically with the previous three chapters, it offers methodological insights and a topical perspective on organizational policy. The analysis showcases unsupervised learning, a novel tool in empirical

[9] With few opportunities in the labor market, self-employment is often the only way for individuals to earn a living. The overwhelming majority of these business fail to grow beyond subsistence and do not contribute to economic growth or provide jobs. Entrepreneurship education has been hailed as a solution by providing the human and managerial capital necessary to establish innovative businesses. Yet, World Bank lead economist David McKenzie and Chris Woodruff of Oxford University (2014) note that the body of work on business and entrepreneurship training in low-income training focuses on middle-aged, existing business owners, and results only show meager impacts.

[10] The U.S. National Academy of Sciences (2019) notes that in the wake of climate change, natural disasters are likely to become more frequent and that their prevalence has broadened.

economics, and how it can be used to reduce the dimensionality of high-dimensional objects. Our specific application analyzes how firms' management style affects their ability to absorb shocks, a highly topical question in light of the ongoing Covid-19 pandemic. Our results show that firms whose management was less rigid and more flexible forewent profits in times of economic expansion but weathered time of economic hardship profoundly better.

1. Minimum Wage Policy and Human Capital Investment

Abstract

Does the introduction of a minimum wage affect students' schooling decisions? Using county-level data from Germany, I show that high school dropout rates increased more in places where the federal minimum wage introduced in 2015 had greater bite. Specifically, I identify the causal effect by exploiting regional variation in pre-2015 wage levels. The difference in minimum wage bite between the county at the 25^{th} and 75^{th} percentile results in a roughly 5 percent increase in dropout rates relative to 2014 levels. The timing of the effect suggests an immediate response in 2015 and 2016, and the effect starts to fade out in 2017. The effect is predominantly driven by male students in rural areas. Permutation tests further corroborate the finding of the minimum wage having affected youths' schooling decisions.

1.1 Introduction

Across an ever-increasing number of nations around the world, minimum wage legislation has become an indispensable tool of economic and welfare policy. The Economist (2020b) recently published a piece on minimum wages, which noted that 90% of countries have some form of statutory wage floor in place but also cautioned that enforcement varies widely. The article, and indeed much of economic research, focuses on understanding employment and wage effects of such minimum wage legislation. Yet, wages play an outsize role in economic thinking, and minimum wage effects likely extend beyond employment outcomes. A nascent recent literature, for instance, links increased household income from minimum wages to improvements in birth weight and child health (Komro, Livingston, Markowitz, and Wagenaar, 2016; Wehby, Lyu, Kaestner, and Dhaval, 2020).

Wages are also central in how economists think about human capital investment. Put simply, the wage individuals could earn in the labor market constitutes the opportunity cost of investing in human capital. By raising the wage individuals could expect to earn in the market, minimum wage legislation may affect individuals' level of human capital investment. Indeed, Card and Krueger (1995, p. 214) state in their seminal book on the economics of minimum wages that "*school enrollment should be treated as outcome measure that is possibly influenced by the minimum wage*".

I focus on precisely this dimension of minimum wage policy and seek to answer the following question: what effect does the introduction of a minimum wage have on students' schooling decisions? I analyze this question in the context of the introduction of the federal minimum wage in Germany in 2015. At its introduction, 4 million people, roughly 10 percent of the labor force, earned less than the minimum wage, albeit there was profound regional variation (Mindestlohnkommission der Bundesregierung, 2016). I combine different sources of publicly available data and assemble a county (ger: *Landkreis*) panel dataset in order to test whether the reform led students to drop out of high school. This setting allows for precisely quantifying the minimum wage's effects on human capital investment at teen age.

The conclusions of this paper are based on exploiting the distance between a county's wage level and the uniform federal minimum wage—the *bite* of the minimum wage. The minimum wage is said to *bite* or *bind* more in counties with lower wage levels in 2014. When setting a uniform minimum wage, a policymaker is unable to account for individual counties' economic conditions. The resulting, plausibly exogenous, variation in the bite allows me to identify the causal effect of the minimum wage introduction on individuals' human capital investment. I measure students' human capital investment using counties' dropout rates, that is, the proportion of students who drop out of high school each year without a diploma. Finally, I draw on the theory of permutation tests to provide exact p-values for the hypothesis that the minimum wage had no effect on dropout rates.

The minimum wage may directly influence individuals' schooling decisions by triggering expectations of higher wages for low-skilled labor. Consider a *marginal* individual whose productivity would only slightly increase with one more year of education: to the extent that this slight increase does not raise productivity above the level of the minimum wage, there is little incentive to pursue such education. Another possibility is that individuals anticipate the minimum wage destroying low-skilled jobs, and thus they invest more into education. I provide *reduced-form* estimates which can be understood as an assessment of which channel plays a more prominent role.[1]

By analyzing dropout rates, I study the education decisions of individuals of about 16 years of age.[2] The minimum wage does not legally apply to underage minors—details on the institutional context in Section 1.2—and those who drop out of high school are not covered by the minimum wage. Thus, dropping out of the school does not immediately lead to remuneration under the minimum wage but only at such time when the individual turns 18. This should lower the incentive to *respond* to the minimum wage law by dropping out of school. Students may still anticipate that their hourly earnings never fall below the minimum, and adjust their education decision ahead of time.

Yet despite the fact that the minimum wage does not normally cover students right after they drop out of school, I do find that the minimum wage introduction caused dropout rates to increase. I estimate that the dropout rate increased by 1.14 percentage points on a base of 5.7 percent, an effect of 20 percent, when comparing the counties with the least and most minimum bite. The implied effect when com-

[1] The German economy was in an expansionary period around 2015—the setting of the present study. Therefore, it is unlikely that young individuals—those who can adjust their schooling decisions—would think that the minimum wage impedes their chances of finding work. For completeness it should also be noted that the minimum wage could give rise to an *income effect*. To the extent that education is a normal good, the increase of future earnings (assuming a constant probability of finding a job) would lead an individual to consume more education.

[2] I would like to note that I do not use any individual-level data. All data analyzed in this work is aggregated data that does not allow for tracking individuals, and was published by official sources of the German government.

paring counties at the first and third quartile is 0.3 percentage points, or 5.3 percent of 2014 levels. Effects are precisely estimated and statistically significant at the 5 or 1 percent level.

The results suggest a non-permanent short run response as the effect is concentrated in the first two years after the reform. In 2017, the effect size is no longer statistically significant, and the point estimates for 2018 suggest that the effect disappeared. In the immediate aftermath of its implementation, the minimum wage was most salient and thus most likely to feature in individuals' decision making. Moreover, as marginal jobs began to disappear individuals may have updated their beliefs about the minimum wage's effect on their labor market prospects (Caliendo, Fedorets, Preuss, Schröder, and Wittbrodt, 2018; Garloff, 2019).

The second set of findings uncovers relevant socio-economic margins of effect heterogeneity. First, the effect is largely driven by male students. Moving from 25^{th} to the 75^{th} of minimum wage bite implies a 6.3 percent effect on male students' dropout rates (base = 6.7 percent). This is about twice times the effect on female students' dropout rate (base = 4.5 percent). The effect for male students is significant at the 1 percent level while the effect for female students is marginally insignificant ($p = 0.15$). Theoretically, one would only expect that only those marginal individuals described earlier would adjust their education decisions. If more men than women are in the group of such individuals, then the "complier" group of men is larger. Research has shown that this is likely to be the case. First, young female have higher cognitive ability compared to their male counterparts (Becker, Hubbard, and Murphy, 2010; Fortin, Oreopoulos, and Phipps, 2015). Second, young females are likely to have more elaborate educational expectations (Bertrand and Pan, 2013).

I further address two margins of spatial heterogeneity that are of interest to policymakers. First, I show that the effect is driven in equal parts by urban and rural areas. Second, I evaluate whether the effect is different in those states that formed the German Democratic Republic (GDR) until 1990, and I am unable to reject that the effects are equal.

To further corroborate the finding of a link between the minimum wage introduction and educational investment, I conduct permutation tests. Permutation tests are an established statistical method that have gained increasing attention in Economics, and allow the researcher to test the null hypothesis of no *individual* effect (Fisher, 1935; Young, 2019). The tests are based on a large number of random permutations of the "treatment"-vector— the minimum wage bite—to construct a randomization distribution of effect sizes under the null hypothesis of no effect. I obtain exact p-values below 0.01 based on 999 permutations. These results serve to corroborate the existence of a link between the minimum wage reform and dropout rates.

This study contributes to the literature on understanding minimum wage effects beyond first-order employment effects. In doing so, I build on sparse and geographically concentrated evidence on a link between minimum wage legislation and skill acquisition by the youth. Research exploiting cross-state variation in the US suggests a link between minimum wages and educational attainment (Neumark and Wascher, 2003; Chaplin, Turner, and Pape, 2003; Sutch, 2010).[3] Similarly, evidence points to an inverse link between minimum wages and school enrollment among teenagers in New Zealand (Pacheco and Cruickshank, 2007; Hyslop and Stillman, 2007). In contrast, Campolieti, Fang, and Gunderson (2005) find no robust association between school enrollment and minimum wages using Canadian data.

The contribution of the present paper is to extend the evidence to a setting outside the Anglo-Saxon context in which an unambiguous measure of teenage skill acquisition is available. While much of the related literature is complicated by enrollment measures depending on labor force participation, the structure of the German high school system permits precise measurement of high school completion rates. Clemens, Khan, and Meer (2018) shows that employers "upskill"—substitute towards higher-skilled labor—following minimum wage increases in the US. If individuals simultaneously reduce human capital investment the ensuing wedge between employers desired and workers acquired human capital reduces employment prospects for those individuals. Thus an understanding of possibly unintended and unanticipated effects of minimum wage legislation on education carries policy relevance.

My setting allows me to exploit finely grained variation in minimum wage bite—*within states*—for identification of the effect of dropout rates of students who are exposed to the same educational policies. In contrast, in US settings the effect is usually identified from variation of minimum wage and dropout rates across states where labor market and educational policies also differ across states.

This research further contributes to understanding short-term effects of the federal minimum wage in Germany. Early research suggests that the minimum wage

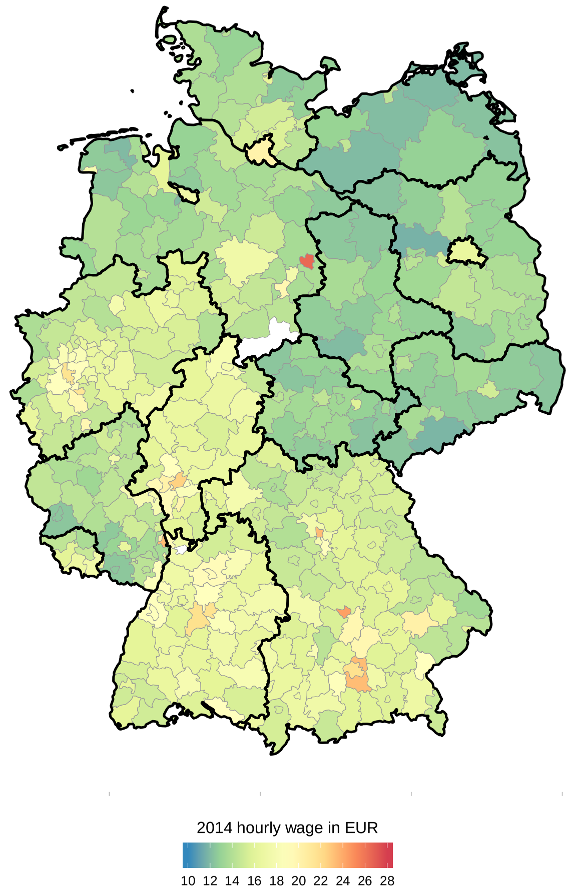
[3] Neumark and Wascher (2003) builds on an earlier article that examined data of the 1970s and 1980s (Neumark and Wascher, 1995). The latter version updates the data to include the Current Population Surveys (US) conducted in the 1990s. The results in the earlier study mirror those in the later one.

only had marginal effects on employment (Bossler and Gerner, 2019; Caliendo, Fedorets, Preuss, Schröder, and Wittbrodt, 2018). Caliendo, Schröder, and Wittbrodt (2019) provide an overview of short-term effects but do not discuss effects on human capital acquisition. Dustmann, Lindner, Schönberg, Umkehrer, and Vom Berge (2019) find that the reform led to significant wage gains at the lower end of the distribution while not reducing employment prospects for low-skilled individuals. I extend this literature by analyzing the short-term educational response. If individuals reduce human capital investment in response to the reform, it may have adverse effects on their medium and long term career prospects, particularly in light of increasing skill requirements in the labor market (Autor and Handel, 2013).

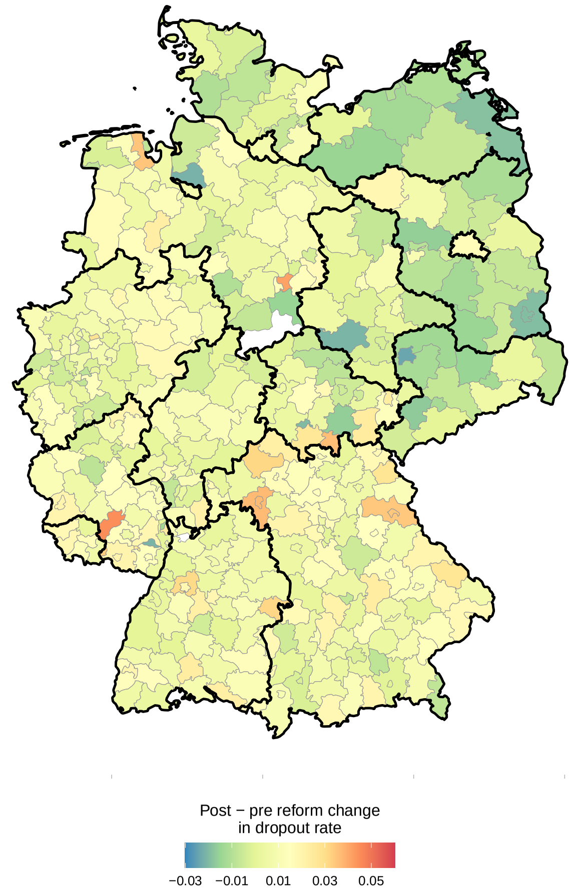
Significant regional variation in wage levels and dropout rates renders Germany an ideal setting for this study. The maps in Figure 1.1 demonstrate this variation. First, the left hand panel shows profound variation in average wages across counties in 2014. Crucially, this variation also exists within states, and is key to the identification strategy which relies on state fixed effects. Second, the right-hand panel similarly illustrates variation in the change in dropout rates. Education policy is administered at the state level, but panel (b) suggests that there is ample within-state variation in dropout rates. Further, the maps illustrate the fact that wages tend to be lower in East Germany—the former (GDR), roughly the north-east of Germany.[4] Visual inspection suggests that dropping out is less frequently observed in those states. Socio-economic discrepancies persist to this day to the extent that there was discussion of different minimum wages for East and West Germany at the time of policy deliberation (Bellmann, Bossler, Gerner, and Hübler, 2015). My results do not suggest a strongly heterogeneous effect, albeit the effect in East Germany is smaller in size and less precisely estimated.

The results presented in this paper are of interest to policymakers as they point to possibly unintended side effects of minimum page policy. As noted above, minimum wage legislation is becoming an increasingly popular tool. While many countries have statutory minimums in place, few countries in the Global South enforce them, or possess the capacity necessary for enforcement. The results presented in the current study carry over to a setting in which a country starts to enforce a minimum wage which it has had in place. My results stress that policymakers may have to take into account that (enforcing) minimum wage policy can have adverse and unintended

[4] The latter states are also sometimes called the "new" states while the former are called "old" states. I will use this terminology interchangeably. In the analysis, I will include state by year fixed effects which accounts for time-invariant underlying differences between East and West Germany.



(a) Implied hourly wages in 2014



(b) Geographical variation in dropout rates

Figure 1.1: Geographical variation in wages and dropout rates. *Notes.* Panel (a) plots the 2014 hourly wage. Darker shades of blue indicate lower values (min = €11.70) and lighter shades of blue indicate higher values (max = €26.50). Panel (b) shows the geographic variation in the change in dropout rates across German counties. The change is computed as the average post-reform level (2015-2018) minus the average pre-reform (2013 and 2014) level. The black borders indicate state borders while the light grey borders show county borders. The white county in the middle is the county of Göttingen which is missing due to incomplete schooling records.

effects on human capital investment.

1.2 Institutional background

This section briefly provides background on the minimum wage legislation and the education system. First, how was the federal minimum wage introduced and which part of the labor force is covered? Second, I discuss the hierarchical structure of the German education system.

1.2.1 Federal minimum wage

Germany introduced a federal minimum wage on January 1^{st}, 2015, establishing a wage floor at €8.50 gross per hour. The minimum wage has since been raised four times and is currently set at €9.35 as of January 1^{st}, 2020. Prior to 2015, wage floors were established by collective bargaining or voluntary agreements which could have stipulated wages higher than €8.50 in some cases. Moreover, a number of sectors (e.g., construction, hairdressing or security services) had sector specific minimum wages prior to 2015. Fitzenberger and Doerr (2016) present a comprehensive overview of sector-specific minimum wages.

Several employment relationships and individuals were exempt from the minimum wage. Self-employed, trainees, specific interns, minors without vocational training, volunteers, inmates and long-term unemployed (exempt for six months) were not covered by the wage floor (Caliendo, Fedorets, Preuss, Schröder, and Wittbrodt, 2018). Despite these arrangements, the minimum wage bound for roughly 4 million (≈10 percent of the labor force) out of the 5.5 million employees who earned less than €8.50 in 2014 (Mindestlohnkommission der Bundesregierung, 2016).

A federally binding minimum wage was put forward after the federal elections in the fall of 2013. The incoming government under chancellor Merkel endorsed a federal minimum wage in its coalition treaty. The government's proposal passed both houses of parliament in July 2014 and went into effect five months later. This timeline suggests that anticipation effects may have played a role, but in Section 1.5.1 I show that there is no evidence of such effects.

1.2.2 Education system

States rather than the federal government are responsible for education. Yet, across all states the high school system can be divided into three tiers. This stands in contrast to, for instance the US, where there is generally one type of high school. The top tier of the German high school system, called *Gymnasium*, is meant to prepare students for tertiary education, and students graduate with a university entry qualification. The middle tier, most commonly called *Realschule*, historically prepares students for vocational training in office jobs and skilled manufacturing occupations. In recent years, the system has become more fluid in that it allows students from Realschule to pursue university education, by attending an intermediate school type upon graduation. Finally, the lowest tier of the school system prepares students for low-skilled occupations. This tier is commonly called *Hauptschule*. There are differences in terminology, and details on how students can move between tiers differ across states; importantly however, this general type of academic hierarchy holds across states.

Crucially, leaving school without any degree is a well-defined concept across all states. This event is defined as leaving school without a *Hauptschulabschluss*, i.e., a passing grade in the final year of the lowest tier of the high school system. Should a student not finish the last year of Hauptschule, commonly the ninth grade, with at least a passing grade, she is considered a dropout. Students are typically between 15 and 17 years of age when they drop out of school.

Broadly speaking, there are three scenarios that result in a student becoming a dropout. First, a student simply fails to complete the last year of Hauptschule and decides against another try. Second, she completes the mandatory years (see below) of high school education prior to reaching the last year of high school. This occurs if a student has to repeat a year because of a failing grade point average. In that case, she can decide to leave high school. Finally, a student may be prevented from trying to complete the last year for disciplinary reasons if she has completed the mandatory years of schooling already.

Different mandatory schooling laws are in place across states but they share important parallels. All states stipulate that individuals must complete nine or ten years of primary and high school.[5] In most cases those nine or ten years must be followed by two or three years of vocational school. Mandatory schooling laws no longer apply when an individual turns 18 (or 17 in some states). Depending on the number of repeated attempts at a particular grade, students may turn 18 without ever having attended vocational school.

Students who finish (or drop out of) high school without having abode with mandatory schooling laws ought to attend vocational school. This is structured

[5] The German media outlet Süddeutsche Zeitung pubslishes an overview of state-specific regulations regarding the minimum number of years across states under `https://bildung.sueddeutsche.de/schulpflicht/` (in German).

through apprenticeships where students split their time between working in a firm and attending school studying occupation-specific content. Having finished high school, students would typically apply to apprenticeships (based on personal preference) to start job training. Dropouts have access to the same type of system but lack any formal high school qualification. Should they be unable to acquire an apprenticeship, special tracks at vocational schools exist in order for them to be able to complete mandatory years of schooling. Upon completion, they are considered part of the labor force but lack any official education or job certificate. Thus, leaving without a Hauptschulabschluss is not a dead end in the education system. It is considered an adverse signal that restricts an individual's labor market choice set since formal requirements prevent dropouts from applying for a range of jobs.

1.3 Data sources

1.3.1 Schooling data

The Federal Statistical Office publishes completion rates for all tiers of high schools in Germany. The data also contains the number of students who left the high school system without having completed the lowest tier. The data is available for all 399 counties that are part of the analysis, and the observation period runs from 2013 to 2018.[6]

Figure 1.2 summarizes the composition of school leavers across years. About five to six percent of students leave high school without a formal degree. This proportion has been growing from 5.6 to 6.8 percent from 2013 to 2018. The modal student (about 45 percent) graduates with the middle tier degree. About one third of students acquire university entry qualification via the top high school tier, while 17 percent of students successfully complete Hauptschule. The share of students completing Hauptschule decreases over time, comparable in magnitude to the increase in the dropout rate. Dropping out is more common among male students (7.4 percent) than among female students (4.7 percent).

Appendix Table 1.A.1 reports summary statistics of dropout rates over time. The spread between counties is about 14 percentage points. The distribution of dropout

[6] There are 401 counties in Germany but two counties are dropped from the analysis sample. The county of Göttingen is excluded since it is the only county with incomplete schooling records. The county of Heidelberg is excluded as it shows an increase from 24 (out of 1491, 1.6 percent) in 2016 to 127 (out of 1568, 8.1 percent) dropouts. This exceeds all other year-to-year changes in the data by an order of magnitude.

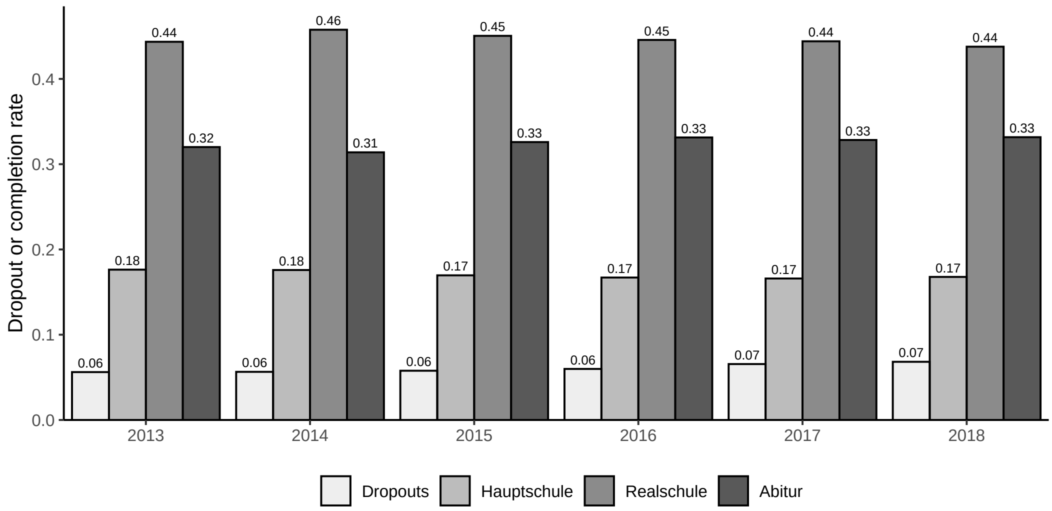


Figure 1.2: Composition of school leavers across years. *Notes.* This figure shows the average dropout rate and type-specific completion rates across counties in Germany. The number of dropouts and the number of graduates of the respective school type is divided by the total number of school leavers in a given year. The labels above the columns indicate the numerical value.

rates appears well-behaved with the inter-quartile range being equally spaced around the median. In absolute numbers 46,733 students dropped out of the high school system in 2014. On average, 117 individuals dropped out of school in each county in 2014.

1.3.2 Hours worked

Data about hours worked is a necessary ingredient to calculating hourly wages from monthly earnings. This data comes from the Structure of Labor Earnings Survey 2014 (ger: *Verdienststrukturerhebung*, VSE) which is conducted by the German Federal Statistical Agency and is representative at the state level (Statistisches Bundesamt, 2014). The VSE is a comprehensive survey regarding employment rates, earnings and non-pecuniary benefits and, crucially, hours worked. The data reports average weekly hours in April 2014 for 19 sectors across all German states. Appendix Figure 1.A.2 shows the average as well as the minimum and maximum (that is, the number of hours worked in the least and most intensive sector) across all states. This data is only available at the state level. There is little variation in hours across states; on average individuals work between 39 and 40 hours a week. At the upper end employees in Sachsen-Anhalt work 40.32 hours while employees in Saarland work only 39.23 hours.

1.3.3 *Counties' economic indicators*

The minimum wage's bite is a function of the pre-reform average prevailing wage level. A lower wage causes the minimum wage to bite more, and vice versa. If few people are directly affected by the new wage floor, the consequences for the local labor market are unlikely to be economically meaningful. In order to construct a measure of minimum wage bite, I employ data from the "INKAR" database which is maintained by the Federal Ministry of the Interior, Construction and Homeland, and provides county level data on gross earnings (ger: *Bruttoverdienst*) and median earnings (ger: *Medianeinkommen*) (BBSR Bonn, 2019). I combine gross income and median earnings with the labor hours data from the previous section to construct counties' implied average hourly wages. Note that this results in two separate measures of counties' implied hourly earnings.[7]

Gross income reflects average monthly earnings of employees in a county. For this measure, all individuals in working relationships (employees, civil servants, soldiers, trainees, etc.) are counted, including individuals in minor employment (ger: *geringfügig Beschäftigte)*. Self-employed individuals are not included for the construction of this indicator. The gross wage is then the sum of salary, employer-provided perks and social welfare contributions made by the employer. Only payments by domestic organizations are considered, and hence income received from foreign entities would not show up in this measurement.

Median earnings (ger: *Monatsentgelte der Vollzeitbeschäftigten*) measures the median level of gross labor earnings for full time employees reported to the Federal Agency for Employment (ger: *Bundesagentur für Arbeit)*. Unlike gross income, this variable is only available from 2014 onward. Median earnings are about 20 percent higher than gross income. The fact that only full time employees' earnings are included in the former accounts for this fact.

I then calculate hourly wages implied by both measures of earnings assuming four weeks per month. Thus monthly earnings are divided by weekly hours times four, and Table 1.1 shows summary statistics across states. Average wages based on gross income tend to be highest in Hamburg and Hessen while Baden-Württemberg has

[7] Through the presentation of the results I provide estimates based on both approaches to constructing minimum wage bit. The results are quantitatively similar. The estimates based on median income tend to be larger in magnitude but with the spread in the underlying Kaitz-ratio being smaller, the implied effect sizes relative to 2014 are almost identical.

the second highest implied wage based on median income. Thus there are rank differences depending on the underlying quantity used to construct wages. Throughout the analysis, I will present estimates based on using both versions of implied hourly wages. Column 8 in Table 1.1 reports the number of counties in each state. The states of Hamburg and Berlin only consist of a single homonymous county. Foreshadowing the analysis, these will be absorbed by a state fixed effect.

The key explanatory variable in the analysis will be the bite of the minimum wage as measured by the *Kaitz-ratio.* This ratio is a non-linear transformation of the average prevailing wage relative to the nominal minimum wage. The Kaitz-ratio is given by $\frac{8.50}{w_{i,2014}}$. A larger value implies a stronger bite of the minimum wage. Note that the non-linearity of the transformation implies *increasing marginal bite* as wages decrease. Appendix Figure 1.A.3 shows a histogram of observed Kaitz-ratios in the data. Using wages implied by gross earnings the range is about 0.3 to 0.7 with a mean (median) of 0.55 (0.55) and a standard deviation of 0.076. The distribution based on median earnings is highly comparable with a mean (median) of 0.48 (0.46) and a standard deviation of 0.080.

A necessary assumption for the Kaitz-ratio to be valid measure of minimum wage bite is that it (rank) correlates with the fraction of individuals who earned below € 8.50 in 2014. Suppose that wage dispersion is larger in counties with higher average wages; in this case, the Kaitz-ratio may be low (high wages) and indicate little bite when in fact there is a significant number of individuals earning below minimum. While I cannot empirically test this assumption, comparing states with low implied wages in Table 1.1 and those identified to host a larger share of enterprises affected by the minimum wage in Bellmann, Bossler, Gerner, and Hübler (2015) suggests that they correlate.[8]

Appendix Figure 1.A.1 illustrates the correlation between Kaitz-ratios based on the two measurements of labor income. It plots the quantities' ranks against each other and compares them to the 45°-line which would indicate perfect rank correlation. The figure suggests high, albeit imperfect, rank correlation. The correlation is lowest in the middle portion of the distribution. The coefficient of correlation is 0.88 and indicates high linear dependence. Spearman's ρ of 0.92 further indicates high

[8] An alternative measure for the minimum wage bite would be a county's fraction of employees who earn below €8.50 in 2014. Yet, also this measure requires additional assumption as it does not take into account the distance between an employee's wage in 2014, and the minimum. The fraction would be equally affected by employees earning €8 or €6. Caliendo, Fedorets, Preuss, Schröder, and Wittbrodt (2018) find comparable effects on employment when they use both measures of minimum wage bite in an analysis of 2015 minimum wage reform in Germany.

Statistic	Mean (1)	St. Dev. (2)	Median (3)	Pctl(25) (4)	Pctl(75) (5)	Min (6)	Max (7)	N (8)
	Wages implied by gross income [€]							
Baden-Württemberg	17.15	1.61	16.82	16.04	18.16	14.68	21.84	43
Bayern	16.38	2.05	15.88	15.03	16.96	13.36	24.34	96
Berlin	16.50	0	16.50	16.50	16.50	16.50	16.50	1
Brandenburg	13.46	0.86	13.48	12.89	14.11	11.70	15.24	18
Bremen	16.80	0.39	16.80	16.43	17.18	16.43	17.18	2
Hamburg	20.15	0	20.15	20.15	20.15	20.15	20.15	1
Hessen	17.41	1.81	16.74	16.12	18.81	14.81	22.53	26
Mecklenburg-Vorpommern	12.89	0.97	12.57	12.07	13.31	11.99	14.53	8
Niedersachsen	14.49	2.40	13.88	13.30	15.02	12.00	26.50	44
Nordrhein-Westfalen	16.48	1.77	16.22	15.25	17.29	13.75	21.64	53
Rheinland-Pfalz	15.15	2.23	14.40	13.90	15.73	12.45	23.33	36
Saarland	15.48	1.30	15.33	14.50	16.79	13.67	17.25	6
Sachsen	13.13	0.88	12.74	12.53	13.58	11.84	15.17	13
Sachsen-Anhalt	13.09	0.61	12.99	12.61	13.80	12.06	14.00	14
Schleswig-Holstein	14.28	1.02	14.14	13.76	15.20	12.54	16.08	15
Thüringen	13.25	0.82	13.19	12.61	13.64	12.28	15.74	23
	Wages implied by median earnings [€]							
Baden-Württemberg	20.72	1.77	20.18	19.38	21.75	18.33	27.11	43
Bayern	18.96	2.27	18.64	17.50	19.83	15.74	28.04	96
Berlin	18.36	0	18.36	18.36	18.36	18.36	18.36	1
Brandenburg	14.27	1.25	14.28	13.15	14.73	12.47	17.21	18
Bremen	20.19	0.96	20.19	19.27	21.11	19.27	21.11	2
Hamburg	21.79	0	21.79	21.79	21.79	21.79	21.79	1
Hessen	20.07	2.43	19.26	18.04	21.71	17.03	25.07	26
Mecklenburg-Vorpommern	13.79	1.45	13.07	12.75	14.63	12.36	16.25	8
Niedersachsen	17.99	2.30	17.29	16.82	18.66	15.68	28.43	44
Nordrhein-Westfalen	19.61	1.48	19.43	18.45	20.11	17.37	24.97	53
Rheinland-Pfalz	18.63	2.11	17.86	17.30	19.88	15.58	27.23	36
Saarland	19.53	1.14	19.34	18.53	20.77	18.04	21.16	6
Sachsen	13.71	1.48	12.94	12.69	14.22	12.22	17.32	13
Sachsen-Anhalt	13.97	1.06	13.77	13.07	14.48	12.69	16.25	14
Schleswig-Holstein	17.30	0.98	17.40	16.41	17.96	15.53	19.60	15
Thüringen	13.91	1.28	13.58	13.06	14.28	12.45	17.54	23

Table 1.1: Implied hourly earnings by state. *Notes:* This table reports summary statistics of both measures of implied hourly wages across counties in states. "Pctl(25)" and "Pctl(75)" indicate the 25^{th} and 75^{th} percentile, respectively. The last column indicates the number of observations in a state which is equivalent to the number of counties in a state. Note that the states of Berlin and Hamburg only have one county each.

rank correlation.

Time-varying controls

In order to control for time-invariant factors, I include county fixed effects in the analysis. While this is likely to absorb significant variation—especially in addition to the state by year fixed effects—time-varying county characteristics may still influence individuals' education decisions. To address this concern, I control for counties' socio-economic situation using several socio-economic indicators measured annually. These indicators describe the equilibrium of labor demand and supply, and thus the signal about employment opportunities students receive. Different measures break vacancies down by skill group, as the aggregate unemployment rate may obscure low-skilled individuals' labor market situation. In addition, controlling for population inflows addresses the fact that individuals may feel compelled to invest more into education in order to remain employable.

Specifically, I include the following time-varying variables: i) the unemployment rate measured as a percentage of the labor force, ii) vacancies, divided into two classes; "helper" and "skilled" using a definition of the German Employment Agency, iii) inward (outward) migration, measured as the number of immigrants (emigrants) in a given year per 1,000 inhabitants in the county in the same year, iv) employment in the first (agriculture, forestry and fishing) and second (manufacturing, mining, construction, etc.) sector, measured as the percentage of total employment in the county in the same year respectively, v) gross local product which reflects the value of goods and services per inhabitant (rather than worker) in a locality, vi) transfers, as the sum of payments from state governments to municipalities in a county per inhabitant per year, and finally, vi) the fraction of the population under the age of 65 that receives social welfare payments.

Table 1.A.2 summarizes these covariates using their averages across the years 2013 to 2017.[9] There is ample variation in economic conditions across counties; the minimum local unemployment rate is 1.4 percent while the maximum is 14.5 percent. Similarly, counties are exposed to asylum seekers in varying ways. Finally, there are counties with 23.5 percent social security recipients, while other counties have as few as 1.2 percent.

[9] Data on all covariates is not available for the year 2018 as of August 2020.

1.4 Estimating the causal effect of the minimum wage implementation

1.4.1 Estimation setup

The objective is to identify the causal effect of the federal minimum wage on students' human capital investment measured by the fraction that completes at least basic high school. The level of observation is a county indexed by $i \in \{1, 2, ..., 399\}$. Observations are further indexed by state $s \in \{1, 2, ..., 16\}$.[10] The panel dimension is denoted by year $t \in \{2013, ..., 2018\}$. I then estimate the following equation:

$$y_{i,s,t} = \beta_1 1[t \geq 2015]_t + \beta_2 1[t \geq 2015]_t * \texttt{bite}_i + X_{i,t-1}\phi + \alpha_i + \alpha_{s,t} + \varepsilon_{i,s,t} \quad (1.1)$$

The dependent variable is a county's i dropout rate in year t, that is, the ratio of students leaving school without any degree to the total number of students leaving school. The indicator variable $1[t \geq 2015]$ identifies the years 2015, 2016, 2017 and 2018. The minimum wage went into effect on January 1^{st}, 2015, and thus the student cohort (supposed) to graduate in 2015 was the first to graduate when the minimum wage was in effect. I show that there do not seem to be anticipation effects at the end of Section 1.5.1.

The time-invariant variable $\texttt{bite}_i$ is the measure of *treatment intensity*, that is, a measure of minimum wage bite in county i. The treatment measure is based on the Kaitz-ratio constructed from the average hourly wage level in 2014 and the nominal minimum wage of €8.50. The coefficient of interest, β_2, multiplies the interaction between the post-reform indicator and the continuous measure of minimum wage bite. This specification assumes a linear effect of bite on dropout rates. I relax this assumption by supplementing said "linear-in-Kaitz" specification with a specification in which treatment is measured using quartile indicators for the Kaitz-ratio in 2014. Thus there are three interaction terms of indicators for quartiles 2, 3 and 4 with the post-reform dummy. The first quartile, the 25 percent of counties where the minimum wage bound *the least*, is the omitted group.

All specifications include county, α_i, and state by year, $\alpha_{s,t}$, fixed effects. The vector $X_{i,t-1}$ includes the first lag of time-varying county characteristics. Across states, the school year for those in graduation year classes finishes between late April and June, and prevailing economic conditions around that time are better

[10] Counties are perfectly subsumed in states. That is, there are no counties whose borders stretch across state lines.

captured by the preceding year's economic measurements.[11] Not all indicators are available for the year 2018 at the time of this analysis, and thus using the first lag also ensures that control variables can be included when the outcome measurement runs through 2018.

I distinguish between a *basic* and an *extended* set of time-varying covariates. The former is comprised by the unemployment rate; vacancies classified as unskilled or semi-skilled; and asylum seekers as well as a county's immigration and emigration numbers relative to population. The extended set of covariates additionally includes the share of employment in the primary or secondary sector; transfer payments from the state government to the county; social security recipients; and a county's economic output in terms of gross local product per inhabitant.

I cluster standard errors at the county level and thus allow for arbitrary autocorrelation in the error terms within a county (Abadie, Athey, Imbens, and Wooldridge, 2020). I discuss rationales for alternative ways of clustering based on a recent literature on inference in Section 1.5.4, and show that results are robust.

1.4.2 Identification

Identification of the causal effect relies on parallel trends of dropout rates over time. In the present context this means dropout rates have to evolve in a parallel fashion across counties irrespective of their 2014 wage level. The coefficient of interest is β_2 which captures the causal effect of differential minimum wage bite on dropout rates after 2015.[12]

Key to this argument is the fact that the minimum wage was set at a uniform level of €8.50 for all counties. This rules out concerns that the level of the minimum wage was endogenous to an imaginary county i's economic idiosyncrasies. Therefore, the relationship between a county's wage in 2014 and the minimum wage is plausibly exogenous.

As described in Section 1.2, education policy in Germany is administered by states. If states pursued different education policies around 2015 in ways systematically correlated with the average prevailing wage level in 2014, or economic conditions more generally, β_2 would not identify the causal effect of the minimum wage reform.

[11] In Section 1.5.4, I also estimate specifications in which either the contemporaneous value of the covariates, or their second lag, is included. The results are virtually unchanged.

[12] When minimum wage bite is measured using quartiles, β_2 is a coefficient vector of length three. The three elements capture the effect of quantiles two, three and four relative to the first, respectively.

To address this concern, I include state by year fixed effects, $\alpha_{s,t}$, in the analysis. This relaxes the parallel trends assumption in the sense that one only needs to assume that counties within a state would have seen parallel developments in dropout rates, regardless of their wage level in 2014 (Stephens Jr and Yang, 2014). Alternative specifications demonstrating robustness include state specific linear time trends, or state by post-reform fixed effects. Counties have no power in setting education policy which further establishes the credibility of the argument that education policy is not set in response to wage levels.

Figure 1.3 assess the plausibility of trends in dropout rates being independent of 2014 wage levels. Counties are grouped into quartiles based on their 2014 Kaitz-ratio. Dropout rates have steadily decreased from about eight percent in 2004 to about five percent in 2012. This decrease is highly comparable across groups. From 2012 onwards dropout rates were initially stable and then started to climb in 2014. While the first three quartiles demonstrate plausibly parallel trends in dropout rates, the trend prior to the observation window differs in those counties where the minimum wage bit hardest. Crucially, the development just prior to the reform is comparable. The specification with quartile dummies implicity takes this into account; the results point to an effect that is not driven by a comparison of the first and fourth quartile.

Appendix Figure 1.A.4 motivates controlling for differential time effects across states. There are pronounced differences in the development of dropout rates over time and space; compare for instance Bayern, Nordrhein-Westfalen and Schleswig-Holstein. The former shows a clear downward trajectory while the latter two show a less distinctive pattern.[13]

Another possible source of bias in this type of design stems from changes in the population's composition. Suppose that in counties where the minimum wage binds the most, families selectively emigrate since they fear losing employment opportunities. Assuming a positive correlation between family income and emigration opportunities as well as a positive correlation between family income and offspring's education, families whose members are unlikely to drop out of school are more likely to emigrate. That would tend to decrease the denominator of dropout rates but leave the numerator unchanged, thus increasing the dropout rate. In the analysis, I control for lagged values of in- and out-migration (both, domestic and foreign) to account for this possibility.

[13] Some panels show incomplete time series figures. In some cases this is an artifact of incomplete records in some years prior to the period of observation. In other cases, this is the result of using state-level quartiles to split the data, which can result in unpopulated quartile cells.

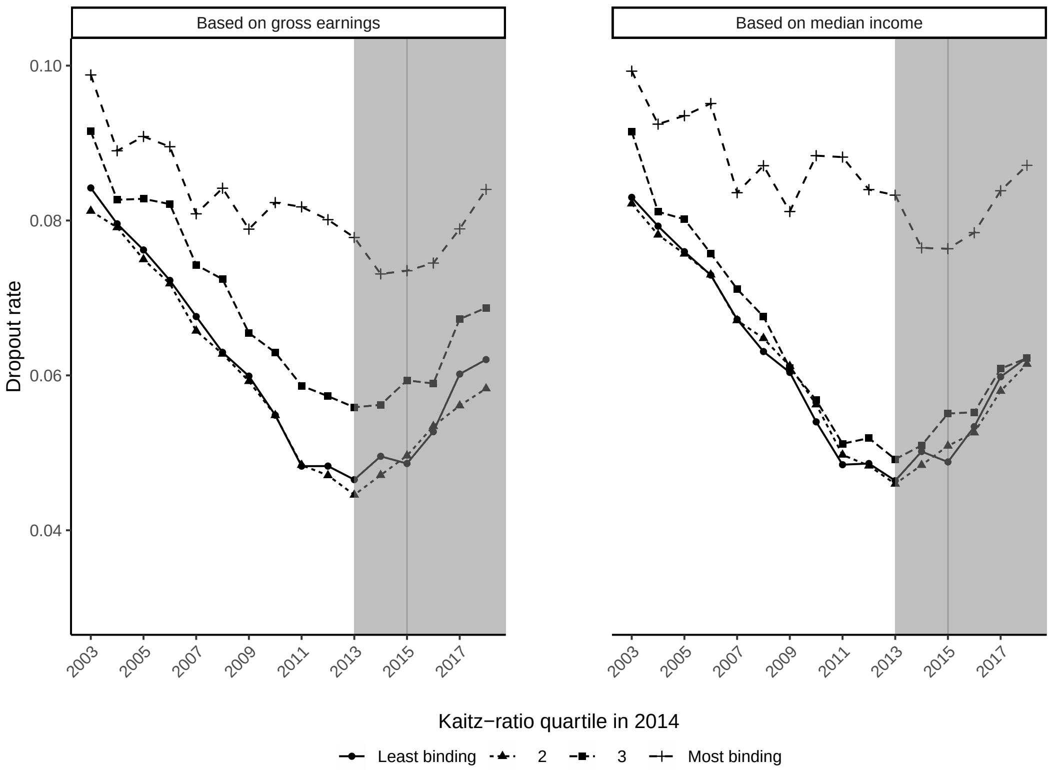


Figure 1.3: Trends in dropout rates by minimum wage bite. *Notes.* This figure shows the average Kaitz-ratio across counties from 2003 to 2017. The sample observation period (2013-2018) is shaded grey and the introduction of the minimum wage in 2015 is marked by a vertical line. Counties are placed into quartiles depending on the value of the prevailing Kaitz-ratio in 2014. The counties where the minimum wage was least binding are in the first quartile (i.e., bottom 25 percent, solid line with dots) while the counties with the most binding minimum (i.e., top 25 percent) wage are displayed by the dashed line with crosses. The left (right) panel shows the trend based on Kaitz-ratio constructed from gross earnings (median income).

Finally, the influx of asylum seekers whose educational status is difficult to establish may affect dropout rates. Teen asylum seekers may enter the school system, thus increasing the denominator of the dropout rate and location patterns may correlate with minimum wage bite. Appendix Figure 1.A.5 shows that immigration is unlikely to confound the effect of the minimum wage reform on dropout rates. It shows that a possible increase in counties' population through migration does not systematically correlate with the dropout rate. I do include the lagged number of asylum seekers as a control variable in the analysis.

1.5 Empirical results

This section first reports the results from the preferred specifications. In the next step, I present insights into the composition of the effect, and discuss margins of heterogeneity. Robustness checks conclude the section.

1.5.1 Benchmark results

This section presents estimates of Equation (1.1) that differ in i) whether the continuous measure or the quartile-based measure of minimum wage bite is used, and ii) whether the implied hourly wage is constructed from counties' average gross earnings or median income.

Table 1.2 reports the main estimates. Columns 1 (4) to 3 (6) use wages implied by gross earnings (median income). The coefficient of interest is the interaction term reported in the second row of the table. The causal effect of the minimum wage on dropout rates is positive and precisely estimated across all specifications. Time-varying controls result in more precise estimates, and generally larger point estimates. Point estimates are smaller in cases when wages implied by gross earnings are used to construct Kaitz-ratios. Overall, time-varying covariates have little explanatory power which is unsurprising in light of county fixed effects. The loss of 4 (12) observations in column 2 (3) stems from non-systemically missing information on time-varying economic conditions in some years for different counties.

Column 3 is the preferred specification, implying a 2.8 percentage point increase in the dropout rate for a one point difference in the Kaitz-ratio. Yet, there are no two counties whose Kaitz-ratios are one point apart; in fact, the maximum distance is 0.41 (0.40) when gross earnings (median income) is considered. These distances imply a causal effect of 1.14 (1.17) percentage points of minimum wage bite on dropout rates. This effect is roughly 20 percent of the average dropout rate in 2014.

95 percent confidence intervals are reported in parentheses and these suggest that more sizable effects cannot be ruled out. Moving from the 25^{th} to the 75^{th} percentile of the Kaitz-ratio's distribution results in a change of 0.11 (0.08); this implies an effect equal to approximately 5.2 (4.1) percent of 2014 dropout rates. The estimate in column 3 is statistically significant at one percent ($p = 0.0070$), while the estimate in column 6 is significant at five percent ($p = 0.018$).

One may be concerned that the linear specification presented above is unable to capture the nature of the causal effect of minimum wages on dropout rates. Figure 1.4 presents estimates that are based on grouping counties into quartiles based on their 2014 minimum wage bite. The first quartile corresponds to the 25 percent of counties with the *highest* average implied wage in 2014. These counties are also the base category, and all estimates ought to be interpreted relative to this group.

Figure 1.4 suggests that the effect size is proportional to counties' bite quartiles. The effect is driven in almost equal parts by counties in the third and fourth quartile; since the minimum wage has less bite in counties of the first and second quartile, the result is conform with economic intuition as a largely non-binding wage floor is unlikely to affect education decisions. The effects in fourth are of comparable magnitude to those in in third quartile. Due to small effective sample sizes within quartiles, the estimates are more imprecise than those in Table 1.2. Therefore, one cannot reject differences in estimates from the second through the fourth quartile. Counties in the top 25 percent of bite had about 0.5 percentage points, or 10 percent of the 2014 average, higher dropout rates than those in the bottom 25 percent due to the introduction of the minimum wage policy. Appendix Table 1.A.3 reports the all estimates on which Figure 1.4 is based.

This specification also serves to alleviate concerns that pre-trends in the outcome are not necessarily parallel across groups, as depicted in Figure 1.3. The fact that the effect of the third and fourth quartile is comparable implies that the average effect is not exclusively driven by differences between the extreme quartiles.

Appendix Table 1.A.4 further reports estimates based on a simple median split of counties based on their Kaitz-ratios. Counties above the median saw a roughly five percent increase in the dropout rates compared to those below the median. These estimates are precise and statistically significant at one or five percent.

In order to understand the effect of the minimum wage policy in greater detail it is instructive to estimate year-specific effects akin to an event-study design. Figure 1.5 does precisely that and suggests that the effect is strongest in the immediate aftermath of the minimum wage's introduction. All estimates ought to be interpreted

	Dependent variable: dropout rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2015	−.0043	−.0022	−.0069	−.0014	.0004	−.0045
	(.0071)	(.0072)	(.0068)	(.0069)	(.0069)	(.0067)
Post 2015 x Kaitz ratio	.0248	.0246	.0280	.0242	.0246	.0294
	(.0109)	(.0104)	(.0100)	(.0129)	(.0123)	(.0119)
Unemployment t-1		.0025	.0011		.0025	.0012
		(.0013)	(.0011)		(.0013)	(.0011)
Vacancies helper t-1		.0002	.0003		.0002	.0003
		(.0001)	(.0001)		(.0001)	(.0001)
Vacancy skilled t-1		−.00003	.00001		−.00004	−0.000000
		(.0001)	(.0001)		(.0001)	(.0002)
Asylum seekers t-1		−.00002	−.00002		−.00002	−.00002
		(.00003)	(.00003)		(.00003)	(.00003)
In-migration t-1		.0001	.0001		.0001	.0001
		(.0001)	(.0001)		(.0001)	(.0001)
Out-migration t-1		−.0001	−.00005		−.0001	−.0001
		(.0001)	(.0001)		(.0001)	(.0001)
Sector 1 t-1			.0010			.0013
			(.0022)			(.0022)
Sector 2 t-1			.00002			.00002
			(.0003)			(.0003)
Transfers t-1			.00001			.00001
			(.000005)			(.000005)
Social security t-1			.0028			.0027
			(.0011)			(.0011)
Gross local product t-1			−.0001			−.0001
			(.0001)			(.0001)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income	Median income	Median income
Fixed effects	County	County	County	County	County	County
State X year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,394	2,390	2,382	2,394	2,390	2,382

Table 1.2: Minimum wage bite and dropout rates. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. "Post 2015" is an indicator for years 2015, 2016, 2017 and 2018. "Post 2015 x Kaitz-ratio" is an interaction term of this indicator and the Kaitz-ratio at the county level. The Kaitz-ratio is constructed from gross earnings (columns 1 through 3) or median income (columns 4 through 6). "Unemployment" is the county unemployment rate (range 0 - 100). "Vacancies helper" is the share of county vacancies which are classified as being routine and simple. "Asylum seekers" is the number of asylum seekers per 1000 county inhabitants. "Vacancy skilled" is the share of vacancies which are classified as requiring special skills and abilities as well as profound training. "Sector 1" and "Sector 2" are the shares of employees in agriculture, forestry and fishing as well as manufacturing. Note that only employees in whose name social security contributions are made are counted (ger: *sozialversicherungspflichtig*). "Transfers" are the total payments made from the state to the county in €1000. "Social security" is the share of the population below 65 that is eligible for social security payments (ger: *bedürftig nach SGB-2*, colloquially called Hartz-IV). Finally, "Gross local product" is the value of all goods and services produced in a county divided by the total number of inhabitants [in €1000]. Each specification includes county and state-by-year fixed effects. Standard errors clustered at the county level are reported in parentheses.

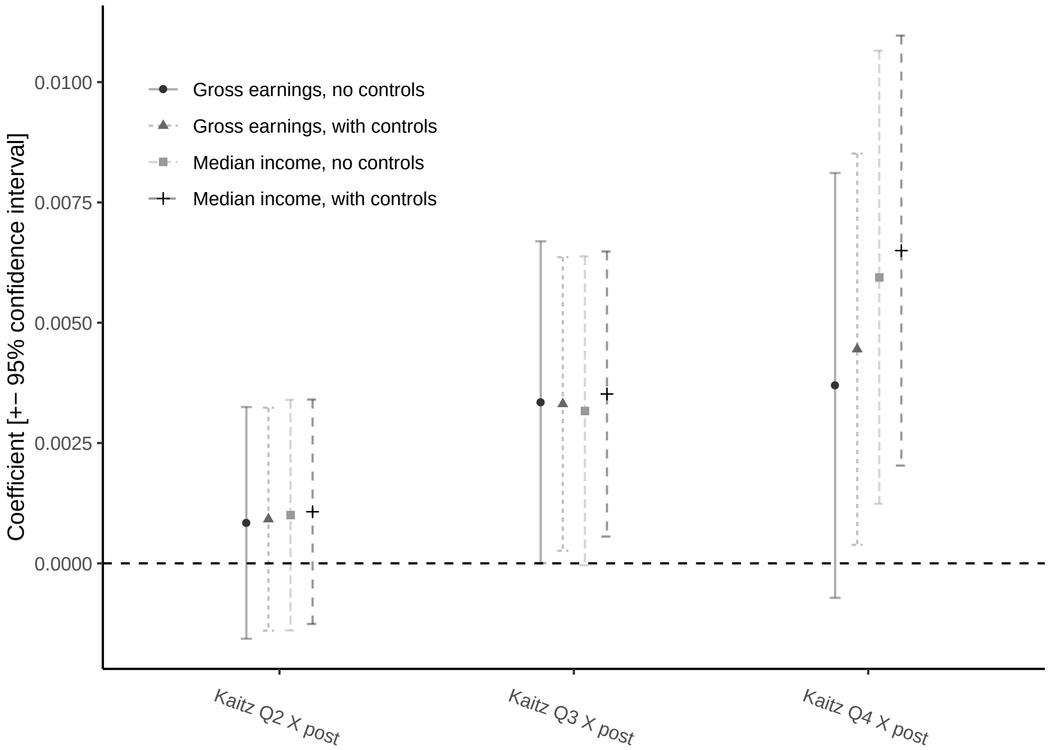


Figure 1.4: Quartile effects of minimum wage bite. *Notes.* This figure plots estimates from a specification similar to Equation (1.1) where counties are placed into quartiles of minimum wage bite. The figure thus plots the interaction term of the post-2015 period and the quartile indicators. The full estimation is reproduced in Table 1.A.3. The 25 percent of counties with the highest average implied wages in 2014 make up the first quartile and are also the base category.

relative to the year 2013.

There is a pronounced increase in 2015 and 2016—relative to 2013 levels—which is also statistically significant at the five percent level across all specifications. The point estimates for 2017 suggest that the effect persisted throughout that year, yet the widening of the confidence intervals foreshadows the effect disappearing. By 2018, the effect is essentially zero albeit with wide confidence intervals.

The minimum wage was raised to €8.84 (a raise of four percent) but this marginal increase is unlikely to have strong effects on education. The increase was much less publicized than the initial reform. The effect starts to disappear in 2017 which is consistent with decreasing salience of the minimum wage. In contrast, the cohorts that were to graduate in 2015 and 2016 were much more aware of the minimum wage. Moreover, the minimum wage did destroy around 200,000 marginal jobs (Caliendo, Fedorets, Preuss, Schröder, and Wittbrodt, 2018); this effect may have become visible and changed individuals' decision environment.

Figure 1.5 is also informative about anticipation effects. Recall that students cannot immediately start working when dropping out of Hauptschule, and that the minimum wage does not apply to apprentices. If individuals knew that a minimum wage would apply once they are eligible for regular employment, they could have adjusted their schooling decision ahead of time. The law was passed by parliament in July 2014 but was agreed upon in the coalition treaty for the 2013-2017 administration in November 2013. Therefore, students could have anticipated its inception and decreased schooling investment in 2014. Figure 1.5 disputes that notion as there is no discernible difference between 2014 and 2013. Thus, anticipation effects do not play a role in understanding the minimum wage's side effects.

1.5.2 Composition of effects

This section discusses the composition of the effect along two dimensions of interest. First, I consider how the reform affected graduation rates for the high school system's other tiers. Second, I show that the effects appear to be driven predominantly by male students.

High school tiers

Section 1.2 describes that Germany's high school system can be grouped into three tiers. If a student fails to complete at least the lowest tier, she is considered a dropout. This is a "zero-sum" calculation in the sense that if somebody drops out,

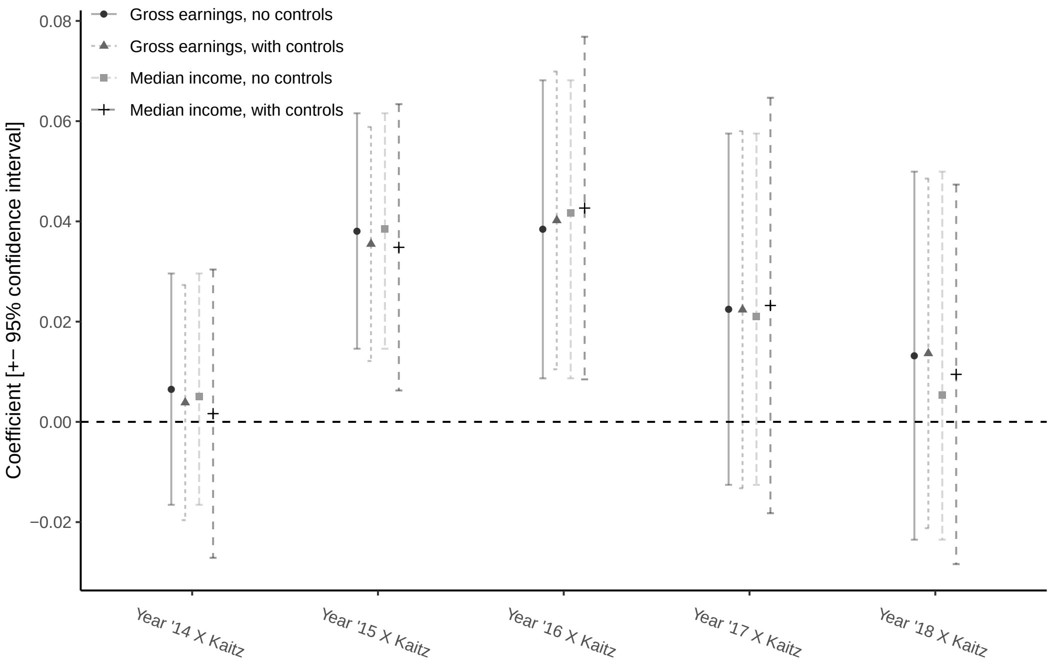


Figure 1.5: Year-by-year effects of the minimum wage on dropout rates. *Notes.* This figure plots estimates from a specification in which the post-2015 indicator in Equation (1.1) is replaced with with dummies for each year. These year dummies are then interacted with the continuous measure of minimum wage bite. 2013 is omitted and serves as the base category. Specification "with controls" (triangles and crosses) include the full set of covariates as in columns 3 and 6 of Table 1.2.

she cannot complete another tier in the same year. Therefore, a natural hypothesis would be that the increase in the dropout ratio coincides with a comparable drop in the graduation rates in Hauptschule, the lowest tier.

Table 1.3 investigates the effect on graduation rates for other school types. The estimates are broadly consistent with this notion and suggest that the increase in dropouts approximately corresponds to the decrease in graduation rates in the lowest high school tier. There appear to be further compositional changes in higher tiers of the school system but these effects are imprecisely estimated. This is unsurprising as one would not expect a policy introducing a wage floor to affect graduation rates for schools that typically channel into jobs paying significantly more than said wage floor.

	Dependent variable: completion rate					
	Hauptschule		Mittlere Reife		Abitur	
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2015	−.0216	−.0154	.0315	.0181	.0268	.0303
	(.0119)	(.0122)	(.0163)	(.0167)	(.0148)	(.0156)
Post 2015 x Kaitz ratio	−.0175	−.0337	−.0483	−.0322	.0235	.0216
	(.0159)	(.0191)	(.0237)	(.0299)	(.0186)	(.0243)
Wage implied by:	Gross earnings	Median income	Gross earnings	Median income	Gross earnings	Median income
Time-varying controls	Extended	Extended	Extended	Extended	Extended	Extended
Fixed effects	County	County	County	County	County	County
State X year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,382	2,382	2,382	2,382	2,370	2,370

Table 1.3: Minimum wage bite and completion rates by school types. *Notes.* This table reports estimates of the minimum wage introduction on *completion rates* by school type. Completion rates are calculated from the bottom-tier school type (*Hauptschule*, columns 1 and 2), middle-tier school type (*Realschule*, columns 3 and 4) and top-tier school degree (*Abitur*, columns 5 and 6). "Post 2015" is an indicator for years 2015, 2016 and 2017. "Post 2015 x Kaitz-ratio" is an interaction term of this indicator and the Kaitz-ratio at the county level. The specifications are analogous to Equation (1.1) and the "extended" set of controls refers to the full set of covariates as in columns 3 and 6 of Table 1.2. The Kaitz-ratio is constructed from gross earnings (columns 1,3 and 5) or median income (columns 2,4 and 6). Each specification includes county and state-by-year fixed effects. Standard errors clustered at the county level are reported in parentheses.

Differential effects by gender

Appendix Table 1.A.5 assesses whether there are any empirical differences in the effect of the minimum wage on dropout rates by gender. Regardless of the specification, the effect on male dropout rates is about twice as large as the effect on female dropout rates. Despite imprecisely estimated coefficients preventing a definite statement, the results suggest that the effect is predominantly driven by male students. There is a sizable effect for male students, about 1.2 times the average effect from Table 1.2. On the other hand, we cannot reject that there is no effect at all on women. Appendix Table 1.A.6 confirms this notion using the quartile-based measure of minimum wage bite.

This heterogeneity can be rationalized by evaluating the composition of the pool of individuals whose productivity is below €8.50, and who therefore have an incentive to drop out of school. There is reason to believe that there are more men than women in this pool; that is, relative to overall enrollment, more young men should be located at the lower end of the productivity distribution. The literature provides at least two reasons for why this is likely to be the case. First, young females tend to have higher non-cognitive ability (Becker, Hubbard, and Murphy, 2010; Fortin, Oreopoulos, and Phipps, 2015). Second, Bertrand and Pan (2013) state that young females tend to have higher educational aspirations than males of comparable age.

1.5.3 Regional variation

This section addresses two margins of effect heterogeneity. First, are there measurable differences for counties located in the former GDR? Second, do the effects differ between urban and rural counties?

West vs East Germany

Recall the discussion from the introduction about persistent socio-economic differences between West and East Germany; these former states tend to have higher wages and incomes even today. Those states that used to form the GDR also tend to have higher levels of unemployment, and Figure 1.1(a) shows higher levels of minimum wage bite in East Germany.

In order to assess whether there is heterogeneity in the way the minimum wage reform affected dropout rates between West and East Germany, I estimate Equation (1.1) and introduce an interaction effect for counties in East Germany. East Germany is geographically smaller, and is comprised of 76 counties. In this type of specification, the interaction term of the Kaitz-ratio with the post-2015 indicator measures the effect for counties in West Germany; the additional interaction with East Germany measures the *difference* in the effects between West and East Germany. Results from the linear-in-Kaitz-ratio specification are displayed in Table 1.4(a) and show that the point estimates for East Germany do not differ from those for West Germany. The triple interaction term for East Germany is estimated to be close to zero. The effect for West Germany matches the average effect from Table 1.2. This suggests that there is no systematically countervailing effect in East Germany.

Rural vs urban areas.

Policymakers have been concerned about a rural exodus and asymmetric development of rural and urban areas (Bauer, Rulff, and Tamminga, 2019). A larger effect on dropout rates in rural counties could contribute to a widening of the gap in economic performance between urban and rural counties. The INKAR database classifies counties along the urban-rural spectrum and there are 196 urban and 203 rural counties in Germany.

First, in the previous paragraph's spirit, I test for heterogeneous effects by interacting minimum wage bite in Equation (1.1) with an indicator for rural counties. Table 1.4(b) presents the results; there is no discernible *additional* effect of the minimum wage introduction on dropout rates in rural counties. The effect for urban counties is comparable to the effect from Table 1.2, while the interaction effect is close to zero and imprecisely estimated. Thus, the results indicate that the effect is at play in a similar fashion in rural and urban areas. Appendix Table 1.A.7 evaluates heterogeneous effect on gender-specific dropout rates. The patterns are consistent with results above; the effect is stronger for male students but, there is no *additional* effect on males in rural areas.

1.5.4 Robustness of results

The results presented in previous sections are robust to alternative specifications. The following paragraphs will discuss alternatives, as well as their relative merits compared to the benchmark model.

Alternative specification of time trends

The results in Table 1.2 include state by year fixed effects which serve to absorb state-specific shocks on a year to year basis. Time trends on the other hand more efficiently account for underlying trends. As Figure 1.3 suggest, dropout rates have generally started to climb in 2013. Not all states follow a comparable upward trajectory—see Appendix Figure 1.A.4—and this modeling choice allows trends to differ across states.

	Dependent variable: dropout rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2015	−.0046	−.0018	−.0065	−.0040	−.0015	−.0064
	(.0075)	(.0076)	(.0072)	(.0082)	(.0082)	(.0078)
Post 2015 X East Germany	.0139	.0088	.0149	.0198	.0186	.0245
	(.0217)	(.0222)	(.0214)	(.0166)	(.0166)	(.0156)
Post 2015 x Kaitz ratio	.0254	.0240	.0274	.0296	.0283	.0331
	(.0117)	(.0111)	(.0105)	(.0159)	(.0150)	(.0143)
Post 2015 x Kaitz ratio x East Germany	−.0060	.0060	.0064	−.0199	−.0136	−.0136
	(.0327)	(.0337)	(.0331)	(.0267)	(.0265)	(.0253)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income	Median income	Median income
Time-varying control:	None	Basic	Extended	None	Basic	Extended
Fixed effects	County	County	County	County	County	County
State X year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,394	2,390	2,382	2,394	2,390	2,382

(a) West vs East

	Dependent variable: dropout rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2015	−.0043	−.0012	−.0054	−.0029	−.0012	−.0055
	(.0071)	(.0084)	(.0080)	(.0087)	(.0087)	(.0084)
Post 2015 X rural		−.0046	−.0055	.0045	.0031	.0022
		(.0095)	(.0092)	(.0081)	(.0080)	(.0076)
Post 2015 x Kaitz ratio	.0248	.0228	.0251	.0272	.0283	.0316
	(.0109)	(.0139)	(.0135)	(.0180)	(.0174)	(.0170)
Post 2015 x Kaitz ratio x rural		.0076	.0095	−.0089	−.0067	−.0046
		(.0173)	(.0168)	(.0173)	(.0171)	(.0162)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income	Median income	Median income
Time-varying controls	None	Basic	Extended	None	Basic	Extended
Fixed effects	County	County	County	County	County	County
State X year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,394	2,390	2,382	2,394	2,390	2,382

(b) Rural vs urban

Table 1.4: Geographical variation in the minimum wage's effects. *Notes.* This table reports geographical variation in the effect of minimum wage bite on dropout rates. In panel (a), the effect is broken by West vs East Germany. Panel (b) analyzes the effect in rural vs urban counties. "Rural" is a dummy variable equal to 1 if the county is classified as rural, and 0 if it classified as urban. "x" denotes interactions of variables. The specifications are analogous to Table 1.2. Specifications in columns 1 and 4 do not contain any controls: specifications in columns 2,3,5 and 6 mimic the specifications in the same columns of Table 1.2. Refer to the table notes of Table 1.2 for definitions of the variables. Each specification includes county and state-by-year fixed effects. Standard errors clustered at the county level are reported in parentheses.

The results in Appendix Tables 1.B.1 and 1.B.2 present results when state-specific linear time trends are included. The results are slightly smaller in magnitude but equally precisely estimated. The estimates for the quartile-based treatment measure suggests an effect of 0.4 percentage points (preferred specification: 0.5).

Yet another way of absorbing variation over time is to include state-by-post-reform fixed effects; that is, fixed effects that control for differential conditions in each states after 2015 as compared to before. This specification would control for a situation in which states chose to systematically pursue different education policy after the minimum wage introduction. If they did not adapt this policy on a year-to-year basis, this specification is more powerful than the more flexible state-by-year fixed effect. Results are presented in Appendix Tables 1.B.3 and 1.B.4 and are larger in magnitude than the effects in Table 1.2.

Using different years for covariates

The school year for those who are supposed to graduate finishes between late April and June. Thus, contemporaneous levels of time-varying controls also constitute a measure of prevailing economic conditions at the time when students make their schooling decisions. Alternatively, one may argue that the decision environment students observe is the product of past economic conditions. The second lag of time-varying covariates may be better able to capture this than the first lag. When using contemporaneous covariates, I cannot use the year 2018 as most time-varying control variables are not observed through 2018.

Appendix Tables 1.B.5 and 1.B.6 present results with contemporaneous covariates. Effects are stronger in magnitude which is unsurprising as Figure 1.5 indicates that the effect starts to disappear in 2018, which is the year not included when using contemporaneous covariates. Further, Appendix Tables 1.B.7 and 1.B.8 condition on the second lag of time-varying covariates. The magnitude becomes visibly smaller and the estimates are less precise. Signs still point in the same direction and the effects along the quartiles are as expected.

Alternative way of clustering

Recall that this study utilizes data on all counties in Germany. To the extent that Germany is the population of interest, it is unclear what underlying sampling mechanism may have resulted in a dataset with German counties and realized dropout rates. Thus, rather than associating the estimates above with sampling uncertainty, one could view this analysis as a finite population analysis. Manski and Pepper (2018) describe this issue in more detail when they discuss their choice to not conduct inference in a cross-state study in the US. Abadie, Athey, Imbens, and Wooldridge (2020) provide an alternative framework that takes into account *design-based uncertainty*. The parameter of interest— the causal effect of the minimum wage introduction—is defined in terms of unobserved outcomes; dropout rates under alternative 2014 wage levels. Using the insights in Abadie, Athey, Imbens, and Wooldridge (2020), the correct way to conduct inference in this application is cluster standard errors at the county level as *treatment* is assigned at the county level. All results presented thus far follow this suggestion.

However, an alternative strand of literature argues that the choice of clustering unit should be guided by beliefs about intra-cluster correlations of covariates and errors. Cameron and Miller (2015) recommend being conservative and use more aggregate clusters, and state that ignoring within-state cross-county correlation of the regressors and the errors could result in biased standard errors. It is conceivable that both, dropout rates and economic conditions, correlate within states, and that this correlations spans years.

Therefore, following the advise in Cameron and Miller (2015) I provide estimates where inference is based on alternative ways of clustering. I consider four additional cases: first, one-way clustering using state-by-year cells. Second, two-way clustering across states and years. The justification for the latter would be an argument that economic conditions and education outcomes correlate within any given year across all counties and, at the same time, across time within any given state. Further, I provide estimates based on two-way clustering at the county and year level. Finally, the most conservative specification clusters at the state level. Clustering across states (eleven) and years (six) profoundly reduces the number of clusters which in turn may lead to inconsistent estimation of standard errors.

Appendix Figure 1.B.1 presents the results from those four alternative cluster specifications. Unsurprisingly, standard errors are larger when clustering at more aggregate levels. Results based on gross income remain statistically significant at five percent across all specifications. In the most conservative specification – one-way clustering at the state level – the 95 percent confidence interval for the point estimate based on gross income (median earnings) is $[0.0065, 0.050]$ ($[-0.0001, 0.060]$). These results corroborate the findings of an economically meaningful effect of the minimum wage introduction on dropout rates.

1.6 Randomization inference

Exposition

Randomization inference, sometimes referred to as permutation inference, is a tool to simulate the distribution of a test statistic under a designated null hypothesis (Fisher, 1935). It has recently received increasing attention in Economics, especially in the evaluation of randomized experiments (Heß, 2017). Young (2019) re-analyzes a number of experimental studies and provides a perspective on seemingly statistically significant results through the lens of randomization inference. The procedure is also applicable to non-experimental settings and allows the researcher to calculate an exact p-value of a *sharp* null hypothesis.

For the ease of exposition, suppose treatment is binary and a county is treated if its Kaitz-ratio is higher than the median Kaitz-ratio in 2014. Appendix Table 1.A.4 shows that the effect of being above the median is about 0.3 percentage points—6 percent of 2014 averages. Randomization inference tests a sharp null hypothesis; for instance, *no individual* treatment effect. If, under the null hypothesis no individual effect is assumed, one can impute the missing outcome for each unit. That is, under H_0, $y_i(d_i = 0) = y_i(d_i = 1)$, where y_i is the outcome of county i under treatment $d_i \in 0, 1$. By means of this equality, one could take any two counties, and "flip" their treatment assignment which would have no effect on observed outcomes. The logic generalizes to a continuous measure of treatment.

Thus, randomization inference starts with generating a large number of possible permutations of the treatment vector and calculating the respectively implied effects. The distribution of resulting treatment effects is referred to as the randomization distribution.[14] It approximates the universe of possible effect magnitudes that could have resulted by chance despite there being a zero individual effect. Finally, by comparing the randomization distribution to the observed treatment effect, one can calculate an exact p-value for a two-sided hypothesis using the ratio of the number of more *extreme* test statistics to the number of permutations.

[14] In this application, I am calculating the treatment effect—that is, the point estimate in a model such as Equation (1.1)—across all treatment permutations. The procedure also lends itself to a different choice of test statistics, such as the t-statistic for instance or the inter-quartile range.

Results

In order to conduct randomization inference in this context, one needs to generate permutations of counties and their minimum wage bite in 2014. Note that, all observations of a county take on the same treatment value. Thus, in an experimental setting this would correspond to a clustered design in which counties are randomly drawn, and all observations in a cluster are assigned the same treatment value.

The following results are based on the following procedure. I create 999 bootstrap samples (i.e., drawn *with* replacement) of counties in 2014 and their treatment value. Treatment is either given by an above-median indicator, or the continuous value of the Kaitz-ratio. In each of these 999 samples, I obtain an estimate $\hat{\beta}_2$ using Equation (1.1). Finally, I plot a histogram of the implied randomization distribution and mark the value of the actually observed treatment effect.

Appendix Figure 1.A.6 plots the randomization distributions of four "treatment effects". First, I consider the benchmark models in which treatment is a county's Kaitz-ratio in 2014 as implied by either, gross income or median earnings. Second, treatment is based on a median split on those quantities. The vertical, dashed lines indicate the actually observed effect in each of those four scenarios.

The randomization distributions of the effect of the minimum wage introduction on dropout rates are such that I can reject the null hypothesis of no individual effect in all four cases. Read clockwise, the p-values of the specifications in Figure 1.A.6 are 0 (top left), 0 (top right), 0.002 (bottom left) and 0.001 (bottom right). I interpret this as corroborating evidence that the minimum wage reform did in fact have an effect on high school dropout rates.

1.7 Conclusion

This paper presents evidence that the German federal minimum wage of 2015 causally increased dropout rates among high school students. The minimum wage reform resulted in 0.085 percentage points, or roughly 5 percent of 2014 levels, more dropouts in the county at the 75^{th} percentile compared to the county at 25^{th} percentile of minimum wage bite. I decompose the effect along margins of interest to policymakers and show that the effect is stronger for male than for female students. A number of robustness checks corroborate this result, and permutation tests lend further credibility to the causal link.

This result may be the consequence of individuals with low productivity reducing their investment into education. Specifically, if individuals expect that further investment is unlikely to raise their productivity above the level of the minimum wage, abstaining from further education may make sense. The fact that the minimum wage does not immediately cover minors who drop out of school should "stack the deck" against finding such an effect. Thus, the estimates can be interpreted as a conservative lower bound for the relationship between minimum wages and education.

The present paper extends the literature on minimum wage effects in three directions. First, it adds to the literature on human capital effects of minimum wage legislation by using a measure of schooling that is free of labor-force participation concerns. Second, it extends existing evidence beyond the Anglo-Saxon environment. Third, the identification structure improves on existing work by exploiting finely grained variation in minimum wage within geographical entities, states, where individuals are subject to the same education policy.

Nonetheless, my findings are not free of limitations. While existing work on this subject has noted that the Kaitz-ratio correlates well with other measures of minimum wage bite, its appropriateness to measure minimum wage bite rests on specific distributional assumptions.

Two theoretical mechanisms may be at play in generating a link between minimum wages and education; the opportunity cost of education is affected, but also jobs at the lower end of the productivity distribution are likely to be destroyed. My study identifies the composite of these two effects. A structural treatment of this link may be helpful in separately quantifying these mechanisms. This constitutes a fruitful avenue for further research.

This research is potentially important for policymakers. Minimum wages are an increasingly popular tool of labor policy, and increasingly so in developing countries. Introducing or profoundly increasing a minimum wage may result in individuals decreasing their human capital investment. Policymakers ought to be aware of this possibly unanticipated side effect. Individuals who do alter their human capital investment in response to minimum wages risk falling behind in a labor market that is increasingly putting a premium on non-routine analytical tasks (Autor and Handel, 2013).

2. Charitable Donations to Natural Disasters Evidence from an Online Platform[1]

Abstract

We investigate how demand for and supply of charitable donations on a large online platform respond to natural disasters. We find that donations are concentrated in a small fraction of the 1,720 disasters in our 5-year period of observation. These tend to be severe disasters that receive media coverage. Event study estimates indicate that the pattern of disaster donations is consistent with donor fatigue over time and across disasters, but is inconsistent with crowding out of other charitable causes. Instrumental variable estimates suggest that charities could raise additional donations on the platform for disaster relief, had they fundraised for them.

[1] This chapter is based on joint work with Rajshri Jayaraman (University of Toronto, Canada, and European School of Management and Technology (ESMT) Berlin) and Marrit Teirlinck (Humboldt University and ESMT Berlin).

2.1 Introduction

Major natural disasters tend to attract generous charitable donations, while a large number of smaller-scale disasters go effectively unfunded. For example Typhoon Haiyan, which struck the Philippines in November, 2013 killing over 7,000 people, received extensive media coverage and raised $33 million from U.S. donors within the space of two weeks, on top of the $37 million pledged by the White House (Hicken, 2013; The White House, 2013). Incredibly, Haiyan was already the twenty-third major storm or typhoon to hit the Philippines that year.[2] None was as deadly as Haiyan but put together, storms that battered the Philippines and other South East Asian countries in 2013 wrought considerable devastation. So too did floods that hit East African countries that year. Yet, these smaller natural disaster events barely registered on private donors' radars.[3]

This has real consequences for the alleviation of human suffering in the aftermath of natural disasters. Charities, who are major players in relief activities for disasters of all shapes and sizes, get 90% of their funding for these activities through charitable donations from private individuals (Stirk, 2014). Understanding which disasters are successful at raising donations and why some disasters may be unsuccessful in doing so is important in this context. When a natural disaster fails to attract donations, "donor fatigue" is often blamed. This refers to the casual observation that after a major disaster, which initially elicits generous charitable contributions, donations tend to dry up. Donors "tire" of giving.

In this paper, we ask three questions. First, which natural disasters tend to receive charitable donations? Second, is the pattern of donations consistent with donor fatigue? Third, how much funding could charities have raised for disasters, had they solicited donations? We begin with a descriptive exercise, which asks which types of natural disasters are successful in attracting donations, by examining the correlates of disaster relief fundraising and donations.

Then, using an event study design, we investigate whether the pattern of do-

[2] Storms and typhoons in the country are named alphabetically and Haiyan's official name was "Yolanda". The Guardian Newspaper quipped, "In 2013, the Philippines had so many typhoons that it ran out of letters to name them." (The Guardian, 2015).

[3] In South East Asia, our data for 2013 reports 1,177 casualties from natural disasters, excluding Typhoon Haiyan. Cumulatively, these events are recorded to have affected nearly 900,000 individuals in that year. In East Africa, floods alone are reported to have affected nearly 1,400,000 individuals, and killed 323. In India and China, natural disasters (not including the June 2013 North Indian floods, which received widespread media coverage) inflicted 2,460 casualties and affected 43,753,779 individuals.

nations is consistent with donor fatigue. We look at three dimensions commonly alluded to in the popular press as culprits for why disasters fail to attract donations. The first is what we call *temporal* fatigue: after an initial surge in generosity following disaster onset, donations to the disaster tend to dry up. For example, following Typhoon Haiyan Time Magazine reported, “While the international community came together for an unprecedented mobilization of relief, [temporal] donor fatigue has now set in” (Time Magazine, 2014).

The second manifestation of donor fatigue is *crowding out*: a surge in donations for disaster relief is thought to crowd out donations to other charitable projects because people have limits to how much they want to, or can, give. For example, The New York Times reported that “After Hurricane Katrina, Congress was so concerned that donations for hurricane relief efforts would cut into other charitable giving that it passed one of the biggest temporary tax breaks in history” (New York Times, 2006). Anticipation of crowding out, which would compromise their ability to successfully fundraise for other, non-disaster-related, projects, may make charities reluctant to fundraise for disaster relief.

The third category of donor fatigue is commonly called *disaster* fatigue: after a major disaster that received generous donations, subsequent disasters that arrive in close proximity to the major disaster may fail to attract donations. For example, in 2017, this variant of donor fatigue was blamed for the fact that The American Red Cross raised less than one tenth of what it raised after Hurricane Harvey for Hurricane Maria, which hit Puerto Rico only one month later (Penta, 2017).

In the final step of the analysis we ask how much more charities could have raised for disaster relief by more actively engaging in fundraising, while conservatively accounting for donor fatigue. In this step, we use an instrumental variables (IV) approach to estimate the causal effect of fundraising on disaster relief.

The setting we explore is Betterplace.org, which is by far the largest platform for online fundraising and giving in Germany.[4] It is a rich and active digital ecosystem wherein charities *demand* donations by posting fundraising pages for a wide range of charitable projects, and individuals *supply* donations to these projects. During our observation period, from 2013-2017, the platform hosted fundraising pages for nearly 20,000 projects, posted by 13,000 charities of all sizes, and to which 600,000 individuals made almost 700,000 donations totaling €35 million. We match these

[4] In Germany, there are no real competitors to Betterplace. Comparable platforms that collect funds for charities are for example GlobalGiving.org and charity.gofundme.com, both of which are U.S.-based.

real-time administrative data to a census of 1,720 natural disaster events. This allows us to assign each time-stamped project fundraising page and donation to a unique natural disaster or a non-disaster-related cause.

There are three main advantages to this setting. First, it allows us to investigate both the demand and supply of donations to natural disasters. For each of the 1,720 natural disasters that occurred over our observation period, we are able to track whether a fundraising page for a disaster relief project was posted by a charity on Betterplace; this captures donation demand for disaster relief, and in what follows we often refer to donation demand as “fundraising” or “project entry”. We are also able to track individual donations to each project. This measures donation supply, which we often refer to as “donations” or “disaster relief”. Supplementing these administrative and disaster census data with additional disaster characteristics, including media coverage, enables us to examine the correlates of disaster relief donations. Second, we know whether charitable projects and the donations they attract are devoted to disaster relief or non-disaster-related charitable causes. This allows us to examine crowding out. Finally, time stamps combined with high-frequency data permit the use of an event study design that exploits the plausibly exogenous onset of disaster events to investigate causal donation responses. This allows us to explore whether time series patterns in donations are consistent with donor fatigue.

The questions of whether, and which, disasters attract donations, and whether disaster relief funding may be subject to donor fatigue are important. Natural disasters, which affect tens of millions each year, are chronically underfunded. In 2019, for example, the UN Office for the Coordination of Humanitarian Affairs (UNOCHA) reported a 44% funding shortfall—the gap between estimated need and provision of funding—in humanitarian disaster relief efforts (United Nations, 2018, p.8). Charitable donations from private individuals play an important role in filling this gap. Of the $27.3 billion in international humanitarian assistance in 2016, about a quarter ($6.5 billion) came from private donors, the majority of whom are private individuals. Moreover, private donations are growing at a faster rate than those from government sources. Whereas humanitarian assistance from government sources saw only a modest increase of 1.5% from 2016 to 2017, contributions from private donors increased by an impressive 8.3% (Development Initiatives, 2018). Online giving is especially important in this context, since it is growing at a much faster rate than overall giving. In the U.S., for example, overall giving grew by 9% between 2016 and 2018 whereas online giving grew at almost twice this rate (Backbaud Institute, 2018).

Against this backdrop, online platforms such as Betterplace present a real opportunity to raise money for disaster relief from private individuals. On the donation demand side, they provide a relatively inexpensive way for charities to fundraise for their disaster-relief activities in a timely manner. The monetary cost of posting a fundraising page for a charitable project on most online donation platforms is negligible, and the marginal cost of soliciting donations once a page has been posted is effectively zero. This marks a striking departure from charities' traditionally high marginal costs of solicitation, including direct mailing, door-to-door fundraising, and telemarketing (Andreoni and Payne, 2011; Name-Correa and Yildirim, 2013). On the donation supply side, online platforms provide a relatively easy way for would-be philanthropists to give to charities' disaster relief projects. Giving remotely at the time of one's choosing from a computer or mobile device is likely to impose lower transaction, psychic, and social costs compared to traditional modes of giving, such as over the telephone, by mail or in person. Thus, at least in principle, online donation platforms provide an opportunity to channel private donations to disaster relief by lowering the barriers to both fundraising and giving.

We find that online giving fails to cater to the bulk of natural disasters. Heterogeneity analysis, which follows the seminal contributions of Strömberg (2007) and Eisensee and Strömberg (2007), indicates that disasters which see traction on the platform tend to be the usual suspects, namely severe disasters which receive media coverage. Concretely, 80% of disaster relief donations on the platform go to natural disasters that account for only 20% of casualties. This is eerily reminiscent of the so-called 80/20 Pareto rule that characterizes traditional product markets, where the bulk of sales revenue (≈80%) is generated by a small proportion of products (≈20%).[5] It is surprising in view of the role online platforms such as eBay and Amazon have played in catering to the remaining 80% of niche products—akin to the smaller scale disasters in our context. The contrast between product market platforms and this donation platform could not be starker: overall, we observe demand and supply for only 4% of disasters on Betterplace. Interestingly, this outcome is characterized by a paucity of demand. Although disaster relief projects that are posted to the platform tend to be successful in raising donations, very few charities fundraise for disaster relief on the platform.

[5] This is sometimes referred to as the "long tail" phenomenon—a term popularized in the business context by Anderson (2006), with reference to the ability of digital markets to cater to niche product offerings in the "long tail of the product distribution". See also Brynjolfsson, Hu, and Smith (2006) and Brynjolfsson, Hu, and Simester (2011).

Donation patterns are consistent with temporal and disaster fatigue, but not with crowding out. More specifically, fundraising and donations kick into gear within a week of disaster onset. The speed of the uptick underscores the low barriers to fundraising and donating online. By the same token, both dissipate by the third week after disaster onset. This pattern is consistent with temporal fatigue. Furthermore, fundraising and donations are effectively absent for disasters that occur within a two-month window of large disasters which have attracted massive funding; by contrast, events that transpire outside this two-month window do receive donations. This is consistent with disaster fatigue.

At the same time, disaster relief fundraising and donations do not appear to crowd out other causes. The evidence suggests that any increase in the former comes from successful *additional* fundraising rather than a substitution away from other charitable causes.

In a final step of our analysis we provide a causally interpretable estimate for how much more money could be raised on the platform if providers were to solicit funding for more disasters. Endogeneity is an obvious concern when estimating how the demand for donations affects the supply of donations.

We address this by using an IV strategy, instrumenting for demand using historical charity presence in a country affected by the disaster. The rationale is that historical charity presence plausibly reduces costs of initiating local disaster relief operations, thus making it more likely for charities to engage in fundraising for these activities (i.e. demand donations).

Our estimates suggest that, had they posted disaster-relief projects on the platform, charities could have raised roughly €1,000 in additional funding for marginal events. Although this sum is unlikely to make or break a disaster relief project, the amount itself is non-trivial. The marginal cost of posting a new fundraising page on Betterplace for the thousands of charities who are already active there is close to zero. Yet, the amount they would raise for a project devoted to a marginal disaster is almost two times the median amount (of €600) historically raised for disaster relief projects posted on the platform. This indicates the presence of fundraising potential on this platform for disaster relief, which remains untapped.

2.2 Related literature

This paper lies at the intersection of two distinct strands of literature: one on disaster relief funding, and the other on charitable donations. We make two contributions to these literatures. The first is to understand how an online platform for charitable donations is, and could be, used to fundraise for disaster relief. The second is to provide systematic evidence regarding donor fatigue.

The literature on disaster relief funding has primarily focused on either overseas development assistance (ODA) (Strömberg, 2007; Eisensee and Strömberg, 2007; Fink and Redaelli, 2011; Becerra, Cavallo, and Noy, 2010; David, 2011), or remittances (Mohapatra, Joseph, and Ratha, 2009; Yang, 2008). We add to this literature by examining private donation responses to disaster events in the context of online donation platforms, which have become an increasingly important vehicle for donating to charitable organizations.

More generally in terms of the charitable donations literature, millions of dollars are spent every year on fundraising; see List (2011) and Andreoni and Payne (2013) for excellent reviews. Andreoni and Payne (2011), for example, report that in the U.S. the average fundraising-to-donation ratio is 12%, with the average charity spending $100,000 per year on fundraising. Online fundraising has driven these costs down dramatically. There are two main channels through which this opportunity has been operationalized by charities. The first is through fundraising on their own websites. Perhaps the most famous example of this is Wikimedia, which successfully raises millions of dollars within the matter of weeks through fundraising banners on Wikipedia.[6] The second online fundraising channel is the use of advertisements to solicit donations. For example, in a thoughtfully designed recent field experiment, Adena and Hager (2020) find that video advertisements on Facebook were a cost-effective way for the NGO *Save the Children* to fundraise.

Charities may not be able to avail themselves to either of these two online fundraising channels for at least three reasons. First, successful Wikipedia-style fundraising requires sufficient website visitor traffic and compelling web design, neither of which many (especially smaller) charities have. Second, online advertising requires deeper pockets than most charities possess. Third, many charities have difficulty overcoming bureaucratic hurdles for collecting donations online, such as providing electronic payment options or tax receipts. These barriers can be overcome, but this requires the kind of time and money that are in scarce supply, particularly in the aftermath of a disaster.

[6] See, for example, `https://meta.wikimedia.org/wiki/Fundraising/2018-19_Report`. Chen, Li, and MacKie-Mason (2005) provide a presciently early field experiment designed to evaluate the efficacy of different online fundraising mechanisms on giving on the website of the Internet Public Library.

We contribute to the fundraising literature by exploring whether charities are taking advantage of a third online fundraising channel—online donation platforms—to overcome these barriers. In particular, online platforms exhibit high web traffic; sleek design; a large number of online payment options; customizable "plug and play" fundraising page templates that can be up and running in a matter of hours; and charities can post projects on them at low marginal cost. In fact, Betterplace offers this service, complete with charity customer support for setting up fundraising pages, pro bono to almost all charities. While a number of studies have explored the efficacy of particular fundraising strategies on online donation platforms more generally (for example, Bøg, Harmgart, Huck, and Jeffers (2012) and Payne, Scharf, Smith, et al. (2014) study the U.K.-based platform, justgiving.com; Altmann, Falk, Heidhues, Jayaraman, and Teirlinck (2019) study Betterplace), we are aware of none that has explored their broader role in catering to a wide range of disaster events and charities.

In focussing on not just the supply but also the demand for donations (i.e. fundraising), this paper also fills a gap in both the charitable donations and disaster relief literatures. With respect to the former, List (2011) notes a paucity of studies investigating the entry of charitable organizations in the market for charitable giving. We shed some light on this by examining charities' entry on this online platform in the context of disaster relief. As for the latter, the literature on ODA and remittances typically ignores the demand side of the funding equation, under the entirely reasonable assumption that disaster relief funding will (or should) be absorbed locally. This assumption is arguably less reasonable in our context where supply is contingent on demand: donors can only give to a disaster event if a charity posts a fundraising page soliciting funding for that event. Indeed, equilibrium outcomes in this online platform seem to be characterized by relatively generous donation responses conditional on charities asking for disaster relief funds, but very little fundraising activity for disaster relief in the first place. This suggests that anemic *demand* is part of the explanation for why so many disasters go unfunded. This is, as far as we know, a novel insight in both the literature on disaster relief funding and the literature on charitable giving.

Our analysis examines donor fatigue along three dimensions and in so doing, brings together three separate strands of the charitable donations literature. The first strand pertains to temporal fatigue. Brown and Minty (2008) use data from the websites of six major charities to examine the effect of media coverage on charitable giving after the 2004 tsunami in the Indian Ocean. Although peripheral to their analysis, they show that donations taper off over time. Scharf, Smith, and Wilhelm (2017) document a similar pattern in the context of fundraising appeals for six major disasters. We confirm the findings of these studies with respect to donation supply for a substantially larger set of disasters and charities, and supplement them by examining how demand-side fundraising efforts respond to disaster onset.

The second dimension of our donor fatigue analysis speaks to a large literature in charitable donations on crowding out. A central question in this literature has been whether successful fundraising efforts on the part of charitable organizations raise new donations, or simply divert funds away from other causes. Although this issue was conceptualized decades ago by Rose-Ackerman (1982), systematic empirical evidence is scant (Andreoni and Payne, 2013).[7] This is not altogether surprising. Individuals and charities are likely to decide on portfolios of giving or fundraising for different causes, and this makes it challenging to identify any potential crowding out effect.

Natural disasters have proved handy in identifying potential crowding out, since they are plausibly random events, which generate exogenous variation in both fundraising and giving. A handful of studies have exploited this feature of natural disasters to investigate whether they crowd out donations to other causes. Brown, Harris, and Taylor (2012) and Scharf, Smith, and Wilhelm (2017) use panel data to explore how individual giving patterns to different causes change in the aftermath of (fundraising efforts for) disaster events.[8] Petrova, Perez-Truglia, Simonov, and Yildirim (2020) investigate whether charitable donations to American Red Cross disaster relief efforts crowd out political giving and vice versa.

Methodologically, our donor fatigue analysis is closest to the event-study strategy employed in Petrova, Perez-Truglia, Simonov, and Yildirim (2020). Empirically, we add to the crowding out literature by examining both the demand and supply of donations, and broadening the scope of the analysis to the universe of non-disaster-related fundraising efforts and donations on Betterplace. Conceptually, our paper is distinct from this literature given our interest in understanding whether, consistent with crowding out, successful disaster-related fundraising is countered by less successful fundraising for other causes.

[7] There is, by contrast, a large literature on whether government grants crowd out private donations. See, for example, the contributions of Andreoni and co-authors in Andreoni and Payne (2003), Andreoni and Payne (2011), and Andreoni, Payne, and Smith (2014).

[8] A number of studies have examined individual correlates of disaster-related giving using survey data; see the recent contribution of Devlin and Rowlands (2019) and the references cited therein.

Third, there seems to be widespread consensus in both the popular press and among humanitarian relief organizations that disaster fatigue is commonplace. This claim tends to rest on anecdotal evidence that when a major disaster has attracted generous donations, the next one gets precious little. We provide what the best of our knowledge is the first systematic evidence exploring this phenomenon.

2.3 Data

We use two main data sources covering the five-year period from 2013-2017. The first, EM-DAT, is an inventory of natural disasters and their characteristics. The second, Betterplace, furnishes administrative records on charitable projects as well as real-time individual records on the donations towards each of these projects. Using project descriptions, project locations, as well as donation time stamps, each donation can be broadly classified as going towards a natural disaster or "other", non-disaster-related, causes. We combine these data to create an event-level dataset, which associates each natural disaster with its corresponding Betterplace fundraising activity and donations.

To this dataset, we match a number of disaster- and country-level characteristics, including media coverage, socio-economic environment and bilateral distance measures. These covariates, collected from multiple sources detailed at the end of this section, allow us to investigate sources of heterogeneity in disaster relief donation responses.

2.3.1 Natural disasters

Data on natural disasters is obtained from the Emergency Disaster Database (EM-DAT) maintained by the Centre for Research on the Epidemiology of Disasters (CRED) at the Catholic University of Leuven (Guha-Sapir, 2018). Disaster events in this database fulfill one or more of the following requirements: ten or more people are reported killed; 100 or more people are reported affected, injured, and/or homeless; a state of emergency has been declared; or there has been a call for international assistance. A separate entry is made for each country affected by a disaster (e.g., if a typhoon affects three countries, three entries are made recording the same event).

The three deadliest disasters during our observation period were the 2015 Nepal earthquake, Typhoon Haiyan in the Philippines (2013), and the Indian Flood of 2013. Each killed over 6,000 people and affected over half a million (see Appendix Table 2.A.1). But they are just the tip of the iceberg. In total, the sample comprises

1,720 disaster country-events (referred to as "events" hereafter). These stem from 1,439 unique disasters across 173 countries.

Table 2.1 provides summary statistics for our final cross-section of 1,720 disasters.[9] Over the five-year period of observation, these events reportedly killed over eighty-six thousand and affected nearly one billion people. While the average disaster across all years killed 50 people, it affected over half a million. The severity of disasters varies considerably. Those at the 25^{th} percentile inflict no casualties and affect around a hundred people, while the most severe disasters kill thousands and affect hundreds of millions (see Appendix Figure 2.A.1). We use the natural logarithms to account for right skewed distributions (Smith, Ottoni-Wilhelm, and Scharf, 2017).

	# Casualties (1)	# Affected (2)
Total	86,333	967,185,633
Mean	50	561,990
Median	6	3,600
p25	0	105
p75	21	42,500
Max	8,969	330,000,000

Table 2.1: Natural disaster severity: Casualties and number affected. *Notes.* This table shows the number of casualties from and the number of people affected by natural disasters in the years 2013-2017. The table reports the total; mean; median; the 25^{th} and 75^{th} percentiles; and the maximum.

2.3.2 Donations: demand & supply

Data on the demand for and supply of charitable donations are obtained from Betterplace, Germany's largest online platform for charitable giving. Betterplace is a fundraising platform on which charities post fundraising pages for charitable projects to which individuals can donate. The only formal requirement for charities to post project pages is that they are registered in Germany. Project pages themselves are eclectic in terms of geography, funding request, and cause, including natural disaster relief. Betterplace accounts for under 1% of overall private donations in Germany, but in proportion to domestic giving it is more than 12 times as large as

[9] Although EM-DAT also records an economic estimate of damage caused, we omit it from the analysis as it suffers from severe measurement error and is plagued by missing values. This is exemplified in the zeros in Appendix Table 2.A.1.

Global Giving, the world's largest online donations platform, based in the U.S.[10] As online giving has grown in popularity, so too has Betterplace, with total donations increasing by 43% from 2013 to 2017. Betterplace also plays an outsized role in disaster relief donations. For example, in the five weeks after Typhoon Haiyan, Germans donated €13.7 to relief (SIR/dpa, 2013); donations on Betterplace totaled €520,000 (4%). In sum, Betterplace provides us with a rich and diverse private donations ecosystem, which allows us to study platform-based charitable giving.

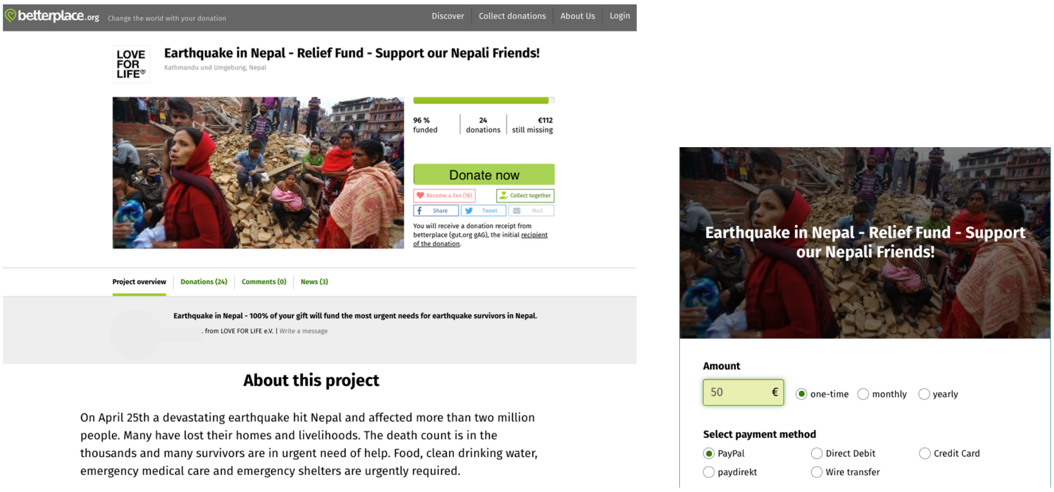
Although not directly material to our analysis, it is worth noting that donors on Betterplace are not representative of the overall donor population. In the interest of data privacy, Betterplace does not track users' personal data. However, a comparison of Betterplace's (aggregated) Google analytics data to German Socio-Economic Panel (G-SOEP) data indicates that donors on Betterplace tend to be disproportionately young and female—a pattern that is consistent with survey evidence on the demographic profile of online donors in the U.K. (Just Giving, 2012). As such, the donation patterns studied in this paper are more likely to pertain to digital natives rather than the donor population at large.[11]

Crucially, administrative data from Betterplace capture both the demand for donations (fundraising pages for projects posted by charities), and the supply of donations (from individual donors). Both are illustrated in Figure 2.1. Panel (a) presents a typical fundraising page posted on Betterplace, in this case, by a charity engaged in disaster relief operations following the 2015 Nepal earthquake. It exemplifies our data source for donation demand. Administrative data records each project description ("About this project") as well as when the project was posted; by which charity; where it is located; and how much funding was requested. Visitors who wish to donate to this project click on the "Donate now" button to be redirected to the donation page depicted in Panel (b), which exemplifies our data source for donation supply. Here again, Betterplace records how much was donated and to which project in real time.

Roughly 80% of visitors to Betterplace land on a fundraising page like the one

[10] In 2017, Global Giving raised $69.5 million, a year in which Americans gave $390 billion to charity. In the same year, Betterplace.org raised €11.1 million, with Germans giving a total of €5.2 billion to charity.

[11] There is, to the best of our knowledge, no representative survey of charitable giving in Germany. We use Betterplace's Google analytics web tracking data from June-October 2018 (which covers 51% of Betterplace's donors) with a question on donations administered to a representative sample of Germans in the 2014-15 wave of the German Socio-Economic Panel (G-SOEP) data (see Appendix Figure 2.A.2).



(a) Project Page (b) Donation Page

Figure 2.1: Project fundraising & donation pages on Betterplace. *Notes.* Panel (a) shows a project fundraising page posted by the charity "Love for Life", to raise money for a disaster relief project in Katmandu, Nepal, and neighboring regions, in the wake of the 2015 Nepal earthquake. Clicking on the "Donate Now" button redirects project page visitors to the donation page on Panel (b), where "Amount" is a free text field into which the donor can type the Euro amount they wish to donate. The bottom half of the project page continues with the project description; the bottom half of the donation page requires the donor to fill in personal information required to complete the donation. Both have been suppressed in these screen shots.

in Figure 2.1(a). Most are directed there by charities soliciting funds from potential donors, typically via a link embedded in an email. These charities all have multiple fundraising methods, of which posting a fundraising page on Betterplace is only one. As such, for most if not all of these charities, successful fundraising efforts on the platform raise *additional* revenue with which to finance projects.

In order to understand how donations respond to natural disasters, we need to be able to classify each fundraising project and each donation as being either disaster- or non-disaster related. We accomplish this by matching each project, and by extension each donation, on Betterplace to a particular disaster event using double entry with reconciliation. We proceed in four steps detailed in Appendix Section 4.6 and summarized here. First, each project record (comprising a description, location, and dates) is manually read by two separate individuals. On the basis of this reading, each project is classified as either disaster relief or "other". Second, for projects classified as disaster relief, information from the project description on the disaster location, the type of disaster, and (if applicable) the disaster name is recorded. Third, any discrepancies in the double-entries for each project are reconciled by a third person. Finally, disaster relief projects are matched to entries in the EM-

DAT database, based on the recorded disaster characteristics, and time and location stamps. At the end of this exercise, we are able to match each project and donation record from Betterplace to a unique disaster event from EM-DAT, or a non-disaster-related cause.

Table 2.2 presents summary statistics for the entire observation period on charitable projects' demand for donations in Panel (a), and individuals' supply of donations in Panel (b), disaggregated by disaster- (column 1) and other, non-disaster-related, causes (column 2). Appendix Table 2.A.2 provides a breakdown across years. On the demand side, the platform witnessed the entry of 19,488 fundraising pages, through which 12,986 different charitable organizations requested on average €6,361. The large number of charitable organizations reflects the fact that, in addition to hosting household names such as *Doctors without Borders* and *CARE*, Betterplace also hosts thousands of smaller charitable organizations that might otherwise lack visibility, but have strong roots in local communities on the ground.[12]

Disaster relief fundraising accounted for just under 3% (543) of projects on the platform. Roughly 77% of these projects received some funding, and about 50% met their funding goals. Despite the fact that they requested only half as much as disaster relief projects on average (roughly €6,000 vs. €12,000), non-disaster-related projects had less success in fundraising: 73% of them received any funding and 43% met their funding targets. There is pronounced skewness in fundraising requests as the median request for disaster relief projects is €4000, only about a third of the mean. The median disaster relief project raised about €600, double what projects for other causes raised. Note that since fundraising requests can be adjusted dynamically by charities, and a project receives all donations given to it regardless of whether the fundraising request has been met, we do not use this variable in our subsequent analysis.

On the supply side, Panel (b), roughly 681,000 unique donations from 590,000 individual donors, totaling €35.2 million were given to projects on Betterplace. Of this total, just over €2.5 million (7%) went towards disaster relief. This came from a disproportionately large number of donations and donors, giving smaller amounts. More specifically, 17% of donations and 19% of donors contributed to disaster relief, with an average donation of €22, which is less than half the average donation to

[12] In fact, catering to the "long tail of charity" is Betterplace's raison d'être. See `https://www.betterplace.org/c/about-us/history`. The charitable organizations on Betterplace are overwhelmingly NGOs, although organizations like the ICRC and the U.N. also regularly post fundraising pages on the platform.

	Disaster relief (1)	Other (2)	Total (3)
# Fundraising pages	543	18,935	19,488
# Charities	412	12,720	12,986
Mean requested amount in €	12,242	6,188	6,361
SD of requested amount €	29,997	25,777	25,918
Median requested amount in €	4,000	2,190	2,243
Proportion receiving donations	0.772	0.727	0.728
Proportion request fulfilled	0.502	0.427	0.429

(a) Fundraising pages—demand side

	Disaster relief (1)	Other (2)	Total (3)
# Donations	115,146	565,755	681,114
# Unique donors	112,686	477,300	588,573
Mean donation in €	22	58	52
SD donation	89	232	215
Median donation in €	9	20	20
Total donations in €	2,543,445	32,646,194	35,196,791

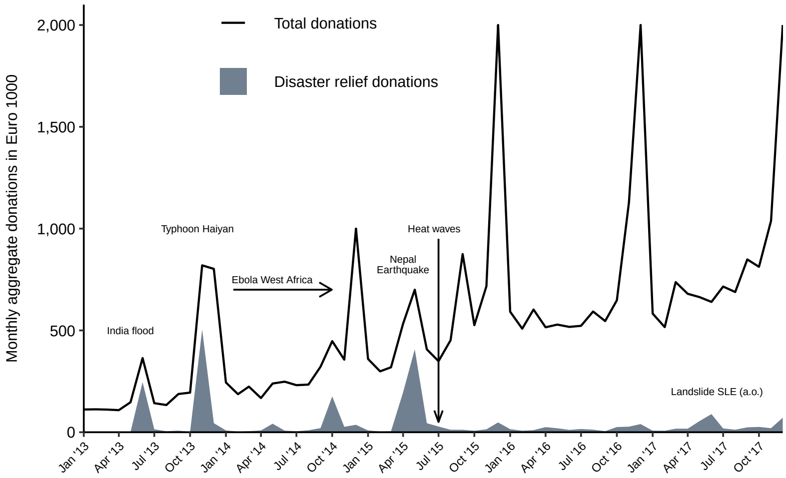
(b) Individual donations—supply side

Table 2.2: Donation demand & supply: Summary statistics. *Notes.* This table reports summary statistics of fundraising and donation activity on Betterplace broken down by disaster relief and non-disaster relief activity. Panel (a) summarizes the demand side and reports the number of fundraising pages; the number of charities that posted them; mean, standard deviation and median of the amount requested by these pages; proportion of pages donated to; and the proportion of pages that raised at least the amount they requested. Panel (b) summarizes the supply side and reports the total number of donations; the number of unique donors; summary statistics of donations (mean, standard deviation, and median); and the total volume of donations.

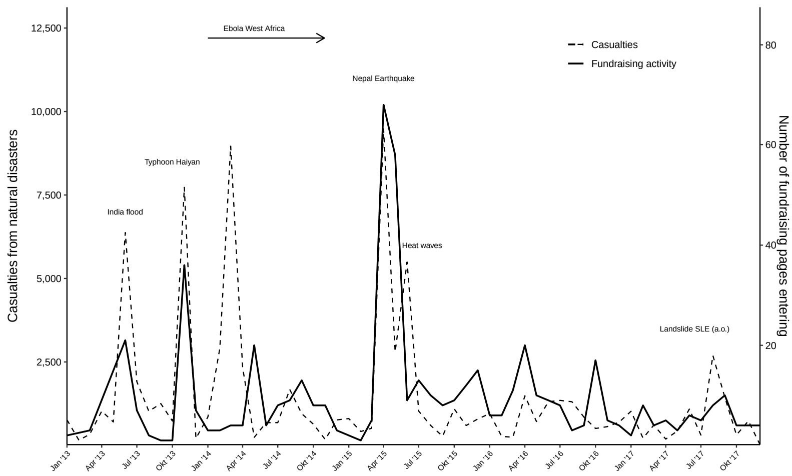
non-disaster-related projects (€58). The same holds true for the median donation (€9 for disasters vs. €20 for other causes). The donations distribution for both types of projects are right-skewed, with a large mass of individuals donating single- or double-digit amounts and a few very large donations. We account for this by transforming donations using the natural logarithm.

2.3.3 Time series patterns

Figure 2.2 describes the time series pattern of disaster-related casualties from EM-DAT and disaster relief donation demand and supply on Betterplace. The top panel depicts monthly time series of overall donations (solid line) and donations to natural disasters (shaded area). The bottom panel captures the time series of fundraising activity, measured by the number of disaster relief projects entering the platform (solid line), and natural disaster casualties (dashed line). Total donations on the platform have grown steadily over the years, reflecting the growing popularity of online donations in general and Betterplace in particular. Overall donations exhibit classic seasonality, with roughly 30% of giving transpiring at year-end, which coincides with both traditional Christmas giving and the end of the tax year. Our econometric models later on will account for this by including time fixed effects.



(a) Donations



(b) Casualties & fundraising activity

Figure 2.2: Disaster relief fundraising activity, donations, and casualties. *Notes.* Panel (a) depicts aggregate monthly disaster relief (grey filled area) as well as total donations on Betterplace (in €1000, black line). Total donations in the figure are censored at €2 million, although the underlying data are not; this only binds during year-end giving in 2015, 2016 and 2017. Panel (b) shows the time series of casualties across the 1,720 events in our sample (dashed line, left axis), as well as the number of disaster-relief fundraising pages posted on Betterplace (solid line, right axis). Major events are illustrated using their earliest start date. The abbreviations “SLE” and “a.o.” stand for Sierra Leone and among others, respectively.

Donations to disaster relief do not display the same time series regularities as overall donations, but do not appear to be entirely idiosyncratic either. For example, large spikes in both disaster relief fundraising activity and donations coincide with some major disasters, such as Typhoon Haiyan in the Philippines and the Nepal earthquake. But disaster severity cannot be the sole explanation because contemporaneous donation responses are virtually absent for other major disasters, such as the 2015 heatwaves in India and Pakistan (see Appendix Table 2.A.1). Beyond this, there is not much we can say. Given that disaster events overlap, it is impossible to attribute donations or project entry on the platform to particular disaster events using time series variation alone.

In order to understand the heterogeneity in donation responses, the relevant unit of analysis is an individual disaster. In what follows, we use this unit of analysis in our core sample to explore how the characteristics of individual disasters correlate with the demand and supply of charitable donations.

2.3.4 Additional data sources

In order to understand what accounts for variation in donation demand and supply, we supplement the core event-level dataset capturing donation activity as well as disaster severity with a number of data sources which we match, as appropriate, on the basis of country or event. First, we measure media attention devoted to the disasters in our sample by scraping the twitter feeds of five major German news outlets and applying regular expression matching using disaster types (storm, typhoon, etc.) and names (Haiyan, etc.) to identify tweets related to natural disasters.[13] Our measure of media attention an event received is the total number of tweets devoted to it by these media outlets; 97% of these tweets link directly to the corresponding coverage in the media outlet.

Next, we construct a measure of whether any charity present on Betterplace was active in a country prior to the respective country experiencing a disaster event using the project registry on Betterplace. We also measure the number of charities present in a disaster-afflicted country in 2011, two years before our observation window begins. We rely on the latter variable for identification in the IV estimator of Section 2.6.

[13] We scrape the entire twitter feed during our period of observation from 2013 to 2017 of *Bild Zeitung* (its print version is the daily tabloid newspaper with the highest circulation in Germany), *Frankfurter Allgemeine Zeitung* (FAZ), *Süddeutsche Zeitung* (SZ), *Welt* and *Zeit.*

	Mean (1)	S.D. (2)	Median (3)	25^{th} (4)	75^{th} (5)	N (6)
# of tweets	0.6	4.6	0	0	0	1,720
# of active charities	49	80	25	11	46	1,720
Trade with GER [1 Mio EUR])	22,649	35,809	2,914	204	26,099	1,682
Distance from capital to Berlin [1000km]	7.3	3.2	7.4	5.5	9.5	1,665
Freedom of press [/ 100]	0.52	0.23	0.5	0.33	0.71	1,699
Ease of doing business [/ 100]	0.61	0.13	0.6	0.53	0.69	1,690
Corruption perception [/ 100]	0.41	0.18	0.36	0.31	0.44	1,677
# of active charities 2011	23	37	11	4	21	1,720

Table 2.3: Correlates of giving—summary statistics. *Notes.* In this table we report the summary statistics (mean, standard deviation, median, 25^{th} and 75^{th} percentile, and the number of observations) of correlates of donation demand and supply. Each variable is described in Section 2.3.4. Note that the indices are divided by 100 and are all in the range of zero to one. For all indices, higher values represent a more benevolent political climate. " # of active charities 2011" is the stock of charities present in countries observed in the event data as of 2011. This is the instrument we employ in Section 2.6.1.

Third, we employ dyadic distance measures between Berlin, Germany's capital, and recipient countries' capitals using data from CEPII.[14] Economic integration is measured via total trade—the sum of imports and exports between a given country and Germany—using the UN's comtrade database.

Finally, we employ a number of indices to measure the socio-political environment of countries. We utilize the *Freedom of the Press* report compiled by *Freedom House*, a watchdog organization that compiles annual reports to systematically document threats to press freedom. Next, we supplement our data with the World Bank's *Doing Business* report which measures business regulations and their enforcement. Finally, we include records from the *Corruption Perception Index* (CPI) published by *Transparency International* on perceived corruption using surveys, business people's opinions and expert assessments. For all indices we use observations from 2012 or the earliest year after 2012 for which data is available.

Summary statistics of these variables are presented in Table 2.3. About 11% of events are tweeted about but the distribution of tweets is lopsided, with the median disaster getting no tweets at all. On average, 49 charities are active in a country prior to the onset of a disaster, and the distribution is right-skewed with a median of 21.

[14] CEPII is the *Centre d'Etudes Prospectives et d'Informations Internationales*, a leading French institute for international economics. We also matched Hofstede's index to our disasters sample to measure cultural distance between Germany and recipient countries, but do not present the results in this paper on account of the sample loss it entails. The results, with sample loss, are statistically insignificant in all our econometric specifications.

2.4 Which natural disasters get funding?

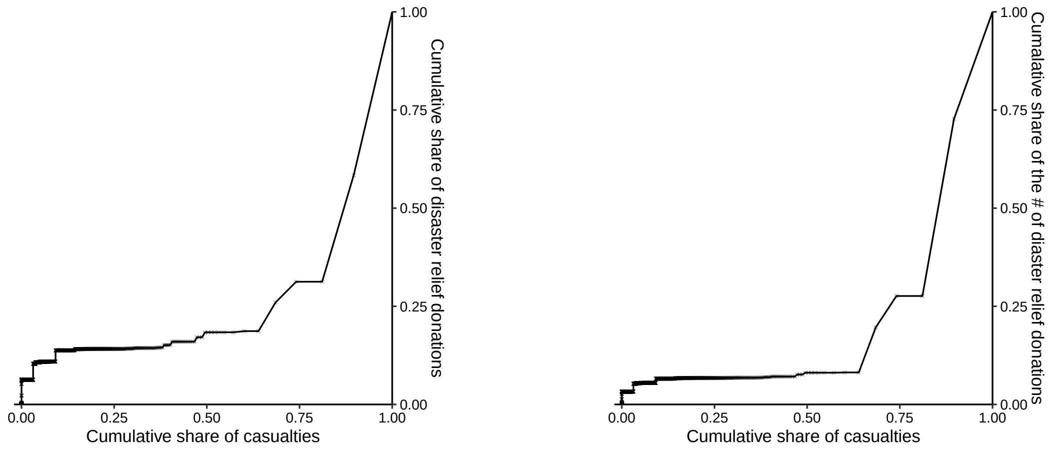
In this section, we ask which natural disasters get funding on this online donation platform. In order for a disaster to attract donations on Betterplace, two things must happen. First, one or more charities must solicit donations for this disaster by posting a fundraising page on Betterplace. Second, donors must give to these projects. What we observe in our data is the equilibrium demand and supply of donations of this two-stage process.

The analysis in this section essentially describes this equilibrium and its correlates, on both the donation demand and supply side. We begin with a graphical analysis and then estimate a simple regression model in order to examine correlates of donation demand and supply. In Appendix Section 4.6, we report additional results based on lasso regression, which investigate the most predictive covariates of fundraising activity following an event.

2.4.1 Graphical analysis

The pattern of disaster relief donations is extremely lopsided. The Lorenz curves in Figure 2.3 depict this graphically, by plotting the cumulative share of casualties on the x-axes against the cumulative share of disaster relief donations (Panel (a)) and the cumulative share of the number of donations (Panel (b)). Panel (a) shows that almost 80% of disaster relief funding accrues to a tiny number of events, which together account for barely 20% of casualties—a disaster relief version of the Pareto rule. The number of donations, depicted in Panel (b), mirrors this pattern. Disasters that together account for the lion's share of casualties, received approximately 20% of overall and individual donations.

This equilibrium outcome is not characterized by fundraising projects failing to attract donations. Rather, there is negligible demand for donations: the bulk of disasters do not experience any fundraising efforts on the platform. Figure 2.4 depicts the 1,720 disasters events (circles and triangles) that occurred between 2013-2017, plotted by the number of casualties (x-axis) and the number of affected people (y-axis). Of the 1,720 natural disasters, only 67 (3.9%) witness any fundraising activity (triangles). At the same time, 84% of events with fundraising pages on Betterplace (56 of 67) receive at least one donation (solid triangles). It would appear, therefore, that donors are potentially inclined to give to disaster relief. Unfortunately, giving is conditional on having a disaster relief project to give to, and disaster relief fundraising activity on the platform is anemic. The upshot is that, of the 1,720



(a) Cumulative share of disaster relief donations

(b) Cumulative share of the number of donations

Figure 2.3: Pareto rule in disaster relief donations. *Notes.* This figure shows the relationship between cumulative disaster casualties and cumulative disaster relief. Panel (a) plots the cumulative share of disaster relief donations (y-axis) against the cumulative share of casualties from natural disasters (x-axis). Panel (b) plots the cumulative share of the number of donations to disaster relief on the y-axis. Thus, an exemplary point in Panel (a) can be read as: "events that together account for x% of all casualties (x-axis) received y% of total disaster relief donations on Betterplace (y-axis)". A 45° line would correspond to a situation in which disasters received donations in perfect proportion to the number of casualties.

natural disasters that struck during this period, only 3.3% received *any* donations on Betterplace.

The paltry equilibrium response to disasters begs the question of why a handful of disasters elicit substantial donation responses, while the vast majority is effectively absent on this online platform. Figure 2.4 indicates that more severe disasters—in terms of casualties or number affected (or both)—are more likely to get traction on Betterplace by stimulating both demand and supply side responses. Figure 2.5 shows that media coverage, as measured by the number of tweets devoted to each disaster by Germany's five largest media outlets (see Section 2.3.4) is also correlated with event-level activity on Betterplace. Many disasters receive no media coverage. The most severe disasters in terms of casualties tend to receive considerable media coverage and, as we saw in Figure 2.4, these disasters are also likely to generate a donation response on Betterplace. In short, the descriptive evidence suggests that donations on the platform do not really cater to the mass of disasters which is relatively less severe in terms of casualties or number affected, and receives little media coverage.

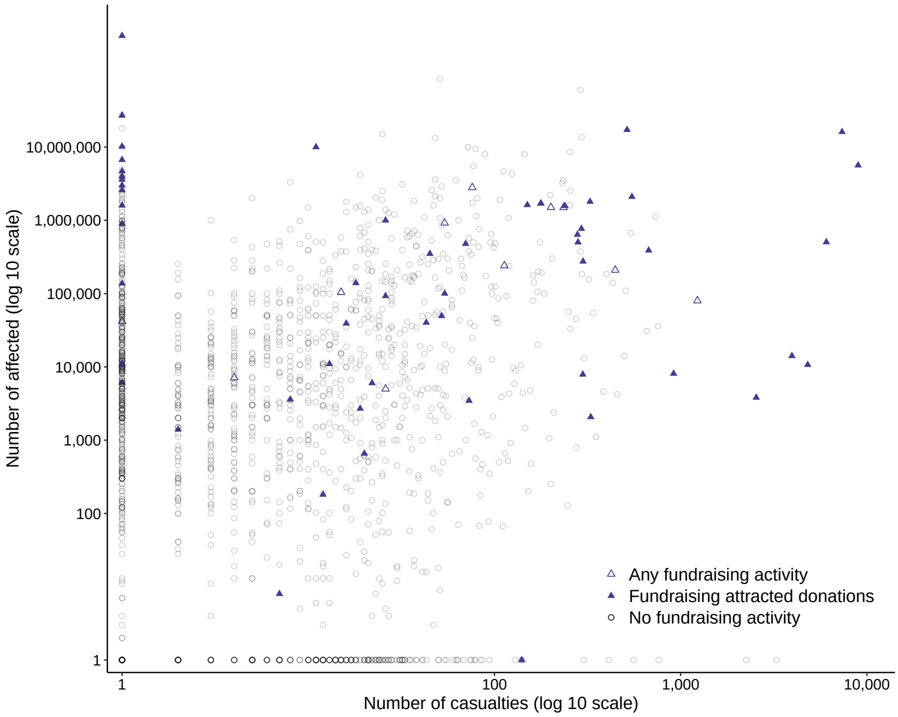


Figure 2.4: Disaster severity, fundraising, and donations. *Notes.* This scatter plot describes the demand- and supply-side responses on Betterplace to disaster events that transpired from 2013-2017. Disaster event casualties are on the x-axis, and the number of people affected is plotted on the y-axis. Both axes are scaled using a $\log_{10}$ transformation. Each marker on the scatter plot pertains to an individual disaster event ($n = 1,720$). Triangles denote events for which at least one disaster relief fundraising page was created on the platform ($n = 67$). Solid triangles capture fundraising pages which received donations ($n = 56$). Empty triangles denote events which did not receive donations ($n = 11$) despite there being fundraising activity for them on Betterplace. Circles denote events which generated no traction on the platform ($n = 1,653$).

2.4.2 Regression model

In order to more systematically explore the heterogeneity in demand and supply of donations to disasters, we follow an approach analogous to that of Strömberg (2007), who investigates the correlation between ODA for disaster relief and disaster severity, recipient country characteristics, news coverage, and donor-recipient relations. Concretely, we estimate the following model:

$$y_i = \beta_0 + X_i'\beta + C_i'\gamma + \alpha_{t(i)} + \varepsilon_i \tag{2.1}$$

where $i \in \{1; ...; 1,720\}$ denotes a particular disaster event, and y_i denotes demand- or supply-side outcomes. The demand-side response is captured by a dummy variable indicating whether or not a fundraising page devoted to disaster relief for event i was

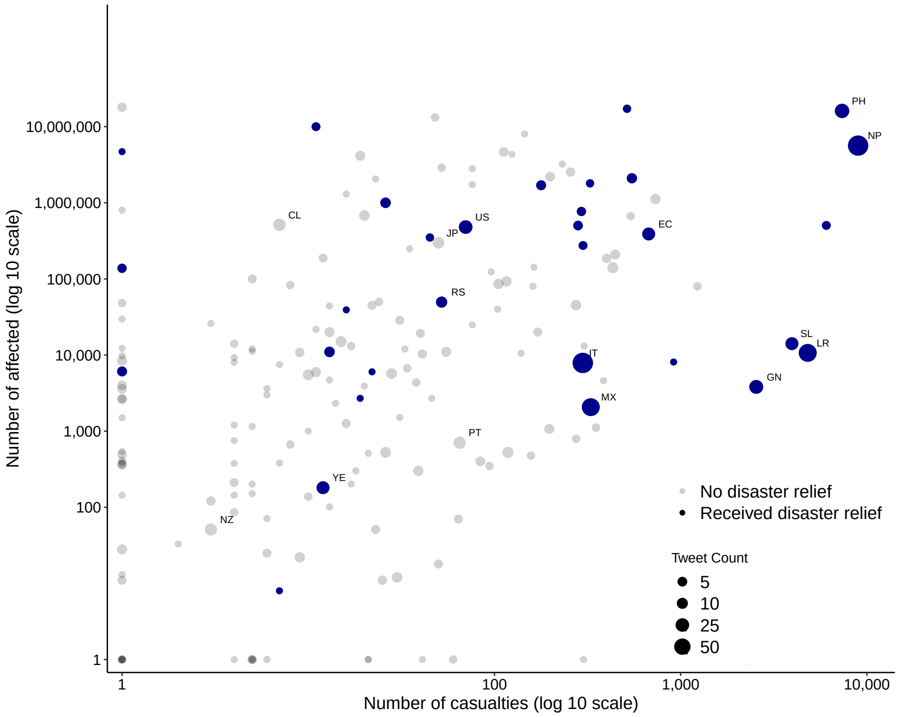


Figure 2.5: Media coverage of disaster events. *Notes.* This scatter plot shows the events that received media coverage and their severity in terms of the total number of casualties and affected. The solid blue circles depict events that received at least one donation on Betterplace. Media coverage is measured by tweet counts from Germany's major news outlets (n=181 disasters receive media coverage). The circles with neighboring country code labels (country codes per the ISO-α-2 standard) indicate the events with the top 1% mentions in the media. The size of the circles is proportional to the number of tweets that mentioned the respective event.

posted on Betterplace. The supply-side response is captured by the amount donated to (charitable projects devoted to) disaster i, measured in natural logarithms, to account for the right-skewed donations distribution evident in Table 2.1. We estimate logit models for binary outcomes, and linear models when the dependent variable is (the log of) the amount donated.

The coefficient vector of interest, β, captures the heterogeneity in donation response with respect to disaster characteristics X_i, which includes disaster severity; media coverage; distance metrics relative to Germany; socio-political indices; and prior charity presence in an affected country (see Tables 2.1 and 2.3 for summary statistics).

The vector C_i contains basic controls. These comprise dummy variables capturing whether the event (i) is listed with a precise onset date by EM-DAT, as opposed to only the month or even the year, since such events tend to be more likely to grab media and attention; (ii) span multiple countries, which is a measure of severity; and (iii) have no recorded deaths or people affected, both of which are likely to reflect measurement error. The controls also include a variable measuring the number of days between an event's onset and the end of the observation period (Dec 31^{st}, 2017), since events occurring later mechanically have less time to generate a response on Betterplace.

Finally, seasonality and the overall time trend, evident in Figure 2.2, are captured in quarter fixed effects, $\alpha_{t(i)}$, denoting the quarter $t = \{1; ...; 20\}$ in which event i took place.

2.4.3 Correlates

Table 2.4 presents estimates of the regression parameters in Equation (2.1). Odds ratio the presence of any demand for donations are reported in columns 1 and 2, although results are qualitatively identical with a linear probability model. The last two columns contain OLS estimates for the supply of donations, measured in logs. Columns 1 and 3 contain the full sample of 1,720 disaster events, with variables drawn from EM-DAT, Betterplace, and Twitter. Columns 2 and 4 include additional correlates. This entails some sample loss, and we ascertain in Appendix Table 2.A.3 that this is not what is driving the (negligible) differences in correlation patterns. The pattern of correlations is, unsurprisingly, identical for both the demand and supply of donations regardless of specification.

Both the number of casualties and the number of people affected are significantly positively correlated with donations, as is media coverage. These correlations, previewed in the descriptive analysis in Figures 2.4 and 2.5, are entirely consistent with results pertaining to ODA in Strömberg (2007), Eisensee and Strömberg (2007), and Fink and Redaelli (2011).

In the full sample, the demand estimates in column 1 indicate that an increase in the number of casualties by 10% is associated with a 5 percentage point increase in the odds of a fundraising page for disaster relief project being posted on Betterplace; the analogous change for the number of people affected is a 3 percentage point increase in the odds. The estimates suggest that, in proportionate terms, a tweet is "worth" considerably more than a casualty. A 10% increase in the number of tweets increases the odds of seeing fundraising activity on the platform by 12 percentage points. Column 3 indicates an elasticity of donation supply with respect to casualties of 0.1, indicating that disaster relief donations increase by 0.1% for a 1% increase in the number of casualties. The elasticity with respect to the number of affected people is of similar magnitude.

Prior charity presence in a country is positive and highly significantly correlated with demand and supply. Row 5 indicates that disasters occurring in countries with stronger trade ties with Germany and those that are more distant from Germany's capital are less significantly likely to witness donation demand (since they have an odds ratio less than one) or supply. The distance correlation is consistent with Strömberg (2007), while the trade correlation is not. This may be because strategic economic considerations are more likely to play a role in ODA decisions than they are in the market for private charitable donations.

Next, we see that although socio-economic indicators such as freedom of press, ease of doing business, and corruption may reasonably be expected to be related to charities' ease of operation in affected countries, or donors' willingness to give, none of these indicators is significantly correlated with either demand or supply of donations.

We report correlates of additional measures of both demand and supply, in Appendix Table 2.A.4. There, we ascertain that on the supply side, the correlates of the number of fundraising pages and the amount requested by these projects follow very similar patterns. The same is true for an indicator of whether an event received any donations, and the number of donations that went to all projects associated with an event.

Appendix Section 4.6 reports additional results based on regularized regression. It points to the most predictive covariates of donation activity for disasters on the platform. The analysis suggests that event severity and media coverage are indeed among the key predictors of such activity. The analysis also suggests that contemporaneous charity presence in a country is important in understanding the demand side response.

In general, the evidence presented in this section indicates that severe disasters with media attention tend to attract donations. As such, the promise of online platforms to cater to smaller-scale, more obscure disasters seems largely unrealized.

	Donation demand		Donation supply	
	1[Any fundraising] Logit odds ratios		Log total disaster relief OLS	
	(1)	(2)	(3)	(4)
Log # casualties	1.7	1.6	.1	.099
	(.23)	(.22)	(.043)	(.042)
Log # affected	1.3	1.4	.07	.07
	(.082)	(.1)	(.02)	(.018)
Log # of tweets	3.2	4	1.2	1.2
	(.58)	(.9)	(.17)	(.18)
Log # of active NGOs	1.6	2.2	.059	.13
	(.32)	(.4)	(.029)	(.034)
Log of trade with GER		.67		-.077
		(.069)		(.015)
Distance from capital to Berlin [1000km]		.76		-.023
		(.053)		(.0088)
Freedom of press [/ 100]		.46		.21
		(.65)		(.19)
Ease of doing business [/ 100]		8.3		.17
		(24)		(.45)
Corruption perception [/ 100]		.85		.58
		(2.2)		(.31)
Basic controls	Yes	Yes	Yes	Yes
Quarter-of-year FE	No	Yes	Yes	Yes
# positives dep var	67	62	56	51
Pse/Adj R-sq	.41	.49	.27	.29
N. of observations	1720	1604	1720	1604

Table 2.4: Correlates of disaster relief donation demand and supply on Betterplace. *Notes.* This table presents parameter estimates for Equation (2.1). The dependent variable in columns 1 and 2 is an indicator equal to one if at least one fundraising page for a given natural disaster was created on Betterplace, and zero otherwise. In columns 3 and 4, the dependent variable is the natural logarithm of total disaster relief for a given event on Betterplace. We show odds ratios from logit models in columns 1 and 2, and OLS estimates in columns 3 and 4. The covariates are described in Section 2.3.4, and "Log" refers to the natural logarithm. "Basic controls" contains indicators for when the date of the event can be pinpointed (rather than the month, only); when more than one country was affected; when the database indicates no casualties; when the database indicates nobody was affected; and the time in days from the time of the event until the end of the observation period (12/31/2017). We report standard errors clustered at the country level in parentheses.

2.5 Donor fatigue

In the previous section we saw that very few disasters attract donations. Donor fatigue is commonly blamed for this. In this section, we ask whether the pattern of donation demand and supply is consistent with donor fatigue. Concretely, we explore patterns along three dimensions. First, if donations spike following a disaster and then taper off to zero over time, this would be consistent with temporal fatigue. Second, if donations to disaster-related causes displace donations to other charitable causes, this would be consistent with crowding out. Finally, if disasters occurring in the direct aftermath of a major disaster that received generous donations fail to receive funding; whereas those that occur after a sufficient lag do receive donations, this would be consistent with disaster fatigue.

To test whether and to what extent the evidence is consistent with these three dimensions of donor fatigue we use an event study design, which exploits daily observations and the exogenous timing of disaster events. Concretely, for disasters that received donations, we estimate (i) temporal variation in donations to the disaster itself; (ii) contemporaneous variation in donations to other causes; and (iii) donations to other disaster events of varying temporal proximity to major disasters that received generous donations. These dimensions of variation map directly into the three dimensions of donor fatigue, namely, (i) temporal fatigue; (ii) crowding out; and (iii) disaster fatigue. The main identifying assumption is the exogenous onset of natural disasters, which is likely to be satisfied.

The analysis in this section is naturally restricted to the 67 events that witnessed fundraising activity on the platform, because the question of Betterplace donor fatigue only arises if disaster donations have been made in the first place. We further restrict our attention to the subset of 44 disasters with a precise onset date recorded in EM-DAT, which witnessed fundraising activity after this date on Betterplace. Our identification strategy relies on the sharp timing of natural disasters, such that the subsequent donation response can be clearly mapped. The final sample of disasters of 44 events captures the bulk of disaster-related activity in the full sample, accounting for 94% of total disaster relief donations on Betterplace, and 43% of disaster-related casualties.

We proceed by building a balanced panel containing these 44 disaster events, extending over all $t =\in \{1; ....; 1,826\}$ days of observations from 2013-2017 (2016 was a leap year). Around disasters' onset dates, we construct 42-day windows covering two weeks before and four weeks after disaster onset. Given that these disasters are plausibly exogenous natural hazards, we interpret statistically meaningful patterns in donations following (the timing of) a disaster event as a causal response to the event. These donations patterns, in turn, furnish evidence that is (or is not) consistent with donor fatigue.

2.5.1 Model

Our assessment of the three dimensions of donor fatigue is predicated on variants of the following generic event study model:

$$y_{it} = \sum_{\lambda=-14}^{27} 1_{[t=k(i)+\lambda]}\beta_\lambda + \alpha_\tau + \alpha_d + \alpha_i + \varepsilon_{it} \tag{2.2}$$

where the subscripts refer to disaster event i and day t; event i occurs on day $k(i)$. The indicator function $1_{[t=k(i)]}$ is equal to one when the condition in the index is satisfied. The dependent variable, y_{it}, varies depending on relevant dimension of donor fatigue. For temporal fatigue, it is the demand or supply of donations to event i on day t. When evaluating crowding out, it is donations to other causes in the wake of event i on day t. Finally, for disaster fatigue, it is the donation response to other natural disaster events that happen in the aftermath of i.

Naturally occurring time series variation in donations over the observation period is captured by week fixed effects α_τ, where $\tau \in \{1; ...; 260\}$. To account for day-of-week patterns in giving we include day-of-week fixed effects α_d, where $d \in \{1; ...; 7\}$; α_i is an event fixed effect, which accounts for event specific unobserved time invariant effects; ε_{it} is the error term. Standard errors are robust to heteroskedasticity and serial correlation.

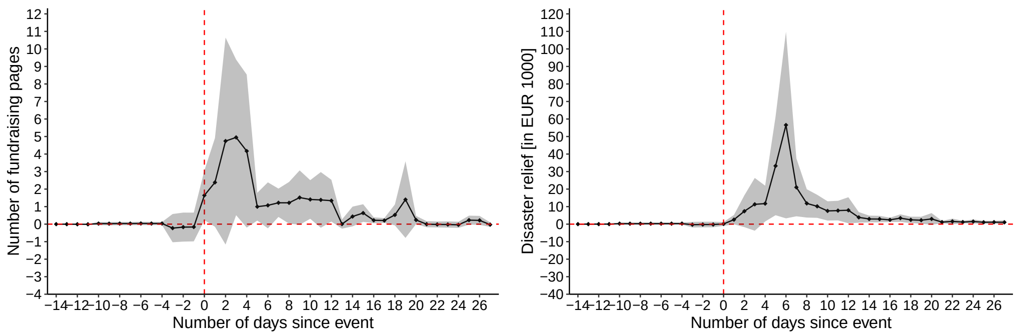
The coefficient vector of interest is β_λ. For $\lambda \geq 0$, β_λ captures the dynamic response to an event i that occurred on day λ relative to the onset date, $k(i)$, in its four-week aftermath. As previously explained, the interpretation of these coefficients as the causal response to an event rests upon the identifying assumption that the timing of events and corresponding fundraising effort decisions on the platform are uncorrelated with market conditions. Pre-trends during the two weeks prior to events' onset are captured by β_λ where $\lambda < 0$. We expect these coefficients to not be statistically different from zero.

2.5.2 Results

In what follows, we present coefficient plots of the event study estimates, which speak to whether fundraising and donation responses are consistent with temporal fatigue, crowding out, and disaster fatigue in turn. The plots report point estimates for β_λs in Equation (2.2) along with their 95% confidence intervals. In these figures, the coefficient estimates for $\lambda < 0$ furnish a visual test of our identification strategy, which relies on the assumption that outcomes are not driven by pre-event donation market trends. This implies that coefficients for $\lambda < 0$ should not be statistically different from zero. This is exactly what we observe for for all our outcome variables.[15]

Temporal fatigue

Figure 2.6 examines whether fundraising and donation patterns for the 44 disasters in the event study sample are consistent temporal donor fatigue. Because these disasters vary greatly in severity and time patterns can be more clearly mapped for larger events, the regressions are weighted by a severity index. This index is equal to the mean of the normalized number of casualties and affected individuals (subtracting each series' minimum, and dividing by its maximum), and lies in the range $[0, 1]$.



(a) Disaster relief fundraising

(b) Disaster relief donations

Figure 2.6: Temporal fatigue. *Notes.* Panel (a) depicts the daily fundraising activity, measured as the number of new disaster relief fundraising project pages posted on the platform in response to a disaster event. Panel (b) depicts the daily disaster relief donations response, measured as total donations going to disaster relief fundraising projects. The underlying data structure and the model are described just before and in the beginning of Section 2.5.1. The error bands show the 95% confidence interval.

Panel (a) of Figure 2.6 shows that disaster relief fundraising activity is concentrated in the one-week aftermath of an event, with an immediate up-tick in fundraising activity following disaster onset, reaching a peak with four to five fundraising pages being posted by charities each day. It highlights the strength of online fundraising, namely the speed with which it can kick into gear. However, in week two this

[15] The fact that β_λ is not always a precisely estimated zero for disaster relief fundraising and donations just prior to event onset stems from the fact that week fixed effects span days on both sides of event onset for some disasters.

activity abates, reverting to zero in week four.

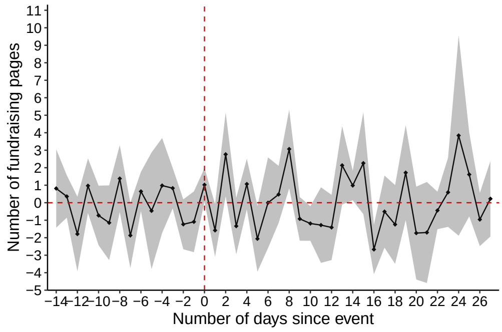
Panel (b) describes the supply-side response. This is more gradual than the demand response, which is to be expected since disaster-relief donations in this context are conditional on disaster-relief fundraising activity. Donations witness a profound increase near the end of the first week following an event. Point estimates indicate a daily donation response of about €60,000 at its peak. Donations fall equally precipitously, settling back down to zero by the end of week two.

Although the estimates are noisy, the time series pattern of fundraising and donations towards disaster relief are consistent with temporal fatigue. The speed with which fundraising efforts and donations are mobilized is impressive. It indicates that charities are able to leverage this online platform to generate funding in a timely manner for their disaster relief activities. At the same time, donations dry up within a couple of weeks of disaster onset, indicating that there is a narrow window of opportunity for charities to fundraise for disaster relief before temporal fatigue sets in. This is consistent with anecdotal accounts of donor fatigue being responsible for the shortfall in disaster relief funding months after disasters like Typhoon Haiyan.

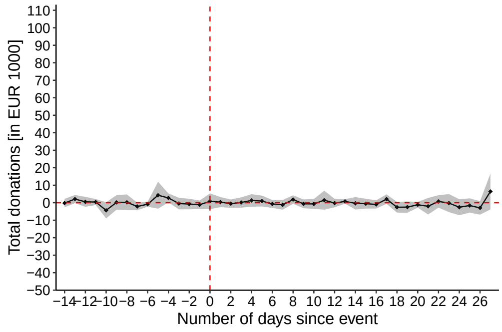
Crowding out

Figure 2.7 describes contemporaneous changes in fundraising and donations to non-disaster-related—other—causes. It shows a wildly different pattern, wherein estimates varying insignificantly around zero over the event window. Crucially, there is no discernible change in the noisy fundraising pattern, Panel (a), before and after the occurrence of a disaster event. In other words, there is no evidence that disaster relief fundraising efforts crowd out fundraising efforts for non-disaster-related causes on the platform. Neither is there any evidence, Panel (b), that there is crowding out in terms of giving to other causes. Here again, we estimate precise zeros over the entire event window.

It is worth noting that the sample used in this analysis is selected on the basis of an endogenous variable, namely whether a fundraising page for a disaster was posted on Betterplace. This constitutes a threat to identification to the extent that fundraising activity is correlated with unobserved factors that drive donation outcomes. For example, if projects are likely to enter the platform due to a surge in generosity that coincides with a disaster occurrence, then we are likely to underestimate potential crowding out. We believe that strategic entry of this sort is likely to play a role only for marginal disasters, which are at risk of receiving no donations.



(a) Other fundraising



(b) Other donations

Figure 2.7: Crowding out. *Notes.* Panel (a) depicts the daily fundraising activity, measured as the number of new non-disaster relief fundraising project pages posted on the platform after a disaster event. Panel (b) depicts the corresponding daily volume of donations to causes not related to disaster relief. The underlying data structure and the model are described just before and in the beginning of Section 2.5.1. The error bands show the 95% confidence interval.

In addition, unobserved heterogeneity should be captured in the pre-event window. Such pre-trends in donations and fundraising activity are not evident in Figure 2.7.

Nevertheless, we address this concern in robustness checks by restricting the attention to the most severe events that are least likely to be marginal (i.e., highly likely to attract donations). Appendix Figure 2.A.3 presents estimates based on using only the most severe events for identification. We use the top 10 and top 2 events—see Panel (d) of Appendix Table 2.A.1—and show that there was no crowding out by these more severe events either. This also addresses, and rejects, the concern that since the majority of funding goes to a handful of events, evidence of crowding out is diluted by the large number of events that received little funding.

In the previous section we saw that for those disasters that receive funding, fundraising and donations spike in the direct aftermath of a disaster event. The results in this section indicate that this increase in disaster relief funding does not come at the expense of other charitable causes. Together, these results suggest that on this platform, donor fatigue is not manifested in crowding out, at least not contemporaneously. This would seem to confirm anecdotal evidence that fears regarding crowding out in the wake of generous donation responses to major natural disasters may be misplaced. For example, the tax break aimed at offsetting potential crowding out of charitable donations after the outpouring of donations for Hurricane Katrina is widely thought to have been completely unnecessary (New York Times, 2006).

Disaster fatigue

In this section, we explore disaster fatigue by asking whether disasters that occur in close temporal proximity to a major disaster are less likely to spur fundraising and donations than more temporally distant disasters. In order to conduct this exercise, we need to make a judgement call regarding two questions: what is a “proximate” disaster and what is a “major” disaster? We define proximate as a two-month window following a major disaster. The rationale is that, as we saw in Section 2.5.2, all the disaster relief action on the platform takes place within a month of disaster onset. The second month both allows us to explore persistence of disaster fatigue over a longer time horizon, and gives us more statistical power. Beyond the two months, we are faced with potential confounding changes that compromise causal inference.

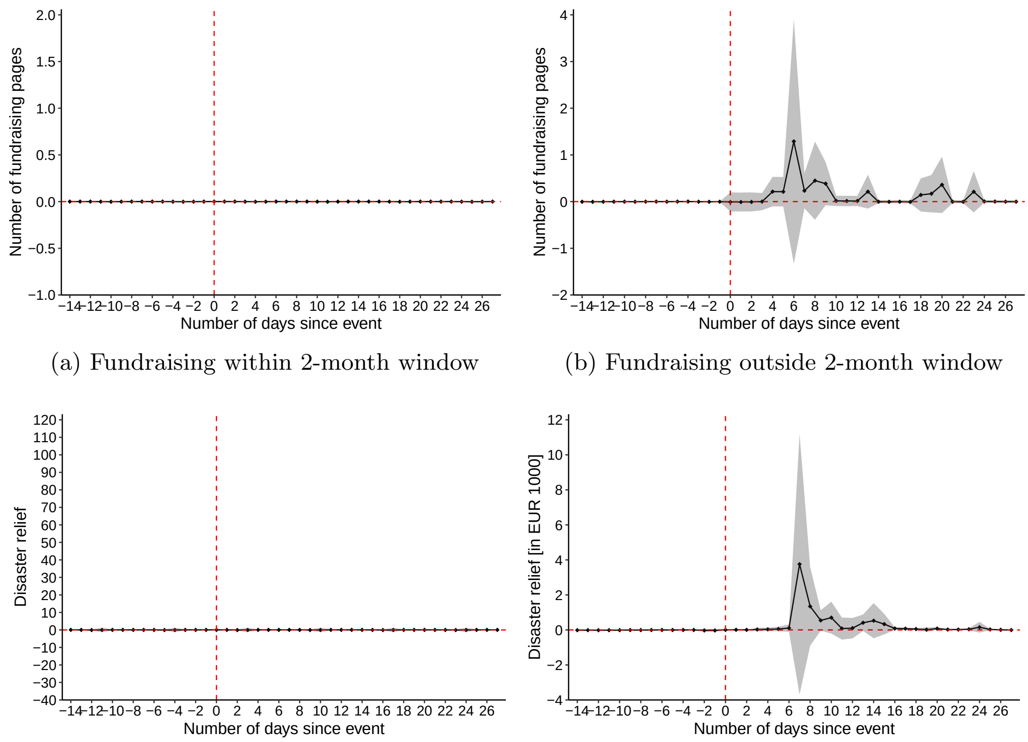
As for what constitutes a major disaster, we focus on the two events in our full sample that (i) have a precise onset date; (ii) inflicted the first and second largest number of casualties; and (iii) together attracted 70% of overall donations. These are the Nepal earthquake in April of 2015, which killed almost 9,000 people; and Typhoon Haiyan in November of 2013, which killed over 7,300.

Two patterns in the data would be consistent with disaster fatigue. First, the 118 disasters that occur within two months after the Nepal earthquake or two months after Typhoon Haiyan should receive meager, if any, funding. Second, disasters that occur outside of these two 2-month windows should witness relatively more funding. The challenge we face is that disasters that are proximate or distant are, by definition, different. This raises the concern that differential outcomes in proximate and distant windows may be attributable to differences in disasters rather than differences in proximity to major disasters.

To address this concern, we construct a matched sample of disaster events that occur within two months of the two large events, and those that occur outside this window. We then examine how much fundraising and donations transpires for events within two months—*treated* events—as opposed to outside these two months—*control* events. The matched samples of treated and control events is generated using a classic propensity score matching procedure (Rosenbaum and Rubin, 1983).[16] We estimate the propensity using only the natural logarithms of casual-

[16] The propensity score is estimated using a logit-model in which the dependent variable is an indicator equal to one if a disaster is treated, and zero otherwise. The propensity is given by the predicted probability from this model. Prior to matching, all events whose propensity score falls outside the common support are discarded.

ties and the number of affected people, as well as, indicators for an event having zero deaths or affected. Since we have complete data for these variables, this allows us to use the full sample of treated events for the match. We then leverage the scope of our data and match five control events to each of the 118 treated events using nearest neighbour matching without replacement in random order. Appendix Figure 2.A.4(a) shows matching balance.



(a) Fundraising within 2-month window (b) Fundraising outside 2-month window

(c) Donations within 2-month window (d) Donations outside 2-month window

Figure 2.8: Disaster fatigue. *Notes.* This figures compares fundraising activity for and disaster relief donations to events within two months—Panels (a) and (c)—of Typhoon Haiyan and the Nepal earthquake to a matched sample of events outside that two month window—Panels (b) and (d). Panels (a) and (b) show fundraising activity, measured as the number of new disaster relief fundraising project pages posted on the platform in response to disaster events. Panels (c) and (d) shows the daily disaster relief donations response, measured as total donations going to disaster relief fundraising projects. The matched sample contains five control events for each treated event using propensity scores based on events' severity. The procedure is detailed in Section 2.5.2. The underlying data structure and the model are described just before and in the beginning of Section 2.5.1. The error bands show the 95% confidence interval.

Figure 2.8 investigates whether fundraising and donation patterns are consistent with disaster fatigue. The outcome variable in Panels (a) and (b) are daily fundraising project entry on Betterplace for disasters *other than* the 2015 Nepal earthquake and Typhoon Haiyan. Panels (c) and (d) show the corresponding estimates for the amount of daily donations. The panels on the left capture fundraising and donation activity for treated events—those that occurred in the proximate two-month window after the major disasters. Those on the right show the analogous outcomes for control events—those that occurred outside this two-month window.

The results are broadly consistent with disaster fatigue. In particular, the two left hand panels show that disaster events occurring within two months after the Nepal earthquake and Typhoon Haiyan do not experience *any* fundraising activity (Panel (a)) and, consequently, do not receive *any* donations (Panel (c)). The latter would seem to fit the American Red Cross' experience with Hurricane Maria after Hurricane Harvey. By contrast, events that occur outside this proximate window, experience both fundraising (Panel (b)) and donation activity (Panel (d)). The coefficients are imprecisely estimated, but the contrast between the patterns on the left and right are patently obvious. Interestingly, the disaster fatigue evidenced here is not mirrored in media fatigue. As shown in Appendix Figure 2.A.5, there is media coverage of disaster of both treated and control events.

Robustness checks presented in the appendix show analogous results to Figure 2.8 using a different specification for the propensity score. There, we estimate the propensity using all the covariates used in Section 2.4, which reduces the sample of treated events from 118 to 111, as we lose some observations due to incomplete records on covariates. Because the model is more saturated, making it difficult to do one-to-many matching, we conduct one-to-one nearest neighbour matching without replacement in random order. This procedure is successful in obtaining a match for each treated observation, and Appendix Figure 2.A.4(b) shows matching balance. The results from this exercise are presented in Appendix Figure 2.A.6. The donation patterns are qualitatively identical, consistent with disaster fatigue.

2.6 Fundraising for disaster relief

As alluded to in the introduction, charities are major players in local disaster relief. They are almost completely reliant on private individuals for charitable donations to these projects; and these are chronically underfunded. Online platforms such as Betterplace provide an opportunity fundraise for disaster relief at a relatively low cost. Yet, we saw in Section 2.4 that donation demand for bulk of smaller scale disasters is anemic. As we just saw in Section 2.5, disaster fatigue is one possible explanation.

Nevertheless, two factors suggest that by failing to post disaster relief projects on

Betterplace, charities may be missing an opportunity to raise money for them. First, disaster fatigue does not characterize large intervals of our observation period. For example, 2016 and 2017 did not experience any major international disasters whose donation response aftermath could reasonably have been characterized by donor fatigue. Second, the descriptive evidence in Section 2.4 indicates that, conditional on demand, donation supply tends to be forthcoming. Of course, as an equilibrium response this is merely suggestive.

In this section, we ask how much additional disaster relief funding could be raised on Betterplace if charities posted fundraising pages on the platform.[17] Providing a causally interpretable answer to this question entails estimation of the donation supply function: we want to know how much *would* have been donated to disaster relief projects if charities had only fundraised for them.

2.6.1 Model and estimation

Concretely, we would like to estimate the donation supply for disaster event i:

$$s_i = X_i'\alpha + \rho d_i + \varepsilon_i \tag{2.3}$$

where X_i are event-specific characteristics and $d_i \in \{0,1\}$ denotes there being any fundraising effort, i.e. demand for donations towards disaster i on the platform.

Estimation of ρ is complicated by the fact that events which witness demand for donations are likely to be the same events to which people feel inclined to supply donations to, so $E[d\varepsilon] \neq 0$. Measures of disaster severity and country characteristics captured in X go some way in addressing the issue, but are bound to be inadequate.[18] As Eisensee and Strömberg (2007) note, disasters tend to have an "unobserved salience". *How* a tropical storm affects people, or is perceived to affect people, is imperfectly captured by a headcount of how *many* people are affected. The term "affected" itself is an elastic concept. Similarly, a measure of media coverage captured in a tweet count is unlikely to adequately convey the same sense of severity

[17] Our underlying assumption here is that there exist charities that are in the position to provide disaster relief. While we have no way of directly verifying this, it seems to be a reasonable assumption given the fact that charities are the principle providers of disaster relief in practice, and 20,000 of them, including organizations like the German Red Cross and Doctors without Borders are active on Betterplace.

[18] Intuitively, one would expect the estimate of ρ to be larger when X is not included in the regression. The reason is that X contains elements, such as event severity, which correlate positively with both the demand and supply of donations. The data confirms this intuition.

or extent of distress imparted in a turn of phrase or a picture in the news. This type of unobserved heterogeneity is likely to be positively correlated with both charities' disaster relief activities and individuals' donations, and would thus lead to upward bias in $\hat{\rho}^{OLS}$.

To identify the causal effect, we instead rely on an instrumental variable to generate plausibly exogenous variation in fundraising activity. As the instrument, we use *historical* charity presence (in its natural logarithm) in the country where event i occurred. This variable is measured in terms of the number of charities present in a disaster-afflicted country in 2011, two years before our observation window begins. Instrument relevance is built on the premise that general charity presence in a country likely reflects lower costs of providing disaster relief, and generally an environment more conducive to charity operations, which in turn requires funding. Higher charity presence thus increases the likelihood that charities initiate and provide disaster relief in a given country. As we saw in Section 2.5.2, charities only have a narrow window of opportunity within which to raise funds for disaster relief. The need for a timely response further motivates the instrument as charities are more likely to be able to respond quickly when they have had "boots on the ground" in the disaster-afflicted country.

The main identifying assumption is that charity presence in 2011 only affects *demand* for donations by triggering fundraising activity on the platform after having conditioned on disaster severity, media coverage, and country's socio-political climate, including corruption and freedom of press. Furthermore, employing the 2011 rather than a more proximately lagged stock of charities contributes to ensuring instrument exogeneity.[19]

In order to account for the disaster fatigue documented in the previous section, we restrict the sample to events that happened either (i) at least 30 days *before*, or (ii) at least 90 days *after* one of the two major events. The former restriction is aimed at dampening concerns over temporal fatigue, given the finding from Section 2.5.2, that donations can persist until up to four weeks following disaster onset. The latter minimizes exposure to disaster fatigue, which we saw in Section 2.5.2 impedes fundraising in the two months following disaster onset. Together, these relatively conservative thresholds reduce the initial set of 1,720 events to 1,443. The sample with full coverage on covariates of 1,604 is reduced to 1,346—with 50 events that

[19] When we instead use the stock of charities present in the month prior to an event as instrument, the results are unchanged. The instrument is valid by conventional thresholds, and estimates are about 1% smaller than those we report in Table 2.5, and remain highly statistically significant.

witness fundraising activity.

In this sample, the mean (median) country has about 13 (4) active charities in 2011. More importantly, as the map in Appendix Figure 2.A.7 shows, there is considerable variation in the value of the instrument both across countries, and within regions. This variation is key to identifying the causal effect of disaster relief fundraising activity on donation supply,

Estimation proceeds in two steps. In the first step we estimate a logit regression, which accounts for the binary nature of the endogenous regressor indicating any fundraising activity for an event ($d_i \in \{0, 1\}$), on the instrument and the vector of covariates (X_i). We use this to obtain the predicted probability, $\hat{d}_i$, of fundraising activity for event i.[20] In the second step, predicted values are used as instrument for d_i in Equation (2.3), which is then estimated using the standard 2SLS procedure. The resulting estimates of ρ are consistent (Angrist and Pischke, 2008; Wooldridge, 2010).

2.6.2 Results

Table 2.5 presents parameter estimates for Equation 2.3 for the (log of) the total amount donated and the number of donations towards disaster events. Columns 1 and 4 present OLS estimates, columns 2 and 5 present 2SLS estimates, with a first stage logit model, and columns 3 and 6 present 2SLS estimates with a first stage linear probability model (LPM). The first thing to notice is that the OLS estimates for ρ in columns 1 and 4 are *smaller* than the corresponding 2SLS estimates. This may seem surprising in view of our prior that unobserved disaster characteristics are likely to be positively correlated with both fundraising activity (demand) and giving (supply). However, it likely reflects the fact that the OLS estimate averages the effect of fundraising activity across *all* events in the data; by contrast the IV estimates capture the effect of fundraising activity on the *sub-population* of disaster events in countries with prior charity presence large enough to expect disaster relief operations and fundraising among charities who have (at least historically) been active on Betterplace. Disaster events in this sub-population are likely to be more successful at generating donations than the full set of events, which includes disasters that are not especially "salient" to potential donors and therefore unlikely to attract

[20] Our results are robust to choosing a probit specification in the first step. In fact, probit models produce slightly stronger test statistics for instrument validity, and marginally larger point estimates in the second stage (about 2% larger). Statistical significance in the second stage is not affected.

donations.

Tests for instrument relevance in the first stage are reported at the bottom of Table 2.5. We provide χ^2-statistics for our preferred specifications in columns 2 and 5, and F-statistics for the models in columns 3 and 6. These statistics indicate instrument validity in the former set of models, but only weak relevance for models in the latter set. This is unsurprising as the linear model is ill-suited to capture the skewed nature of the project entry variable. We discuss this at length in Appendix 4.6.

Appendix Table 2.A.5 shows the first stage estimates for both specifications. Appendix Figure 2.A.8 reports the predicted probabilities obtained by estimating the logit model for the demand function. Unsurprisingly there is significant bunching around zero which corresponds to minor events that were very unlikely to see any fundraising activity. However, there are some events with an extremely low probability that did see fundraising pages being created.

The results suggest that demand shortfalls constitute a barrier to disaster relief donations on Betterplace. We estimate a causal effect of fundraising activity on disaster relief donations of about €1,000 ($\approx \exp(6.9)$).[21] This implies that a marginal disaster—an event for which a large enough prior charity presence could warrant fundraising activity—would have been able to raise slightly less than twice the median donation to a disaster relief project of € 590 in this sample, or 3.5% of average disaster relief project donation, conditional on entry.[22] In columns 3 and 4 we estimate that fundraising activity would have attracted around 37 individual donations to those marginal events, with an average donation of € 27. Columns 1 and 4 of Table 2.5 suggest that ignoring omitted variable bias would led us to underestimate the disaster relief fundraising on donations—the OLS estimate indicates that only about € 220 would have been raised.

[21] Recall that these results are based on dropping events within 30 days before and 90 days after one of the two major events. An even more conservative approach would be to only use events in the years 2016 and 2017—years without a major disaster (Typhoon Haiyan happened in November 2013, and disaster fatigue likely extends into 2014). Even in this more restricted sample—2016 and 2017—we estimate a causal effect of a comparable magnitude of about €500.

[22] The qualitative implications of these results are not driven by the largest disasters (left in the sample) attracting a disproportionate share of donations. In Appendix Tables 2.A.6 and 2.A.7 we re-estimate these specification after dropping the ten most severe events; and separately, the events receiving the ten largest donations (in the subset of 1,336 considered). While point estimates become smaller, the estimates still suggest that charities could raise economically meaningful amounts by soliciting donations for marginal disasters.

	Log total donations to disaster relief			Log # of donations to disaster relief		
	OLS	2SLS		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
1[Any fundraising]	5.4	6.9	6.4	2.5	3.6	3.2
	(0.43)	(0.56)	(1)	(0.27)	(0.47)	(0.67)
Log # casualties	-0.014	-0.053	-0.041	-0.0072	-0.036	-0.025
	(0.026)	(0.026)	(0.038)	(0.021)	(0.018)	(0.026)
Log # affected	-0.0044	-0.015	-0.011	-0.0026	-0.01	-0.0072
	(0.0033)	(0.0049)	(0.0088)	(0.0025)	(0.0033)	(0.0055)
Log # of tweets	0.46	0.29	0.34	0.32	0.2	0.24
	(0.11)	(0.089)	(0.17)	(0.086)	(0.074)	(0.12)
Log of trade with GER	-0.018	-0.012	-0.014	-0.015	-0.011	-0.012
	(0.0081)	(0.0081)	(0.0088)	(0.0058)	(0.0058)	(0.0062)
Distance from capital to Berlin [1000km]	0.00046	0.0062	0.0044	-0.0017	0.0025	0.00087
	(0.0063)	(0.0068)	(0.0087)	(0.0037)	(0.004)	(0.0049)
Freedom of press [/ 100]	0.24	0.37	0.33	0.21	0.3	0.27
	(0.12)	(0.14)	(0.13)	(0.09)	(0.1)	(0.098)
Ease of doing business [/ 100]	-0.028	0.0018	-0.0076	-0.11	-0.083	-0.092
	(0.29)	(0.3)	(0.3)	(0.2)	(0.21)	(0.2)
Corruption perception [/ 100]	0.23	0.3	0.28	0.26	0.31	0.29
	(0.17)	(0.19)	(0.17)	(0.14)	(0.15)	(0.14)
Basic controls	Yes	Yes	Yes	Yes	Yes	Yes
Quarter-of-year FE	Yes	Yes	Yes	Yes	Yes	Yes
First stage		Logit	LPM		Logit	LPM
χ^2 / F first stage		11	7.5		11	7.5
Adj R-sq	0.75	0.71	0.73	0.62	0.55	0.59
N. of observations	1346	1346	1346	1346	1346	1346

Table 2.5: Effect of fundraising activity on disaster relief. *Notes.* This table presents estimates of the effect of a project creation on the natural logarithm of total disaster relief (columns 1-3), and the natural logarithm of the number of donations to all fundraising pages devoted to an event (column 4-6). Columns 1 and 4 report OLS estimates, while columns 2,3,5 and 6 report 2SLS instrumental variable estimates. The dependent variable is the natural logarithm of total disaster relief for a given event in columns 1-3, and the natural logarithm of the number of disaster relief donations to all fundraising pages for a given event in columns 4-6. In columns 2,3,5 and 6 we instrument for "1[Any fundraising]" using the natural logarithm of the number of charities present in a country in 2011. Columns 2 and 4 show estimates from the two-step procedure described in Section 2.6.1; columns 3 and 6 show estimates from a standard 2SLS procedure. "χ^2/F first stage" reports the test statistic of instrument relevance in the first stage. See Section 2.3.4 for definitions of variables. We report standard errors clustered at the country level in parentheses.

2.7 Conclusion

This paper has investigated the demand and supply of donations to natural disasters on an online charitable donations platform, by asking three questions. First, which natural disasters tend to get funding? Second, is the pattern of donations consistent with donor fatigue? Third, how much funding could charities have raised for marginal disasters, had they solicited donations?

We find that this online platform does not cater the vast majority of natural disasters, which cumulatively wreak the lion's share of devastation. Instead, donation patterns conform to a version of the Pareto rule, whereby 80% of donations go to a tiny fraction of disasters responsible for merely 20% of casualties. Those events that do attract funding activity tend to be severe disasters, which enjoy media coverage. Pertinently, the equilibrium demand and supply of donations on this platform is characterized by a paucity of disaster-relief fundraising activity on the part of charities. Although disaster relief projects in general tend to be successful at fundraising on the platform, very few disaster events witness any fundraising activity.

The contrast between the failure of this online donations platform to cater to smaller-scale disasters, and the success of conventional e-commerce platforms in selling niche products is striking. One possible explanation is that online fundraising in this form is simply not part of charities' business models. This seems unlikely given the large number and variety of charities represented on Betterplace. Another is that charities do not need the extra funding, either because they already have it or because they are not involved in disaster relief operations. This too seems dubious given the large funding gap in disaster relief and the fact that charities are known do the heavy lifting in relief activities on the ground.

This leaves two arguably more plausible explanations for this discrepancy. First, one of the reasons for the success of online product platforms is that they cater to heterogeneous preferences. It is hard to imagine what the analog of heterogeneous preferences over product characteristics would mean in the context of natural disasters or indeed, if such heterogeneity exists at all. Second, the importance of search engines in enabling customers to find products they prefer is another feature that has been credited for the success of online markets to cater to niche products. In other words, reducing barriers to fundraising and donating may not be enough to remedy the neglect of disasters. Additional tools may be needed in order to reduce information frictions and raise awareness. It is probably in acknowledgement of this fact that Betterplace, and its U.S. analog GlobalGiving, have started using social media (e.g., Facebook crisis response pages) or e-mail newsletters to draw attention to more natural disasters. An open question for future research is whether such efforts will be successful in spurring both demand and supply donation responses to neglected disasters.

Using an event-study design, which exploits high-frequency data and the exogeneity of natural disaster occurrences for identification, we find that donation patterns are consistent with temporal fatigue and disaster fatigue, but not with crowding out. The results indicate that this online setting allows charities to fundraise for disaster relief in a timely manner without the worry that successful efforts on that count will crowd out donations for other charitable causes on the platform. But there are two caveats. First, temporal fatigue means charities have a narrow two-week window within which to raise money on the platform after disaster onset. Second, disaster fatigue may compromise their ability to raise money for disasters which follow on the heels of a major disaster.

Finally, our IV estimates suggest that there may be unexploited potential for disaster relief fundraising on the platform. The estimates indicate that charities could have raised about €1,300 in additional funding for marginal disasters, equivalent to twice the median event donation volume, had they solicited donations on the platform. This amount is unlikely to make or break a project. As such, fundraising for disaster relief on the platform seems unlikely to be the kind of "silver bullet" online platforms have proved to be for catering to the long tail of niche products. Nevertheless, the failure to solicit donations on the platform does seem to present something of a missed opportunity.

3. Identifying and Teaching High-Growth Entrepreneurship Experimental Evidence from Academies for University Students in Uganda[1]

Abstract

We present an ongoing randomized control trial in Uganda which disentangles the extent to which entrepreneurial success of university students can be attributed to skill formation and to selection. To study skill formation we randomly accept applications to a training program fostering an entrepreneurial mindset. We study individuals' motivation for entrepreneurship by experimentally varying marketing messages prior to students' application decision, emphasizing either entrepreneurial profit or freedom. Lastly, we non-experimentally describe endogenous self-selection by comparing key outcomes of applicants and eligible students who did not express interest. We track labor market outcomes of all groups for up to three years.

[1] This chapter is based on joint work with Vojtěch Bartoš (LMU Munich), Kristina Czura (University of Groningen, the Netherlands), Timm Opitz (Max Planck Institute for Innovation and Competition (Munich) and LMU Munich and Brendan Shanks (LMU Munich). Please note that the project presented in this chapter is ongoing as of February 2021.

3.1 Introduction

Entrepreneurship is key for economic development (Schumpeter, 1934). While most individuals in low-income countries are self-employed (e.g., 78.1 percent of the working population in Uganda was self-employed in 2019), these are mainly small-scale businesses that are only remotely related to the Schumpeterian entrepreneurship that drives economic growth (Hsieh and Olken, 2014; Porta and Shleifer, 2008). They typically lack capital and entrepreneurial ability, preventing them from reaping the full benefits of high-return investment opportunities (Beaman, Magruder, and Robinson, 2014; Bruhn, Karlan, and Schoar, 2018; De Mel, McKenzie, and Woodruff, 2012). While relieving credit constraints shows some improvement in terms of business profits, it does not result in sustained business growth (Banerjee, Duflo, Glennerster, and Kinnan, 2015). Interventions aimed at improving business practices and managerial capital have not been shown to result in sustained increases in profits or employment (McKenzie and Woodruff, 2014). More promising approaches focus on the role of the psychology of entrepreneurship. Campos, Frese, Goldstein, Iacovone, Johnson, McKenzie, and Mensmann (2017) show that training programs focusing on soft skill concepts, such as *personal initiative* and the *entrepreneurial mindset*, outperform programs teaching accounting, finance and marketing skills.[2]

Most business training studies target existing businesses—with the notable exception of Klinger and Schündeln (2011), Blattman, Fiala, and Martinez (2014), and Premand, Brodmann, Almeida, Grun, and Barouni (2016)—but neglect the importance of selection into entrepreneurship. Levine and Rubinstein (2017) and Levine and Rubinstein (2018) provide evidence that successful entrepreneurs in the USA are positively selected on human capital. Moreover, evidence from high-income countries shows that cognitive and non-cognitive traits predict entrepreneurial success (Andersen, Di Girolamo, Harrison, and Lau, 2014; Koudstaal, Sloof, and Van Praag, 2016; Levine and Rubinstein, 2017). Yet little is known on whether non-cognitive traits are shaped by entrepreneurial activity, or whether people select into entrepreneurship based on these traits. This distinction is important for policy. If relevant non-cognitive traits are malleable, this would favour programs aimed at developing an entrepreneurial mindset. If they are not, interventions designed to identify high-potential entrepreneurs would be more promising.

[2] Entrepreneurial mindset is one's ability to spot and benefit from opportunities that are encountered in daily life. Personal initiative captures one's desire to proactively tackle problems (Frese, Krauss, Keith, Escher, Grabarkiewicz, Luneng, Heers, Unger, and Friedrich, 2007).

We seek to disentangle the extent entrepreneurial success can be attributed to skill formation and to selection. First, we causally identify the effects of a business training program, which develops an entrepreneurial mindset, on business creation and business performance. In our field experiment, training is randomly offered to university students in Uganda who had expressed interest in entrepreneurship, a suitable sample positively selected on human capital. Second, we study how selection into the entrepreneurship training program varies by motives and personality traits. Using panel-data drawn from the same population, we document how students interested in entrepreneurship differ from those that are not with respect to socio-economic, cognitive and non-cognitive factors, as well as labor market outcomes, including self-employment. Third, we causally identify what motives draw students to entrepreneurship training.

We partner with a Ugandan organization, StartHub Africa, that provides extracurricular entrepreneurship training academies at leading local universities. We track three semesters of training academies (henceforth "waves") conducted at eight to ten universities with a combined enrollment of around 2,000 students in our study sample.[3] Each wave consists of a marketing campaign, an application phase, and an entrepreneurship training academy. A wave begins with an untargeted marketing campaign to raise general awareness of the program. Then, to be eligible for the program, students must attend an information session that consists of short presentations that summarize the training program. This is also where the application forms are distributed.

Our experimental design relies on two sources of exogenous variation. First, we randomly vary the motivational message for becoming an entrepreneur that is marketed in the information sessions' video presentations: financial gains or creative freedom. This allows us to causally identify the motivations of applicants. Second, among those who apply, we randomly offer admission to the program to identify the effect of being offered admission on business creation, survival and performance. We complement these analyses by documenting patterns of entrepreneurial self-selection by comparing applicants to those who were aware of the training program but did not express interest along several repeated measures of socio-economic indicators, person-

[3] Two waves have been conducted to date. We plan to include one more wave. We will discuss the feasibility of this extension and base our power calculations both on the status quo and the planned implementation.

ality traits and preferences.[4] The data collection effort includes surveys at different points in the self-selection and application process, as well as surveys administered both before and after the entrepreneurship training academies (Figure 3.1).

This study relates to four strands of literature. First, we contribute to the literature on entrepreneurship and business training in low-income countries by studying a unique sample of highly-educated, high-potential individuals (see Levine and Rubinstein (2017) and Levine and Rubinstein (2018)) who aspire to be entrepreneurs. Despite extensive research on business training interventions, there is a paucity of evidence on the effects of training on high-skilled youths. Interventions in low-income countries typically provide middle-aged, incumbent micro-entrepreneurs with education on business skills and managerial capital, which have not been found to result in sustained increases in revenue, profits or employment (Hsieh and Olken, 2014; Bruhn, Karlan, and Schoar, 2018; Bruhn and Zia, 2013; McKenzie and Woodruff, 2014; McKenzie, 2017; Rigol, Hussam, and Roth, 2018). This population, however, may lack the necessary skills for becoming successful entrepreneurs (Levine and Rubinstein, 2018; Bjorvatn and Tungodden, 2010; Hurst and Pugsley, 2011; Carlson and Rink, 2019) or may be unwilling or unable to change the way they run their businesses (Burmeister and Schade, 2007). With respect to our target population, the most closely related study is Premand, Brodmann, Almeida, Grun, and Barouni (2016) who analyze the inception of an official entrepreneurship track at universities in Tunisia. They document modest increases of one to four percent in self-employment rates but no effect on overall employment.[5] Our setting differs from theirs in that we study an extracurricular program that is more likely to only attract the genuine subpopulation of those interested in pursuing entrepreneurship.

Second, we contribute to the literature on the entrepreneurial mindset. The entrepreneurship training program we study is based on a curriculum that aims to foster an entrepreneurial mindset and personal initiative. Campos, Frese, Goldstein, Iacovone, Johnson, McKenzie, and Mensmann (2017) show that this type of training results in larger increases of profits than a traditional business training program. Ubfal, Arraiz, Beuermann, Frese, Maffioli, and Verch (2019) find transient, short-

[4] We elicit data on the Big-5 personality traits, grit, personal initiative and aspirations. Further, we gather measurements of time and risk preference as well as individuals' degree of loss aversion.

[5] This speaks to substitution from wage employment to self-employment, and does not imply overall employment effects. Alaref, Brodmann, and Premand (2020) present results from a medium term follow-up and show that any effects were short lived: four years after the program, there are no differences in self-employment and wage employment rates between the treatment and control groups.

term effects of this type of training on micro-entrepreneurs in Jamaica. We complement this burgeoning literature by offering further evidence on the merits of non-traditional training programs and enhance it by focusing on nascent entrepreneurs who have been found to benefit from traditional training programs (see Klinger and Schündeln, 2011).

Third, we contribute to the literature on selection into entrepreneurship and predictors of entrepreneurial success. Levine and Rubinstein (2017) show that successful entrepreneurs select along both cognitive and non-cognitive dimensions. Evidence from high-income countries suggests that cognitive and non-cognitive traits are important predictors of entrepreneurial success (Andersen, Di Girolamo, Harrison, and Lau, 2014; Koudstaal, Sloof, and Van Praag, 2016; Levine and Rubinstein, 2017). For example, entrepreneurs are generally more risk-tolerant (Bouchouicha and Vieider, 2019) and display more overconfidence (Åstebro, Jeffrey, and Adomdza, 2007; Herz, Schunk, and Zehnder, 2014). Evidence is scarce on whether non-cognitive traits are shaped by entrepreneurial activity or whether people select into entrepreneurship based on these traits. An established view suggests that preferences are relatively stable (Schildberg-Hörisch, 2018). There is however recent evidence that personality traits, such as grit, may be malleable—at least among young adolescents (Alan, Boneva, and Ertac, 2019). We extend this literature by documenting personality traits, preferences, and beliefs before individuals select into entrepreneurship, how these differ by interest in entrepreneurship, and by identifying how entrepreneurship training affects these characteristics.

Fourth, we speak to the motivations of becoming an entrepreneur, and whether selection patterns differ by motivation. A sparse literature using observational data from the USA stresses that non-pecuniary benefits, such as being one's own boss or having flexible working hours, play a first-order role for business creation decisions and that these independence-oriented workers are willing to forgo higher earnings from wage-employment (Hamilton, 2000; Hurst and Pugsley, 2011, 2015). Guzman, Oh, and Sen (2020) and Ganguli, Huysentruyt, and Le Coq (2018) confirm the importance of motives and differential responses to monetary and non-pecuniary motives resulting in selection patterns into entrepreneurship competitions in randomized field experiments in the USA and the UK, respectively.[6] We complement this recent literature by identifying the differential selection decisions made by high-skilled youth

[6] Ashraf, Bandiera, Davenport, and Lee (2020) vary the salience of career incentives in a recruitment drive for public health workers in Zambia, and also show that the salience of motives affects selection patterns, and later, performance on the job.

in a low-income country using random variation in the salience of different motives for entrepreneurship.

3.2 Research design

3.2.1 Background

StartHub Africa (SHA) conducts the academy at local universities during the academic semester. There is one academy per university which has a target class size of 40 students that spans nine weeks with one three-hour session each week. The academy covers all stages of training for nascent entrepreneurs: developing a business idea, creating a prototype, and implementing the idea. In the curriculum developed by SHA, management skills, such as cost accounting, and basic principles of finance and marketing are included, but emphasis is placed on developing participants' personal initiative to foster their entrepreneurial mindset. In this respect the training program is similar to the program studied by Campos, Frese, Goldstein, Iacovone, Johnson, McKenzie, and Mensmann (2017). Lecturers are encouraged to create an interactive atmosphere, and the standardized materials SHA provides to the instructors require active input from the participants. Finally, the curriculum contains a number of practical exercises outside of the classroom. For instance, students are taught basic principles of market research, then brainstorm product ideas and spend the rest of the session venturing out on campus to assess people's reaction to their product ideas. The training is taught by university lecturers or respected entrepreneurs from the local community that have been extensively trained and are continuously supported by SHA.

The academy is preceded by a marketing and application phase which spans the first three weeks of the semester. During the marketing phase, SHA creates awareness of the program using posters and flyers across campus, and in short pitches in classrooms and at campus events. Students are informed that attending an information session is a prerequisite for applying. Six to twelve of these 30-minute sessions are held per day over two or three days in a central location at each university. The information sessions provide detailed information on the academy's content, the expectations of the participants, in particular the time commitment necessary to complete the academy, success stories from previous participants, and the possibility to ask questions to SHA staff. To harmonize the information sessions as much as possible, the same SHA staff hold the information sessions throughout each day. Moreover, the presentations are video-based and contain the same structure: motivation for the academy, details, deliverables and requirements of the academy, and success stories from alumni. After the information session, students could pick up an application form in person, fill it out (in 10 to 15 minutes) and return it either to the team conducting information session, or to a well-know place on campus indicated on the application form. Application forms were only available to participants of the information sessions.

3.2.2 Experimental design

We exploit two sources of exogenous, experimental variation. First, in the *entrepreneurship training experiment*, admission to the academy is randomly offered to a subset of applicants. We use this variation to estimate the causal effect of being offered admission to the academy on entrepreneurial activity and economic outcomes. We also investigate effects on cognitive and non-cognitive skills. Second, to understand in more detail the characteristics and motivations of these young entrepreneurs, we add a second layer of exogenous variation: In the *selection experiment*, we randomly vary whether marketing for the academy emphasizes financial independence or creative freedom as motivation for becoming an entrepreneur. This variation allows us to identify how motivation impacts the application decision and to study heterogeneous effects based on individual characteristics. Figure 3.1 presents the experimental design. Finally, using a sample drawn from the same population, we document endogenous self-selection by comparing eligible students who did not express interest in the academy to applicants. We also investigate how key outcomes from the *entrepreneurship training experiment* evolve differentially over time between students who did not express interest to those who applied but did not receive training.

We first discuss the *selection experiment* and the complementary observational examination of self-selection, and then the *entrepreneurship training experiment* because this follows the chronological journey of a student from hearing about the training to submitting an application and possibly being offered admission. Nonetheless, the main research question draws on hypotheses about the *entrepreneurship training experiment.*

Understanding selection and motives.

The first layer of experimental variation is induced by randomly exposing clusters of students to different marketing messages during the information sessions. In

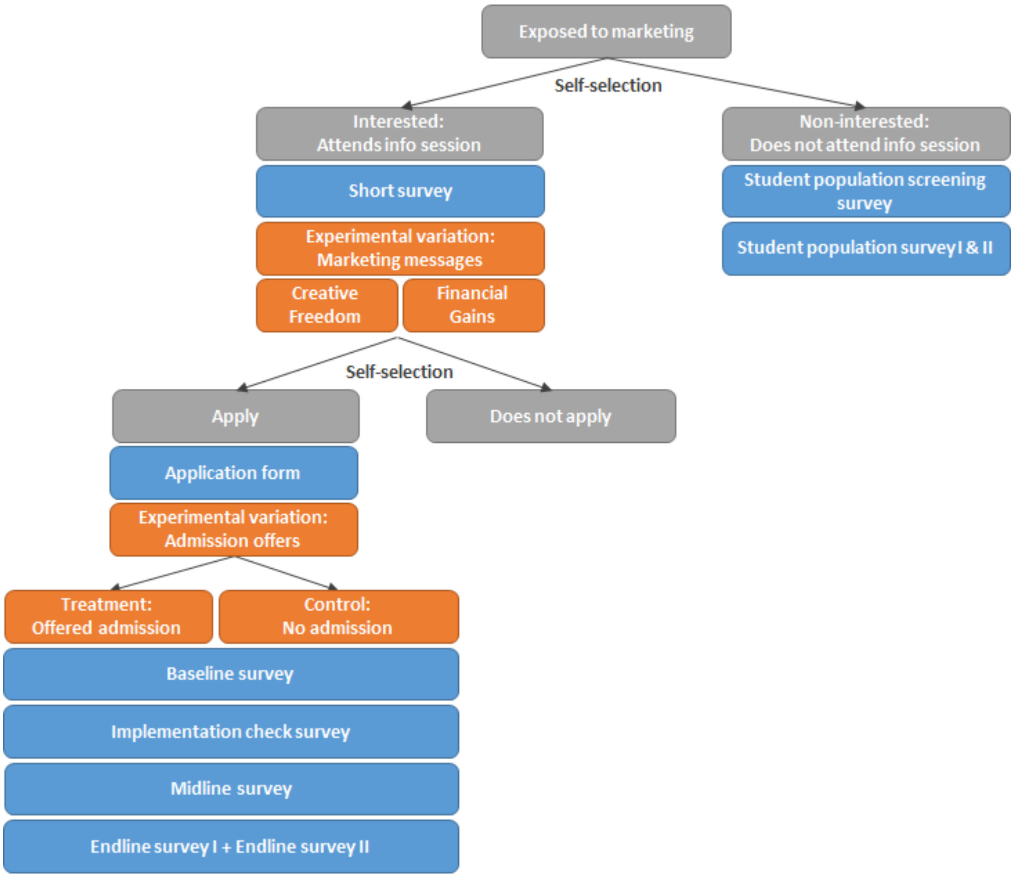


Figure 3.1: Experimental design and data collection. *Notes.* Different phases of the experimental design and self-selection decisions is marked in grey, exogenous experimental variation is marked in orange, and data collection is marked in blue.

the *selection experiment*, the content of two motivational video presentations is randomly varied between emphasizing i) that entrepreneurship offers the possibility of achieving financial independence, and ii) that entrepreneurship offers the freedom to be creative. For this, the respective motives are varied in four of the twelve overall slides reiterating the benefits of becoming an entrepreneur and in the corresponding voice-over of these slides.[7] Everything else is kept constant. Support staff ascertained that no student listened to two information sessions by either staying in the room for the next session or entering early during an ongoing session. This exogenous variation allows us to cleanly identify how the pool of applicants differs across these two messages.

To analyze selection into the academy, we compare students who are interested in entrepreneurship, indicated by applying to the academy, with those who are not interested in entrepreneurship indicated by being aware of the entrepreneurship academy and not attending an information session. We refer to this latter group as the *non-interested subpopulation.* In other words, conditional on having been exposed to the marketing phase, we investigate what drives certain individuals to opt-in to the academy.

Entrepreneurship training experiment

The *entrepreneurship training experiment* allows for causally estimating the effect of the academy on individuals' self-employment probability, as well as on labor market outcomes and personality traits. Having participated in an information session, students decide whether to apply to the academy. A random sample, stratified by year and field of study, is then drawn from the set of all applications and offered admission to the training program — the treatment group. The remainder is placed into the control group.

3.2.3 Hypotheses

Grounded in the results of previous work, there are several hypotheses we seek to test. The first set of hypotheses concerns the effects of entrepreneurship training on economic and business outcomes and inputs. First, as shown by Klinger and Schündeln (2011) for a traditional entrepreneurship training program, we hypothesize that participating in the entrepreneurship academy fosters business creation.

[7] In Appendix Section 4.6 we present in detail how information sessions differed across the two marketing themes.

Yet, as our entire sample consists of highly-educated students that are all interested in entrepreneurship, we may not find significant differences between treatment and control groups at the extensive margin. Therefore, we further hypothesize that participation in the academy will improve business performance, captured by indicators such as monthly sales and profits, measures of capital and labor input, and measures of general economic self-sufficiency. One particular dimension we are interested in is labor input, and whether treated subjects create jobs through the businesses they create. The hypotheses are summarized in Table 3.1, Family 1.1. Positive findings for these hypotheses would provide evidence for entrepreneurial activity being teachable.

Second, we seek to identify channels through which the entrepreneurship training effects the primary outcomes of business creation and performance. Campos, Frese, Goldstein, Iacovone, Johnson, McKenzie, and Mensmann (2017) find that a personal initiative training program can deliver lasting improvements for small business owners and they identified several channels: application of successful business practices, increased personal initiative, increased capital and labor inputs, substantial innovative activity (e.g., in the form of new products originating from own ideas) and product differentiation. We therefore hypothesize that participation in the academy leads to implementing more successful business practices, greater financial professionalization, marketing activities, product and process innovation, and better access to business networks. The hypotheses are summarized in Table 3.1, Family 1.2, Hypotheses 1 to 6. Finding effects along these dimensions would lend evidence to the most effective channels through which entrepreneurship training impacts individuals' economic outcomes.

Moreover, as laid out before, there is evidence that entrepreneurs are positively selected on cognitive and non-cognitive traits. Little is known, however, about whether non-cognitive traits may be shaped beyond adolescence. We therefore test hypotheses that investigate whether participating in the academy shapes non-cognitive traits. These hypotheses are summarized in Table 3.1, Family 1.2, Hypotheses 7 and 8. These hypotheses allow us to test whether — and to what extend — non-cognitive traits are malleable through participation in entrepreneurship training.

The second set of hypotheses concerns selection into entrepreneurship. First, individuals may have different motives for desiring to be an entrepreneur. Guzman, Oh, and Sen (2020) study entrepreneurs and find that women and individuals located in more altruistic cultures are motivated more by social-impact messages than money, whereas men and those in less altruistic cultures are motivated more by money than potential social-impact. Ganguli, Huysentruyt, and Le Coq (2018) document crowding-out between extrinsic, cash-based and intrinsic, social motives for social entrepreneurs. While extrinsic motivational messages affect effort in applications for a start-up grant, they reduce the pool of applicants at the same time. Further, business success was less likely: social entrepreneurs motivated by extrinsic messages worked fewer hours per week, created fewer employment opportunities, and profited less from their venture. We therefore test which marketing message attracts more applicants: whether monetary motives or the promise of independent work better draws young, highly-educated individuals to entrepreneurship. We also investigate the types of individuals that are drawn to the different marketing messages. We consider measures of average cognitive ability, over-confidence and entrepreneurial self-assessment. These hypotheses are summarized in Table 3.1, Family 2.1, Hypotheses 1 to 4. These hypotheses test whether stressing different motivations for becoming an entrepreneur lead to differential application patterns, both in terms of the quantity of applications and the attributes of the applicants themselves. Finding differences between the two messages would also speak to how different motivations to undertake entrepreneurship training are correlated with certain individual characteristics, and how such motivations shape the composition of applicants.

Further, we document selection into entrepreneurship (as proxied by selection into the academy) by comparing those that applied to the academy to those that were exposed to the marketing campaign but did not apply for the program (*non-interested subpopulation*). The outcomes of interest are listed under Hypothesis Families 2.2.1 and 2.2.2, and mirror those in Hypothesis Families 1.1 and 1.2 from the primary outcomes of the *entrepreneurship training experiment*. Comparing baseline characteristics and outcomes between the two groups allows us to identify the dimensions on which individuals select into entrepreneurship. Those and additional measures are investigated at endline to document how the non-interested subpopulation evolves over time compared to those that applied to the training and were not admitted (control group).

The outcome variables and their measurement are detailed in Section 3.3.2, while the empirical analysis is detailed in Section 3.4. Our results will inform to what extent teaching entrepreneurial skills and selection are important aspects for entrepreneurship. This is interesting from an academic perspective as it addresses fundamental questions on skill formation and its potential repercussions for en-

Family	H #	Hypotheses title	Index	Data collection				Sample	Exogenous variation
				Baseline	Midline	Endline	Short		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. Entrepreneurship training									
1.1. Economic	1	Business creation	✓	✓	✓	✓			Random assignment to
outcomes	2	Business success	✓		✓	✓		Applicants	training (*entrepreneurship*
(primary)	3	Capital and labor input	✓		✓	✓			*training experiment*)
	4	Economic self-sufficiency	✓		✓	✓			
1.2. Business	1	Business practices	✓		✓	✓			
and personal	2	Financial professionalization	✓		✓	✓			
input	3	Marketing	✓			✓			
(secondary)	4	Innovation	✓			✓			Random assignment to
	5	Networks	✓	✓	✓	✓		Applicants	training (*entrepreneurship*
	6	Entrepreneurial mindset	✓	✓	✓	✓			*training experiment)*
	7	Owner's non-cognitive traits	✓	✓	✓	✓			
	8	Preferences	✓	✓	✓	✓			
2. Selection									
2.1 Selection into	1	Submitted application					✓		Random assignment to
entrepreneurship	2	Cognitive ability					✓	Attendees at Info	marketing messages (*selection*
(primary)	3	Over-confidence	✓				✓	Sessions	*experiment*)
	4	Entrepreneurial self-assessment	✓				✓		
2.2.1 Economic	1	Business creation	✓	✓		✓		General population	
outcomes	2	Business success	✓			✓		(exposed to marketing	
(non-experimental)	3	Capital and labor input	✓			✓		campaign)	
	4	Economic self-sufficiency	✓			✓			
2.2.2 Business and	1	Business practices	✓			✓			
personal input	2	Financial professionalization	✓			✓			
(non-experimental)	3	Marketing	✓			✓			
	4	Innovation	✓			✓		General population	
	5	Networks	✓	✓	✓	✓		(exposed to marketing	
	6	Entrepreneurial mindset	✓	✓	✓	✓		campaign)	
	7	Owner's non-cognitive traits	✓	✓	✓	✓			
	8	Preferences	✓	✓	✓	✓			

Table 3.1: Overview of families of hypotheses.

trepreneurship. It is also of utmost importance for policy: If entrepreneurial skills can indeed be formed, we offer an evaluation of a cost-effective, relatively easy to implement, and scalable intervention for high-potential, well-educated individuals. We can also document whether the nascent entrepreneurs originate from high-skilled individuals that would otherwise be unemployed or whether they are substituting away from formal-employment. If selection is found as relatively more important for entrepreneurial success, our study would inform policy makers that identifying high-potential entrepreneurs is of first-order importance (see McKenzie (2017) and Rigol, Hussam, and Roth (2018) who seek to identify high-potential entrepreneurs, and Shane (2009) who warns about dragging people into risky, non-growth entrepreneurship). Our results would also offer some guidance on the motives that attract these entrepreneurs-to-be.

3.2.4 Time frame

The proposed project consists of three waves of entrepreneurship training academies. Each wave consists of the implementation of the entrepreneurship academies, the experimental variation introduced in both the *entrepreneurship training experiment* and the *selection experiment*, and the data collection before and after the intervention. As detailed below, there will be a baseline survey, an implementation check survey (one to two months after the intervention), a midline survey (six months later) and two endline surveys. The Endline Survey I takes place 12 months after the intervention, the Endline Survey II then 24 months after the intervention of the last wave.

The first wave started in September 2019, and the second wave started in January 2020. The third wave is scheduled for September 2020. The Endline Survey I will take place in February 2021 (Wave I), July 2021 (Wave II), and February 2022 (Wave III). The Endline Survey II is scheduled for February 2023 for all three waves. We expect to finish the analysis in the summer of 2023. Table 3.2 sets out the detailed time line for all steps in all waves. The implementation of Wave I and Wave II is already in progress, while the data collection for the midline survey (Wave I) and the implementation check survey (Wave II) in 2020 is scheduled. Later data collection and the implementation of Wave III is planned.

Due to the current Covid-19 crisis, we may not be able to implement Wave III as planned in September 2020, but have to postpone it to the spring semester 2021. In this case, all of the following dates will be postponed by around six months. In the

	Stage/Instrument	Sample	Status	Date
Piloting		3 academies, 380 applicants	Completed	March-May 2018
Wave I		10 academies		
	Marketing / information sessions / short surveys	$n = 1019$	Completed	Aug.- Sept. 2019
	Application data / Baseline survey	$n_{app} = 713$, $n_{baseline} = 672$	Completed	Aug.- Sept. 2019
	Enterpreneurship academy	$n = 414$	Completed	Aug. 2019 - Jan. 2020
	Implementation check survey	$n = 625$	Completed	Jan. - Feb. 2020
	Midline survey		Scheduled	Sep. - Oct. 2020
	Endline survey I&II		Planned	Jan. - Feb. 2021 & Jan. - Feb. 2023
Wave II		8 academies		
	Marketing / information sessions / short surveys	$n = 760$	Completed	Feb. - March 2020
	Application data / Baseline survey	$n_{app} = 584$, $n_{baseline} = 562$	Completed	Feb. - March 2020
	Enterpreneurship academy	$n = 313$	In process	Feb. - July 2020
	Student screening survey	$n = 926$	Completed	Feb. - March 2020
	Student population survey I		In process	May- June 2020
	Implementation check		Scheduled	July - Aug. 2020
	Midline survey		Planned	Jan. - Feb. 2021
	Endline survey I&II		Planned	July- Aug. 2021 & Jan. - Feb. 2023
	Student population survey II		Planned	Jan. - Feb. 2023
Wave III		9 academies		
	Marketing / information sessions/ short surveys		Planned	Aug.- Sept. 2020
	Application data		Planned	Aug.- Sept. 2020
	Enterpreneurship academy		Planned	Aug. 2020 - Jan. 2021
	Student screening survey		Planned	Sep. - Oct. 2020
	Student population survey I		Planned	Oct. - Dec. 2020
	Implementation check		Planned	Jan. - Feb. 2021
	Midline survey		Planned	July - Aug. 2021
	Endline survey I&II		Planned	Jan. - Feb. 2022 & Jan. - Feb. 2023
	Student population survey II		Planned	Jan. - Feb. 2023

Table 3.2: Research project timeline. *Notes.* The midline survey of Wave I is scheduled for September and October 2020 research funds that have been secured in June 2020 are being disbursed. Wave III is planned for the fall semester 2020/2021. Due to Covid-19, Wave III may have to be postponed to the spring semester of 2021. All following dates will be postponed by around six months in this case. In we are unable to implement Wave III, endline survey II will be conducted in the summer of 2022 for Wave I and II.

worst possible case, we may not be able to implement Wave III at all. Although we deem this highly unlikely, we are conservative in the statistical power calculations below and account for a worst-case scenario with only the two already implemented waves and a base-case scenario with all three planned waves. The Covid-19 crisis will not have any effect on the scheduled data collection as only the endline survey will be conducted as an in-person survey, all other surveys are conducted via telephone.

3.2.5 Treatment assignment and statistical power

Selection experiment

Each information session presenter was provided with a randomly drawn marketing theme — financial independence or creative freedom — for the first session of the day. This was randomly chosen by the research team using a fair coin. The themes for the remaining sessions were then alternated by the presenter.

Entrepreneurship training experiment

The randomization procedure offered admission to the training program to individuals with complete applications. Within *each* training cohort (i.e., university-semester), the target was to offer admission to 40 students, an optimal classroom size determined by SHA.[8] We targeted a control group of equal size; however, the group sizes were constrained by the number of applications received.

Thus the treatment and control group sizes were a function of the number of applicants. Specifically, if there were over 120 applications, we picked 45 students at random and offered admission, assigned 75 to the control group and omitted the remaining students from the study.[9] We anticipated low demand in some training cohorts and chose to over-sample the control group when possible; in case of low demand, having a sufficiently sized treatment group took priority. When we received between 85 and 120 applications, 45 students were randomized into the treatment group, and the rest was assigned to the control group. In case of 80 to 85 applications, we assigned 40 students to control and offered treatment to the remaining ones. Finally, if there were less than 80 applications we offered treatment to $n_T = min[\mathtt{n}_{\mathtt{applications}}, 40]$, and assigned $\mathtt{n}_{\mathtt{applications}} - n_T$ to control.[10]

Having chosen the experimental group sizes, we implemented the following randomization algorithm which stratifies along two dimensions. First, we grouped students according to how many years they had studied their current degree. This is top coded at three years as this is the modal number of years students require to complete a Bachelor degree.[11] The rationale for this is that students who are close to graduation are more likely to move into (self-) employment in the near future. Second, the algorithm ascertains that the share of business students (students who study business, management, finance, marketing or related fields) is balanced between treatment and control *within* each year of study. Students' prior knowledge about business and entrepreneurship concepts may interact with the training content

[8] SHA allowed for deviations from the optimal size within a range of between 30 to 45 students. In case of excess (insufficient) interest, the classes were larger (smaller).

[9] This is done due to capacity and resource constraints. In practice, it is rare to receive over 120 applications for an academy.

[10] Note that the second term can be zero if less than 40 applications are received.

[11] Most applicants are Bachelor students ($\approx$ 87 percent) and those that are not are almost exclusively enrolled in "certificate" and "diploma" programs, which can either be a preparatory or supplementary degree. These usually take two years and can precede or follow a Bachelor degree.

and business students' responses to the training program would systematically differ vis-à-vis non-business students.

We form six cells based on the program of study: business-related (two dimensions: yes or no), and years into the program (three dimensions: one, two or three years). We first use both cells for third-year students, and within each assign an equal number to either treatment or control. This ensures that all applications from third-year students are used.[12] We then applied the same procedure to second-year students. If not all applications from second-year students were necessary to complete target group sizes, we chose a subset at random. Finally, if group sizes were still not exhausted, we included (a random subset of) first-year students.[13] The exact same procedure will be used in Wave III.

Our calculations show both the worst-case scenario, in which we cannot implement the planned third wave at all, and the base-case scenario, in which we proceed with our project as planned or with some delays. To benchmark the statistical power of detecting effects of the training program on business success, we are conservative and present minimum detectable effect sizes based on the actual training cohort sizes from the first two waves of training conducted in the fall of 2019 and the spring of 2020 as the worst-case scenario. We further provide power calculations for various scenarios of attrition and non-compliance given the realized sample size.

During the first two waves we worked with 18 cohorts, meaning 18 university-by-semester blocks. There are 727 and 497 students in the treatment and control groups respectively. This corresponds to an average treatment group size and control group size of 40.4 and 27.6, respectively, and 68 students per cohort in total.

To incorporate myriad factors such as attrition, non-compliance, varying treatment and control group sizes into the power calculations, we perform simulations. We specify a data generating process and set the magnitude of our treatment effect to be equal to a pre-specified percentage of the standard deviation of a generic outcome; this can be interpreted as an effect size in percentage terms. This maps well into our strategy to deal with concerns from testing multiple hypotheses which rests

[12] In theory, it would be possible to receive applications from third-year students in excess of the experimental group sizes. In such cases, we would have randomly picked the respective number. In practice this was never the cases.

[13] As an example, suppose there are 80 third year applicants; 56 in business-related degrees, 24 in non-business related degrees. The procedure allocates 28 of the business students to *each* of treatment and control; similarly, 12 of the non-business students would be in *each* of treatment and control. Overall, there would be 40 students in treatment and 40 in control, but the shares of business and non-business students would be equal across the groups.

on constructing normalized indices of our outcome variables with a mean of zero and a standard deviation of one.[14]

For the simulations, we estimate the primary specification (see Equation (3.1)) in a simulated sample and conduct a two-sided t-test of the null hypothesis of a zero effect of the treatment using standard errors that are robust to heteroskedasticity for inference. For the simulated sample, we set the number of cohorts, rates for attrition, non-compliance, and percent of sample treated as specified in the next paragraph. Then we vary the sample size per cohort starting from four, going until 122 in steps of four.[15] We draw 1,000 simulation samples per sample size considered. Across all simulated samples, we calculate the share of rejected null hypotheses at $\alpha = 0.05$ which is the measure of simulated power.

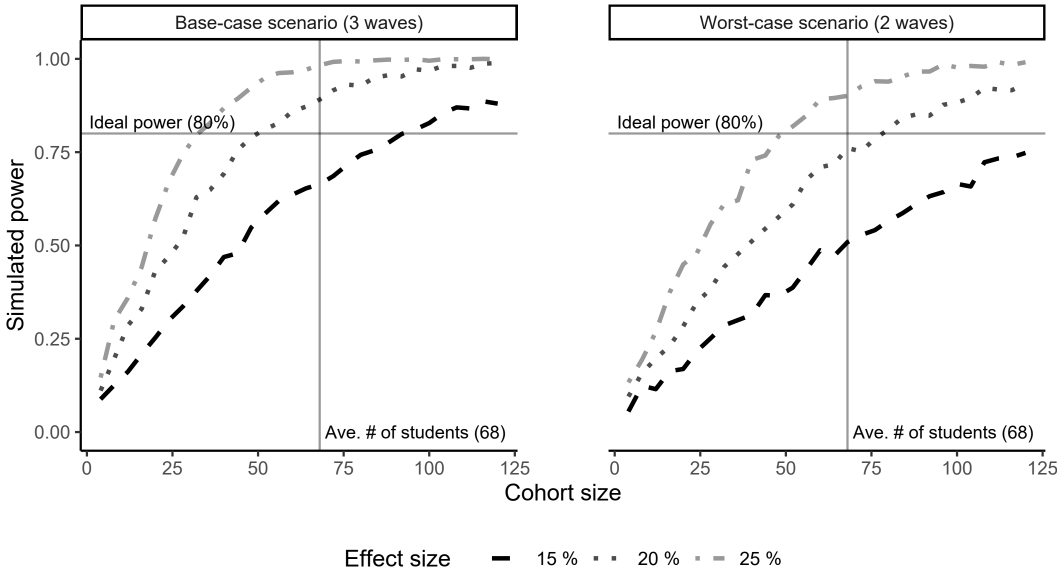
The simulation results are shown in Figure 3.2. Panel (a) presents the base-case scenario based on three waves of academies (left panel) and the worst-case scenario based on the two waves of academies that have been implemented already. We set the following parameters for our benchmark simulations: attrition rate of 5 percent, corresponding to twice the actually observed attrition in the implementation check of the first wave in fall 2019; a non-compliance rate of 25 percent as calculated based on the attendance data for the first wave in fall 2019; and a correlation within training cohorts of 10 percent, corresponding to a generously upward rounded measure from pilot data. The right half of panel (a) indicates that the design is sufficiently powerful (76 percent) to detect an effect of 20 percent (or 0.2 of a standard deviation) even in the worst-case scenario which seems to be a typically observed change (McKenzie and Woodruff, 2014).[16] In the base-case scenario in the left half of panel (a), our design would be well-powered to detect an effect size of 20 percent (89 percent power). If the effect size is actually only 15 percent of our standardized variable, the statistical power of our design reduces to 66 percent.

In panel (b) and (c) of Figure 3.2, we take the worst-case scenario and calculate the power to detect a 20 percent effect considering even more severe scenarios of

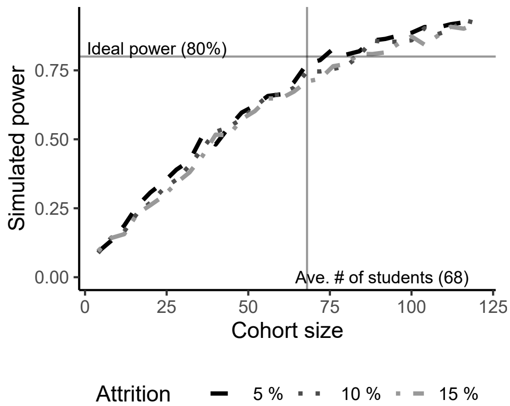
[14] In Section 3.2.3, we detail the procedure. In short, combining several measures into one index measure reduces the number of hypotheses to be tested. Rather than testing one hypothesis per variable, general conclusions are drawn by testing a hypothesis regarding the index.

[15] The lower sample sizes are not realistic, though they help to visualize the trend in power with respect to cohort size.

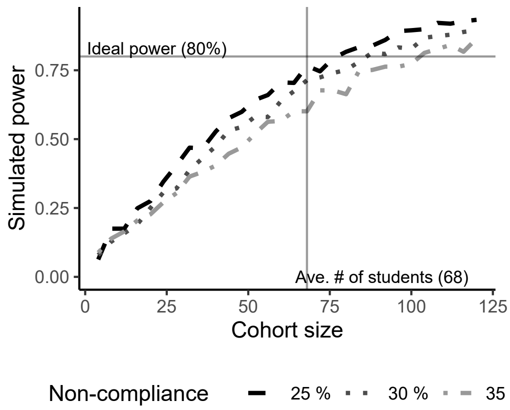
[16] Our study is not only well-powered to detect typical effect sizes, it also improves on existing studies. McKenzie and Woodruff (2014) notes that in most studies the power to detect an increase of 25 or even 50 percent in profits or revenues is well below generally accepted levels of power of above 80 percent.



(a) Effect size in base- and worst-case scenario.



(b) Worst-case scenario + attrition.



(c) Worst-case scenario + non-compliance.

Figure 3.2: Simulated statistical power. *Notes.* The simulations in Panel (a) have the following specifications: attrition rate is five percent, non-compliance is 25 percent, within-cohort correlation is 10 percent and the treatment probability is 59 percent. Statistical power to detect an effect of 15 percent, 20 percent or 25 percent for different average cohort sizes is presented. Cohort size is the sum of treatment and control group individuals. The right hand panel reports the worst-case scenario (two waves) while the left hand panel illustrates calculations for the base-case scenario (three waves). The worst-case simulations vary the attrition rate in Panel (b) and the non-compliance rate in Panel (c) for an effect size of 20 percent.

attrition and non-compliance, holding the other parameters constant.[17] Panel (b) reports that attrition rates of 10 percent and 15 percent would only have a marginal effect on the power of the design. Panel (c) shows that non-compliance rates of 30 percent and 35 percent decrease statistical power to detect an effect of 20 percent to 71 percent and 60 percent, respectively. Overall, our design is well-powered for the base-case scenario with three waves. The conservative, worst-case scenario still yields better power than previous studies despite being below the generally accepted appropriate target of 80 percent power (McKenzie and Woodruff, 2014).

3.3 Data

3.3.1 Data collection and processing

Measuring treatment effects at two levels and describing selection into entrepreneurship requires a multitude of surveys. Figure 3.1 details our data collection efforts, and to which subpopulation surveys are administered. We make all survey instruments available through attachments to the pre-registration in the AEA registry #4502.[18]

Selection into entrepreneurship

The highest level of self-section occurs when individuals select into being interested in entrepreneurship training and attend an information session (see top of the pyramid in Figure 3.1). From this subpopulation, we collect the following data during the information session: pen and paper based *short surveys* eliciting contact details, field of study, measures of cognitive ability using four Raven matrices, student's assessment of how many of these they believed they completed correctly and their assessment of their own entrepreneurial potential on a scale from one through

[17] In results not reported, we can also demonstrate that a correlation of 0.15 within training cohorts has only a negligible effect on the minimum detectable effect.

[18] To ascertain data integrity and safety, and to ensure survey respondents' privacy, we collect, manage and store data in the following way: First, the interview data is collected by experienced local enumerators. Prior to each data collection effort, PIs personally conduct extensive multi-day workshops with the enumerators. Data is collected using Kobo toolbox, and its Android-based mobile device app. Data is stored on secure drives provided by the University of Munich digital infrastructure. When data is collected using pen and paper, data is digitized also using Kobo toolbox in a timely manner and physical records are safely kept at the University of Munich to ensure privacy thereafter.

ten.[19] To reach the non-interested subpopulation we track those classes where the training academy was advertised using short pitches. We classify all students of such a class as having been exposed to marketing. We return to the same classrooms a few weeks later and distribute *student population screening surveys.* These surveys mimic short surveys conducted during information sessions and also elicit students' awareness of entrepreneurship training programs. This allows us to identify students who were aware of the academy based on whether they have heard about our training program or about any entrepreneurship training program at their university.[20] The pool of students who are aware of a training program but did not apply constitutes the sampling frame for the *student population survey.* We then randomly sample 80 students per university, and survey them at two points in time. First, we conduct a phone survey mimicking the *baseline survey* conducted with academy applicants, which allows us to describe predictors of selection into entrepreneurship (*Student Population Survey I*). Second, we repeat this in *Student Population Survey II* to analyze how the subpopulation of non-interested students evolved over time relative to those who expressed in training but were not admitted—the control group. There is no experimental variation at either stage of this comparison.

Selection experiment

Attending information sessions is a necessary requirement for students to be able to apply to the training program since the exogenous variation of the marketing messages in the *selection experiment* is implemented in the information sessions. At the end of an information session, interested students can pick up a paper-based application form. Thus, application form data is only available for the subset of those interested in the training who actually submit a (complete) application form. Application forms contain contact details, demographic information, questions about motivations for and experience with entrepreneurship. We also include questions on students' expected future wage income, as well as expected earnings from entrepreneurship. With the experimental variation of the marketing messages we identify how selection into applying for entrepreneurship training varies with the stressed motives.

[19] A short and standardized illustration on how Raven matrices work in general and how students ought to indicate their answers on the short surveys was provided.

[20] Most universities do not offer alternative entrepreneurship training programs. Thus, it is reasonable to assume that students who were aware of a general academy were aware of *our* academy despite being unable to exactly recall the name of the program.

Entrepreneurship training experiment

To causally identify the effect of being offered entrepreneurship training, admission to the training program is offered on a random basis among those who apply. We gather pre-treatment data by conducting a *baseline survey* prior to individuals being informed about their admissions decisions. After the entrepreneurship training academy, we conduct an *implementation check survey* (around one to two months after the academy ends) and a *midline survey* (around six month later) with the treatment and control groups. Finally, we carry out two *endline surveys*: Endline Survey I will be conducted 12 months after each cohort is finished with their training; Endline Survey II surveys the entire sample around two years after the last round of academies. This survey will be conducted simultaneously for all cohorts and allows us to look at how medium to long-term effects evolve.

While the baseline, implementation check and midline survey are conducted over the phone, the endline surveys will be conducted in person. As detailed below, the surveys elicit information on socio-economic characteristics and main outcome variables, such as prior and ongoing wage and self-employment, preferences measures (risk and time preferences, degree of loss aversion), and non-cognitive traits (Big-5, grit, aspirations and personal initiative). Financial compensation for participation in the endline surveys helps to minimize attrition.

3.3.2 Key outcomes

We use the collected data to construct outcome measures for our five families of hypotheses, as laid out in Section 3.2.3. To test hypotheses we follow the approach by Kling, Liebman, and Katz (2007) and aggregate variables into indices to test each main hypothesis (see Table 3.1) when possible. This reduces the number of tests conducted within each family. For instance, rather than testing for effects across ten business practices, we define an index using adherence to those ten practices and only conduct one hypothesis test. This hypothesis test in turn is part of a family of hypothesis tests. While we focus on indices of outcome measures here to address multiple hypothesis testing, we will also look at individual outcome variables during the analysis. We will clearly mark which results are accounting for multiple hypothesis testing and which are not.

Testing primary Hypothesis Families 1.1 and 2.1 will allow us to draw general conclusions about the entrepreneurship training experiment. Testing hypotheses within Hypothesis Family 1.2 is informative about the mechanisms through which the training program works. Hypotheses 2.2.1 and 2.2.2 set out to analyze dimensions which correlate with entrepreneurial aspirations and success by comparing applicants to the non-interested subpopulation.[21]

To create a summary index from several continuous variables we calculate the unweighted average of those variables' z-scores. Z-scores are constructed using the control group mean and dividing by the control group standard deviation. Thus, each component of the index has a mean of zero and a standard deviation of one for the control group. To create an index of a set of binary variables we calculate their mean; that is, the fraction of "successes" across all component variables. If required, variables that are used to construct an index are reversed so that meaning is consistent.[22] In Appendix 4.6 we describe which variables are used to construct the indices in Table 3.1. The pre-analysis plan details the construction of the specific indices.

Hypothesis Family 1.1 consists of four indices: i) business creation (extensive margin), ii) business success (revenue, profits), iii) labor (employees) and capital (assets, inventory) input, and iv), an index of economic self-sufficiency which aggregates earnings from self-employment, wage employment and other sources.

Hypotheses Family 1.2 consists of six primary indices: i) business practices (we draw on an abbreviated version of the 22-item questionnaire used in McKenzie and Woodruff (2016), and retain ten elements of the original questionnaire (see Appendix 4.6), ii) financial professionalization (contains among others, knowledge and usage of financing instruments, indicators of business registration and licensing), iii) marketing practices, iv) capacity to innovate, v) business networks, and vi) development of an "entrepreneurial mindeset" (a composite index constructed from measures of personal initiative, aspirations and entrepreneurial future and self-efficacy (Frese, Krauss, Keith, Escher, Grabarkiewicz, Luneng, Heers, Unger, and Friedrich, 2007; Campos, Frese, Goldstein, Iacovone, Johnson, McKenzie, and Mensmann, 2017; Bernard and Taffesse, 2014; Streicher, Rosendahl Huber, Moberg, Jørgensen, and Redford, 2019)). The last two Hypothesis Families contain non-cogntive traits, such as the Big-5 personality traits or grit, time and risk preferences, as well as one's degree of loss aversion (Rammstedt and John, 2007; Duckworth and Quinn, 2009;

[21] We can compare the non-interested subpopulation to the full set of applicants using baseline data (pre-intervention). Using endline data, we compare the non-interested subpopulation to the control group (post-intervention).

[22] For example, all variables used to create the "Innovation" index are arranged so that a larger number indicates more innovative.

Falk, Becker, Dohmen, Enke, Huffman, and Sunde, 2018; Fehr and Goette, 2007). In these Families, we create indices where there is a natural grouping (e.g., risk and subjective risk preferences), and investigate sub-indices in other cases (e.g., Big-5 indices).

Hypotheses Family 2.1 is the essence of the selection study and consists of four hypotheses: i) the relative effectiveness of the two randomly chosen marketing messages in terms of attracting applications, ii) whether applicants differ in their cognitive ability (proxied by performance on Raven matrices), iii) whether applicants exhibit differences in over-confidence, and iv) whether applicants self-assess their entrepreneurial potential differently. We construct a measure of over-confidence by comparing individuals' observed and subjective (self-reported) performance on the Raven matrices (Åstebro, Herz, Nanda, and Weber, 2014; Moore and Healy, 2008).

There are two families of hypotheses which we use to study correlates of entrepreneurial aspirations and success in the wider population, Families 2.1.1 and 2.1.2. They mirror the hypotheses from Families 1.1 and 1.2 and therefore mimic the baseline and endline. These two families of hypotheses describe patterns through which students select into being interested in entrepreneurship training. We non-experimentally study baseline and endline differences between students who were interested in entrepreneurship training, and students who were not. First, the baseline comparison sheds light on how the subpopulation that applied to the training program differs from the general student population at large. Second, by comparing those that did not express interest (*non-interested subpopulation*) to interested students who were not offered admission to the training (control group) at endline, we can observe how those groups evolved over time.

3.3.3 Variation from intended sample size

The final sample size depends on the number of applicants and their response rate. To ensure that potentially interested students know about the academy and come to *information sessions*, we closely monitor the marketing campaign. To reduce attrition (i.e., non response) of the applicants over time, we conduct multiple rounds of follow-up surveys to establish frequent contact and trust.

Changing phone numbers represent the highest threat to maintaining contact with the surveyees. Therefore, in addition to students' own phone number(s), we inquire into contact details from a next-of-kin, their classroom coordinator and ask for an email address. In subsequent surveys, respondents are asked to verify or update this information. We achieved a response rate of 97.1 percent in the implementation check of the first wave.

We may use social media groups of the academies as an additional source of information in the future.[23] If those groups retain additional information, attrition could be treatment-specific. We try to account for this by documenting whether the data used to contact surveyees would have been available for both treatment and control. Further, when testing whether attrition is treatment-specific we will be conservative and test for attrition using 10 percent as threshold for statistical significance.

Should treatment status predict attrition, we will additionally provide treatment effect bounds using two approaches recently proposed in the literature. First, the procedure proposed by Lee (2009) quantifies the distribution of those who were induced to "staying in the sample" by treatment and estimates the best and worst-case scenarios. Second, we construct treatment effect bounds using the method suggested by Behaghel, Crépon, Gurgand, and Le Barbanchon (2015). This approach uses the number of attempts (e.g., phone calls) made to reach a person as instrument in a Heckman-type selection model.

3.3.4 Randomization balance

At this point, implementation of the intervention of the first two waves is completed. We have data available for all participants of the information session where we implemented the selection experiment using randomly chosen marketing messages across both waves. Additionally, we have collected baseline data from applicants and randomized admission offers across both waves (see Figure 3.1).

In Table 3.3, we conduct balance checks using the baseline data and compare those individuals who were offered admission (treatment) to those who were not (control) in the *entrepreneurship training experiment*. Columns 1 and 3 report the *unconditional* means for the treatment and control group. In column 5, we regress the respective variable on a treatment indicator, controlling for training cohort fixed-effects, and report the estimated treatment effect (*regression-adjusted difference*). Using heteroskedasticity robust standard errors, we then conduct a two-sided t-test of whether the treatment effect is equal to zero, and report the p-value in column 6. Overall, Table 3.3 suggests that randomization was successful; from 49

[23] Trainers typically create WhatsApp groups to stay in touch with their class members, share materials, and give updates about scheduling and locations.

tests we conduct, only one difference is statistically significant at the five percent level. Specifically, treatment subjects report higher time preference scores which we attribute to random sampling variation.

For the *selection experiment*, we only have the short-surveys of participants at the information sessions as baseline data. The elicited characteristics (gender, field and year of study) were balanced across both randomly assigned marketing themes.

	Treatment		Control		Reg. Adj.		
	Mean	St. dev.	Mean	St. dev.	Diff.	p-value	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
General							
Profit marketing theme (d)	0.44	0.50	0.41	0.41	0.01	0.83	1215
Male student (d)	0.52	0.50	0.58	0.58	−0.02	0.48	1215
Employment							
Working for a wage during the semester (d)	0.10	0.30	0.07	0.07	0.02	0.21	1214
Employer is company (d)	0.04	0.20	0.03	0.03	0.01	0.33	1214
Compensation per month in UGX (ths)	39.37	184.51	35.40	35.40	6.35	0.65	1214
Hours per week working	3.71	13.88	2.42	2.42	1.01	0.19	1214
Business							
Ever owned a business (d)	0.28	0.45	0.25	0.25	0.01	0.73	1214
Currently owning a business (d)	0.22	0.41	0.19	0.19	0.01	0.62	1214
Founder/Co-founder of business (d)	0.26	0.44	0.24	0.24	0.00	0.98	1214
Number of partners in business [*]	0.38	1.82	0.29	0.29	0.10	0.39	547
Business officially registered (d) [*]	0.02	0.14	0.02	0.02	0.00	0.71	546
Business has local trading license (d) [*]	0.03	0.16	0.05	0.05	−0.02	0.13	536
Length of existence of business	0.57	1.74	0.53	0.53	0.04	0.74	1214
Length of work at business in months	0.56	1.72	0.48	0.48	0.07	0.53	1214
Number of full-time employees	0.71	11.25	0.40	0.40	0.48	0.43	1214
Number of part-time employees	0.20	1.20	0.20	0.20	−0.01	0.93	1214
Hours per week working at business	7.71	18.58	6.12	6.12	0.19	0.85	1209
Profit per month at business in UGX (ths)	174.32	689.43	147.19	147.19	31.29	0.43	1214
Number of additional businesses owned [*]	0.06	0.27	0.04	0.04	0.01	0.52	547
Networks							
Personal contacts for business advice	0.70	0.46	0.73	0.73	−0.01	0.71	1210
Number of contacts in family and friends	2.74	3.30	2.86	2.86	−0.02	0.91	1206
Number of contacts outside family and friends	0.80	1.83	1.03	1.03	−0.23	0.12	1210
Contacts can help discussing business ideas (d)	0.67	0.47	0.72	0.72	−0.02	0.51	1202
Contacts helped discussing business ideas in the past (d)	0.40	0.49	0.43	0.43	−0.02	0.42	1202
Contacts can help collecting payments (d)	0.38	0.49	0.41	0.41	−0.02	0.48	1142
Contacts helped collecting payments in the past (d)	0.12	0.32	0.13	0.13	−0.01	0.55	1142
Contacts can help with sharing tools, inputs, employees (d)	0.37	0.48	0.39	0.39	−0.03	0.37	1126
Contacts helped with sharing tools, inputs, employees in the past (d)	0.12	0.33	0.13	0.13	0.00	0.82	1126
Contacts can help with purchasing inputs, stocks (d)	0.36	0.48	0.38	0.38	−0.02	0.51	1129
Contacts helped with purchasing inputs, stocks in the past (d)	0.13	0.33	0.12	0.12	0.00	0.84	1129
Funding							
Ever took loan to fund business idea (d)	0.08	0.28	0.08	0.08	0.00	0.97	1213
Number of known funding initiatives (out of 7)	1.38	1.16	1.41	1.41	0.01	0.88	1172
Non-Cognitive							
Big-5: extraversion	0.88	1.44	0.84	0.84	0.13	0.13	1210
Big-5: agreeableness	1.56	1.27	1.44	1.44	0.03	0.75	1212
Big-5: conscientiousness	1.99	1.22	1.90	1.90	0.00	0.95	1211
Big-5: neuroticism	−1.15	1.35	−1.13	−1.13	−0.07	0.40	1212
Big-5: openness	7.55	1.27	7.50	7.50	0.04	0.56	1213
Grit score (1-5)	3.57	0.44	3.58	3.58	−0.03	0.32	1203
Personal initiative score (1-5)	4.02	0.41	4.01	4.01	0.00	0.99	1208
Stress score (0-16)	6.20	2.25	6.02	6.02	0.11	0.42	1200
Preferences							
Risk preference: scale (1-5)	4.07	0.79	4.06	4.06	−0.03	0.54	1212
Risk preference: final number (1-32)	15.53	11.79	16.09	16.09	0.46	0.52	1213
Loss aversion: Final number (0-6)	4.68	2.02	4.57	4.57	0.16	0.20	1213
Time preference: scale (1-5)	4.03	0.90	3.96	3.96	0.06	0.28	1212
Time preference: final number (1-32)	11.58	12.36	9.79	9.79	1.42	0.05	1213
Entrepreneurial Self-Assessment							
Confidence in ability to start own company (1-5)	4.24	0.66	4.24	4.24	−0.03	0.46	1213
Confidence in ability to pursue self-employed career (1-5)	4.30	0.59	4.22	4.22	0.05	0.16	1213
Confidence in ability to manage challenges of an entrepreneur (1-5)	4.17	0.61	4.17	4.17	−0.03	0.39	1213
Confidence in ability to work in own business one year from now (1-5)	3.90	0.90	3.86	3.86	0.00	0.96	1200

Table 3.3: Balance in entrepreneurship training sample. *Notes.* Columns 1 and 3 report the unconditional mean, columns 2 and 4 the standard deviation for the treatment, who was randomly offered admission to the training program, and control group, respectively. Column 5 reports the regression adjusted mean $\hat{\beta_1}$ estimated using $y_{i,u} = \beta_0 + \beta_1 treat_{i,u} + \alpha_u + \varepsilon_{i,u}$ where α_u is training-cohort fixed effect. Column 6 displays the p-value from a two-sided t-test of $\mathtt{H}_0 : \beta_1 = 0$ using heteroskedasticity-robust standard errors. The last column shows the number of non-missing observations. (d) denotes an indicator variable. Variables marked with a [*] are those that were only measured in the second wave.

3.4 Analysis

OLS will be used if the outcome measure is continuous. We will report results from both logit and OLS regressions for binary outcomes, with the logit specification being our preferred. Inference about treatment effects will be based on two-sided t-tests obtained from using (cluster-)robust standard errors. We precisely state how standard errors are calculated when discussing the empirical specifications for estimating treatment effects. We separately discuss the empirical specifications for the entrepreneurship training study and the selection study. The p-values that govern our conclusions will take into account multiple hypotheses testing by being adjusted to control for the family-wise error rate (FWER). We detail the procedure in Section 3.4.4.

3.4.1 Entrepreneurship training experiment

In the *entrepreneurship training experiment*, we identify the Intention-to-Treat (ITT) effect of being offered admission to the entrepreneurship training. We separately estimate the coefficient of interest $\beta_{1,r}$ for short-term ($r = 2$, Endline I) and long-term effects ($r = 3$, Endline II) according to Equation (3.1):

$$y_{i,u,r} = \beta_{0,r} + \beta_{1,r}\texttt{treat}_{i,u} + \alpha_u + \texttt{strata}_{i,u} + \varepsilon_{i,u,r} \tag{3.1}$$

where $y_{i,u,r}$ is outcome (measured by an index) for individual i, training cohort $u \in \{1, ..., K\}$, and survey round r. The indicator variable $\texttt{treat}_{i,u}$ is equal to one if individual (applicant) i in training cohort u was randomly offered admission, and zero otherwise. Since randomization of admission offers was stratified by field of study and year of study, we include an indicator variable for every combination of the two variables.[24] Since the probability of being assigned to treatment differs across training cohorts, and is a function of the number of applicants, we include a training cohort fixed effect α_u.

Equation (3.1) is our preferred specification, and results from it will be reported first in the analysis. Put differently, estimates of β_1 from Equation (3.1) will be used to address the questions and hypotheses posed earlier. The following specifications are intended to provide more precise estimates in order to help us better gauge the

[24] This results in five indicators included in the regressions, with one reference category omitted. These randomization cells refer to every combination of field of study (business and non-business) and year of study (first, second, and third).

magnitude of the estimated effects.

To improve the precision of $\widehat{\beta_1}$ we run a second set of specifications which includes a set of pre-treatment predictors. We follow the recommendation in Duflo, Banerjee, Finkelstein, Katz, Olken, and Sautmann (2020) and use a variable selection approach. The double post-lasso estimation proposed by Belloni, Chernozhukov, and Hansen (2014) selects a low-dimensional set of predictors which are then included in the estimation. The method uses two separate Lasso regressions; one model to predict treatment assignment, another model to predict the outcome, and each model returns a set of variables to be included. Denote the union of this (as of now unknown) set of covariates by $X_{i,u,r=0}$. We further include the baseline value of the dependent variable $y_{i,u,r=0}$ whenever available:

$$y_{i,u,r} = \beta_0 + \beta_{1,r}\texttt{treat}_{i,u} + \beta_2 y_{i,u,r=0} + X'_{i,u,r=0}\gamma + \texttt{strata}_{i,u} + \alpha_u + \varepsilon_{i,u,r} \quad (3.2)$$

McKenzie (2012) discusses the benefits of a design that uses several post-treatment surveys to obtain more precise treatment effect estimates. Variables central to the analysis, such as profits and revenues, are likely to exhibit little auto-correlation. In this setting, statistical power in ANCOVA specifications is increased by pooling post-treatment observations. Section 3.3 describes that we conduct one midline follow up in addition to two endline surveys, resulting in three ($r \in \{1, 2, 3\}$) post-treatment surveys. Pooling those rounds, we estimate:

$$y_{i,u,r} = \delta_r + \beta_1\texttt{treat}_{i,u} + \beta_2 y_{i,u,r=0} + \alpha_u + \varepsilon_{i,u,r} \quad (3.3)$$

where δ_r is a survey round fixed effect, and $r = 0$ indexes the baseline.

Effect heterogeneity

We are interested in analyzing heterogeneity in the ITT-effects along four independent, preregistered dimensions. First, we explore whether effects differ by an individual's field of study. Students in a business-related degree may have a higher ex ante likelihood of starting (successful) businesses due to higher entrepreneurial intentions or a different skill set (e.g., Solesvik, 2013; Bae, Qian, Miao, and Fiet, 2014). Second, we test whether effects differ by an individual's year in their degree. Students closer to graduation are more likely to move into (self-)employment in the near future. Thus, we test whether effects differ between students in their final (third year) and the remaining students. Third, we assess whether effects are different for students who report having sufficient financial means at baseline. Capital constraints have frequently been cited as the major obstacle to business growth in developing countries, and individuals who already possess the required funds may stand to benefit in a more immediate way (McKenzie and Woodruff, 2014). Fourth, we analyze differential effects by gender (Shinnar, Hsu, and Powell, 2014). Additional exploratory heterogeneity analyses (e.g., along self-reported motives and randomly assigned marketing themes, economic preferences or personality traits) will be clearly indicated as such.

Inference

Inference about the estimates in Equations (3.1) and (3.2) will be based on conventional heteroskedastic-robust Eicker-Huber-White standard errors. In case of Equation (3.3) standard errors will be clustered at the individual level since we use up to three observations per individual. Randomization of admission offers occurs at the individual level, and thus these standard errors are appropriate.

3.4.2 Selection into entrepreneurship

We describe and analyze selection at two steps before being (randomly) offered admission to the training program. In the *selection experiment*, random assignment to marketing messages during information sessions provides us with orthogonal variation which we exploit to study selection into applying for the training program along two salient motivations. Specifically, we use the following specification to analyze the differential effect of exposure to a specific marketing messages on a student's propensity to apply (Hypothesis 1 of Hypothesis Family 2.1):

$$\texttt{applied}_{i,u} = \beta_0 + \beta_1 \texttt{treat\_profit}_{i,u} + \alpha_u + W'_{i,u}\delta + \varepsilon_{i,u}. \tag{3.4}$$

Indices are defined as above; `applied` is an indicator equal to one if an individual submits an application for the training program, and zero otherwise; `treat_profit` is an indicator equal to one if an individual participated in an information session randomly emphasizing financial independence, and equal to zero if theme was creative freedom. The vector $W_{i,u}$ is included to increase the precision of estimates and it contains an individual's gender as well as indicators for years in the current degree (defined as above).

Hypotheses 2 through 4 of Family 2.1.1 capture the idea that selection patterns may differ relative to the underlying motivation for entrepreneurship. Denote a dimension of hypothesized heterogeneity in selection (cognitive ability, over-confidence, entrepreneurial self-assessment, see Hypotheses 2 through 4 of Family 2.1.1 in Table 3.1) with Z_i; we then estimate the following specification to test for different selection patterns:

$$\texttt{applied}_{i,t} = \beta_0 + \beta_1 \texttt{treat\_profit}_{i,u} + \beta_2 Z_{i,t} + \gamma Z_{i,t} * \texttt{treat\_profit}_{i,u} + \alpha_u + W_{i,u}\delta + \varepsilon_{i,t}. \tag{3.5}$$

Conclusions about differential selection will be based on assessing whether the estimated coefficients of our heterogeneity analyses are statistically significantly different from zero ($\mathtt{H}_0 : \gamma = 0$).

Effect heterogeneity

We do not anticipate to have sufficient power to study whether effects are heterogenous by individuals' field of study. However, we do intend to conduct exploratory analyses to assess whether the marketing messages induce differences in the composition of business and non-business students. In this case, we will follow Casey, Glennerster, and Miguel (2012) and label the regressions as unregistered and exploratory.

Inference

The *selection experiment* is a clustered design in which all students participating in a given information session are exposed either to the financial independence or the creative freedom marketing message. Thus, standard errors should be clustered at the session level (Abadie, Athey, Imbens, and Wooldridge, 2020); the level at which treatment varies. However, due to administrative issues, for some individuals we are unable to observe the exact session an individual attended and cannot cluster at this appropriate level. We attempt to overcome this by conservatively clustering at the training cohort level which is the next highest level. In the worst-case scenario of two waves, there are only 18 training cohorts and standard cluster-robust inference may over-reject. We thus pursue the wild bootstrap adjustment proposed by Cameron, Gelbach, and Miller (2008) to calculate standard errors and conduct inference.

Non-experimentally describing selection

Finally, we document selection into entrepreneurship by comparing those who were informed about the training program but did not attend an information session (non-interested subpopulation), to those who applied to the training program using baseline data. In addition, we document trends in how the non-interested subpopulation evolves over time relative to the subpopulation that expressed interest in the training. We do so by comparing them to those who applied but were not admitted—the control group—using Endline I ($r = 2$) data. Both comparisons are based on estimating the following specification:

$$\mathtt{y}_{i,u,r} = \beta_0 + \beta_1 \mathtt{applied}_{i,u} + \alpha_u + \varepsilon_{i,u,r} \tag{3.6}$$

Indices are defined as above and we examine them at the baseline ($r = 0$), and again at the Endline I ($r = 2$). $\mathtt{applied}_{i,u}$ is an indicator equal to one if an individual applied to the training, and zero otherwise. There is no experimental variation at this stage and therefore $\widehat{\beta_1}$ does not measure a causal effect, but is merely informative of a correlation. We calculate heteroskedasticty-robust Eicker-Huber-White standard errors. For completeness, we also show results for individual index components.

3.4.3 Data processing

First, to establish that our results, especially those involving monetary outcomes, are not driven by extreme observations, we will report results with and without winsorizing outcomes at the 99^{th} percentile. Should a variable lack a natural lower bound (i.e., revenues are bound at zero, while profits are unbounded), we also winsorize at 1^{st} percentile.

Second, distributions of variables such as revenue and profits are likely be skewed to the right. We apply the inverse hyperbolic sine transformation to this data which is defined as $f(x) = log(x + \sqrt{x^2 + 1})$ (Burbidge, Magee, and Robb, 1988). Note that this transformation is also defined for $x = 0$ and retains the interpretation of the classic linear-logarithmic regression model for all values of x — except for very small values.

Third, in order to limit noise caused by variables with minimal variation, questions for which 95 percent of observations have the same value within the relevant sample will be omitted from the analysis and will not be included in any indicators or hypothesis tests. In the event that omission decisions result in the exclusion of all constituent variables for an indicator, the indicator will be not be calculated. We explicitly exclude variables in Hypothesis 2 of Family 1.2 for "financial professionalization": Indicators, such as equity investment or business registrations are likely to be rare events and are insightful despite having little variation.

Fourth, whenever a survey's skip logic was triggered by a "yes" or "no" answer, we code the subsequent questions in the logical fashion.[25] Note that we account for the fact that people answer "don't know" or "don't want to answer"; We only impute the logical value if an explicit "yes" or "no" answer triggered the skip logic.

Section 3.3.2 describes how we construct indices to reduce the number of hypotheses tests. Note that the index value is missing if there is one or more missing values in the component variables (e.g., if a person answers "don't know" to one of the questions). We address this problem by providing two estimates in addition to the estimate based on the actually observed number of non-missing cases. First, we impute missing values using the mean value for the entire population, and then generate the index. For robustness, we also provide benchmarks for imputing minimum and maximum values for the entire population. Second, we implement an Inverse Probability Weighting (IPW) estimator in which each non-missing index value is weighted by the inverse probability of having data observed (Seaman and White, 2013). We model the incidence of observing an index value using a logit model with complete baseline characteristics (sex, employment status, self-employment; see Table 3.3), and use the predicted probability.

Fifth, in order to compare monetary values across time, we adjust values using Consumer Price Index data published by the Uganda Bureau of Statistics.

3.4.4 Multiple hypotheses testing

We construct several indices within each family of outcomes as detailed in Section 3.3.2 and Appendix 4.6. We employ two approaches to control the FWER, that is, controlling the probability of a false positive within each family. First, we implement the approach used by Aker, Boumnijel, McClelland, and Tierney (2016) who use a traditional Bonferroni-type adjustment but account for correlations across variables used to test hypotheses.[26] Their method nests the classic Bonferroni ad-

[25] For instance, if somebody does not know any entrepreneurs, then the number of friends and family members who are entrepreneurs is zero — although the skip logic would have result in this being a missing value.

[26] In our cases, we employ the correlation between index measures within each family.

justment when outcomes are uncorrelated. Second, we also employ the method outlined by Barsbai, Licuanan, Steinmayr, Tiongson, and Yang (2020) who develop a regression-adjusted version of List, Shaikh, and Xu (2019). This is a bootstrap-based stepwise procedure designed to control the FWER in settings with multiple hypotheses.

Thus, for each hypothesis across our five families we obtain two p-values which control the FWER, on top of standard p-values. The p-values that correct for multiple hypothesis testing are of interest for researchers with no priors on the specific hypotheses we test. Our preferred procedure is the one by Barsbai, Licuanan, Steinmayr, Tiongson, and Yang (2020) and our main conclusions will be based on being able to reject null hypotheses using those p-values. We report p-values using the procedure by Aker, Boumnijel, McClelland, and Tierney (2016) for comprehensiveness.

3.4.5 Test for reporting errors being treatment independent

In business training interventions whose overall effectiveness is — among others — judged through financial outcomes and adherence to "good" management practices, reporting errors may not be independent of treatment assignment. Individuals who have gone through the training program may be better at accurately judging profits and sales. Alternatively, they may intentionally overstate profits (to suggest the training was helpful) or positively report on business practices because they are more likely to know what the "correct" answer is (McKenzie and Woodruff, 2016).

To address this concern, we construct a measure of sales minus profits which should equal costs and thus be weakly larger than zero. Should it be lower than zero, it likely signals a reporting error as costs cannot be negative. We then test whether treatment assignment predicts the incidence and magnitude of observed reporting errors. In a second step, we calculate implied revenue per customer, and compare the implied prices of the goods and services across treatment and control and cross check with market prices.[27]

Conditional on detecting statistically significant treatment differences in reporting errors, we will conduct detailed in-person audits with a randomly selected subset of 100 treatment and 100 control group subjects. The audits will take place shortly

[27] We are aware of the possibility that new businesses may create goods and services of higher quality which command above-market prices. Nonetheless, implied prices should be largely comparable to market prices, assuming they are free of reporting errors.

after the endline data collection in the spirit of McKenzie (2017). We focus on business experience and business performance. This allows us to establish bounds of reporting errors for each of the variables studied (difference between endline self-reports and the audit data, separately by treatment and control groups). We will present the bounded results as robustness checks.

4. Management Practices and Firm Performance Evidence from Spanish Survey Data[1]

Abstract

This paper employs unsupervised learning to estimate management styles as latent objects using survey data collected in 2006 ,and examines how they affected firm performance during the Great Recession. First, we estimate styles using Latent Dirichlet Allocations and describe each firms' choice of styles using a scalar index. Second, we employ independently collected balance sheet data and establish a positive correlation between a style reflecting structured management with performance prior to the crisis. Third, we then show that those same firms were more severely affected by the crisis. Results point to relatively higher holdings of non-liquid assets and lower employee turnover as mechanisms.

[1] This chapter is based on joint work with Florian Englmaier (LMU Munich), Jose E. Galdon-Sanchez (Universidad Pública de Navarra, Spain) and Ricard Gil (Queen's University, Canada).

4.1 Introduction

The study of management has been part of economics almost since day one. Adam Smith prominently discusses in his books *The Wealth of Nations* and *The Theory of Moral Sentiments* various management topics, such as the division and organization of labor, wage setting, incentivizing employees, or interpersonal authority. Yet, rigorous empirical economic research, which documents differences in management and its effect on performance, has only recently become the focus of a growing literature (Ichniowski, Shaw, and Prennushi, 1997; Ann, Ichniowski, and Shaw, 2004; Helper and Henderson, 2014; Bloom, Lemos, Sadun, Scur, and van Reenen, 2014). This literature has shown that management structure and quality as an input of production varies profoundly across countries, across firms within a country and even across plants within the same firm (Bloom, Brynjolfsson, Foster, Jarmin, Patnaik, Saporta-Eksten, and van Reenen, 2019). Differences in management can help explain some of the variation in firm productivity within and across countries as well as over time (Bloom and van Reenen, 2007). Therefore, understanding the consequences of good and bad management practices for firm performance, productivity and income inequality has clear policy implications.

In this paper, we aim to speak to three aspects of studying management in the realm of economics. First, we empirically document what *bundles* of management practices firms adopt. In the second step, we assess whether and how these bundles affect productivity and firm performance. By evaluating the interplay between management bundles and firm performance before and during the Great Recession, we also speak to whether the effect of management is invariate to changing economic environments. In order to address these questions, we combine two independently collected data sources from Spain. First, we employ a firm survey conducted in 2006 which provides extensive information on firms' human resource policies. Second, we match the firms from the survey to a panel (2001-2010) of balance sheet data from *Bureau van Dijk*, a commercial data provider, to obtain measures of productivity and firm performance. The firm survey happened to take place in 2006—just before the Great Financial Crisis—allowing us to study the relationship of management with performance during the expansionary period before 2006 as well as during the ensuing Great Recession.

A challenge for empirical studies of management practices have been the supposed complementarities between individual practices. This leads to to sets of practices being adopted together by firms which complicates assessing specific management practices' impact in isolation (Milgrom and Roberts, 1990, 1995). Our approach to the topic embraces this complementarity and leverages unsupervised machine learning, in particular *Latent Dirichlet Allocations* (LDA), to retrieve low-dimensional latent objects which we term *management styles* from highly dimensional survey data of Spanish manufacturing firms collected in 2006 (Blei, Ng, and Jordan, 2003). Intuitively, the algorithm identifies groups of practices that tend to appear together across firms but whose presence also distinguishes firms from one another. This approach of applying unsupervised machine learning to management data is inspired by the work of Bandiera, Prat, Hansen, and Sadun (2020) who use LDA to classify CEOs according to their usage of time.

In the first step of the analysis, we estimate these latent styles from the survey data on single-plant firms. We focus on single-plant firms for there is an immediate match between the entity that decides on the adoption of practices and its performance. We estimate and define two "pure" styles and describe every firm as a linear combination of these two pure styles. Note that the estimated styles do neither carry natural labels, nor are they ordinal. In order to work towards an interpretation of these abstract styles estimated by LDA, we then compare the single-plant and multi-plant firms. The latter generally exhibit a more structured style of management, which provides us with a benchmark (Bloom, Sadun, and van Reenen, 2012b,a). We document that single-plant firms whose management loads more heavily on abstract *Style 2* are similar to multi-plant firms in terms of management practices they employ, and consequently label this Style 2 "structured". This classification is also consistent with the practices most indicative of the respective style.

In the second step, we combine the survey data with administrative balance sheet data that allows us to relate our measure of management style to firms' performance. We report two key results. First, we find a systematic and significant positive correlation of the structured management style with firm productivity prior to the Great Recession. Second, this correlation turns statistically significantly negative for firms' performance during the Great Recession. These findings would be consistent with an interpretation that structured management helps firms strive in economically benevolent environments, In times of crisis though, more flexible and informal management styles may have an edge as they are more conducive to short-term adjustments. While, in terms of exploring this interpretation, we are somewhat restricted by our data, we document patterns along two margins. First, we document that more structured firms are adjusting their workforce to a lesser degree during the crisis. Second, that prior to the crisis more structured firms hold relatively more fixed assets than more informally run firms, ceteris paribus.

Employing LDA, i.e., unsupervised machine learning, enables us to utilize all available dimensions of the survey data without prior conceptions of what constitutes *good* management. It also allows us to retrieve a simple measure of management style that can be related to performance during times of economic expansion or crisis using conventional methods (OLS).

Even though data science methods are increasingly used in economics—see for instance Currie, Kleven, and Zwiers (2020)—many economists are still uncomfortable with the application of (unsupervised) machine learning tools. This is possibly due to the fact that it can at times be considered atheoretical, and many applications favor short-term predictions over economic content. Moreover, there is an obvious risk of ex-post rationalization of findings through data and story mining. We are acutely aware of this but still believe that settings, such as ours, lend themselves well to the application of these techniques. Applying the algorithm allows us to leverage all available data without us pre-imposing structure on the components of the data. Our results pass key sanity checks in that the retrieved management styles are meaningful; interpretable; not trivially explained by observable firm characterstics (size, sector, region, etc.); and, in line with existing literature, correlate significantly with firm productivity.

From a methodological point of view, it is part of our contribution to show that automated methods applied to firm surveys can be useful in capturing management styles. Leveraging plentifully existing survey datasets and combining them with powerful algorithms allow us to cost-effectively address open questions before starting new and costly—in terms of money, and especially, research time—data collection initiatives.

Our paper contributes to various streams of literature. These are extensive literatures and, therefore, in this section we focus on those that appear, to the best of our knowledge, most directly connected to our contribution. First and foremost, our paper contributes to the literature investigating what management practices work best. Bloom and van Reenen (2007) and all other papers derived from their original work related to the World Management Survey (WMS hereafter) systematically collect information on management practices across firms, document differences across firms, industries and countries, and finally examine their relationship with outcomes. Their work with many other co-authors has studied management practices in manufacturing, the service industry, and even health care to name but a few. Their work has been highly influential because it has shaped a modern view of management practices as being ordered along a uni-dimensional score ("good management"). Bloom, Lemos, Sadun, Scur, and van Reenen (2014) show some associations detailing the role and impact of WMS measures that validate our findings. Not surprisingly, they find higher scores of management practices with multinational companies and their subsidiaries. This is similar to our findings because our management Style 2 (more structure and larger plants) resembles the style typically present in multinationals.

Methodologically speaking, we contribute to an emerging literature using unsupervised machine learning to retrieve meaningful information from highly dimensional data in the spirit of Bandiera, Prat, Hansen, and Sadun (2020). Extant data on firm policies come in the form of highly dimensional surveys with no obvious way of aggregation to a single score. We show that machine learning can be effective in identifying patterns and clusters of management policies across a large number of establishments. Most importantly, the use of machine learning to study management styles allows economists to tackle and advance their knowledge of an old question in economics, that is, the role of complementarities within organizations (Ann, Ichniowski, and Shaw, 2004; Ichniowski, Shaw, and Prennushi, 1997). Yet, Brynjolfsson and Milgrom (2013) describe challenges in the empirical assessment of interdependencies between organizational practices, stating that the opportunities to run designed experiments in firms are "underexploited" in this respect. Unsupervised machine learning allows for complementarities of a large number of management policies, summarizing all information in low-dimensional space with complementarities embedded in each style.

Finally, our paper also contributes to work on the impact of the 2008 financial crisis on firms' management and performance. Almunia, Antras, Lopez-Rodriguez, and Morales (2018) use firm-level Spanish data to investigate changes in export policies of Spanish firms before and after the crisis. They find that those firms hit hardest in their domestic sales are also the firms that increase their exports the most after the crisis. The paper by Aghion, Bloom, Lucking, Sadun, and VanReenen (2020) is the closest paper to ours in that they investigate the optimal organizational form during "bad times". They find that firms that delegated more power from central headquarters to local plant managers prior to the Great Recession outperformed their centralized counterparts in sectors that were most severely affected by the subsequent crisis. Also close to our findings, Yang, Christensen, Bloom, Sadun, and Rivkin (2020) find that CEOs use a wide range of markedly different processes to make strategic decisions; some follow highly formalized, rigorous, and deliberate processes while others heavily rely on instinct and habit. In their analysis, more structured strategy processes are associated with larger firm size and faster employment growth. Our findings align with the results in these two papers in that we find that those firms with a more structured management style outperformed those firms with less structure prior to the crisis but not during the crisis.

4.2 Data

In this paper, we use two distinct sources of data. On the one hand, we measure management policies through a survey administered in 2006 to a sample of 1003 manufacturing plants in Spain. On the other hand, we use independently collected accounting data from *SABI* to measure plant and firm performance.[2] In what follows, we describe the survey and its matching with the SABI data.

4.2.1 Survey data

We estimate the latent structure of management styles using firm survey data collected in Spain in 2006. This survey on human resource (HR) practices was administered to a sample of Spanish manufacturing firms, and is representative of the population of manufacturing plants in Spain with 50 or more employees. The survey was conducted at the establishment level, and collected through computer-assisted personal interviews with the general managers of those plants.[3] The responses from this survey have been used in earlier work although with a different focus and employing different methods (Bayo-Moriones, Galdon-Sanchez, and Martinez-de Morentin, 2013, 2017). Bayo-Moriones, Galdon-Sanchez, and Martinez-de Morentin (2017) discuss sample selection and sampling in more detail and Appendix 1 of said reference details the full questionnaire.

The entire survey contains 1003 observations; 534 single-plant firms (SPFs) and 469 plants that belong to a superior organization. We refer to the latter group as *multi-plant firms* (MPFs). We restrict our analysis in this paper to the sample of single-plant firms. In single-plant firms, the link between management practices and firm performance is *direct* in the sense that no superior entity can interfere with

[2] SABI stands for "Sistema de Análisis de Balances Ibéricos" and a quick translation into English would be "System of Iberian Balance Sheet Analysis"

[3] Throughout the paper we use the terms *plant* and *establishment* interchangeably. Single-plant firms are equivalent to firms that only have one establishment. A multi-plant firms consists of multiple plants or establishments.

Sector (1)	% in sample (2)	% in population (3)
Food, beverages and tobacco	15.5	15.9
Textile industry, wearing apparel, leather and footwear	6.9	8.6
Wood and cork	3.4	2.6
Paper, editing and graphic design	7.0	8.1
Chemical industry	8.0	7.2
Rubber and plastic products	6.7	6.0
Non-metallic mineral products	10.8	9.7
Metallurgy and fabricated mechanical products	15.4	15,4
Machinery and mechanical equipment	7.5	8.0
Electrical, electronic and optical products and equipment	7.1	6.3
Transport equipment	6.0	6.5
Other manufacturing industries	5.7	5.5
TOTAL	100	100

(a) Percentage of firms by sector of activity.

	$50 \leq$ workers < 100 (1)	$100 \leq$ workers < 500 (2)	≥ 500 workers (3)	TOTAL (4)
% in sample	48.4	46.4	5.3	100
% in population	54.2	40.7	5.1	100

(b) Percentage of firms by size.

Table 4.1: Percentage of firms by size. *Notes.* These tables report the sample composition in terms of sector of activity—Panel (a)—and number of employees—Panel (b).

decisions in a potentially unobserved manner. Thus, the unit of analysis is the firm or the establishment which is equivalent under the sample restrictions.

The survey asks plant managers to provide information on a host of administrative information and HR practices. It can be broadly divided into eight dimensions: (i) plant and firm characteristics such as number of employees, and multinational and multi-plant status; (ii) HR's policies for blue-collar workers (demographic information, hiring and promotion processes, on-the-job training, etc.); (iii) compensation policies for blue-collar workers (incentive provision, evaluation criteria, etc.); (iv) workplace organization (hierarchical levels and supervisors' roles); (v) labor conflict and cooperation among blue-collar workers; (vi) governance and authority in the implementation of human resource strategies; (vii) profile of other (white) workers and occupations in the plant; and (viii) plant manager characteristics (education, demographics, skill set, etc.).

We discuss summary statistics of firms in the sample in more detail in Section 4.3.3 when we analyze correlates of firms' management style.

4.2.2 Measuring management practices

The unsupervised algorithm we employ to construct a low-dimensional measure of management style requires categorical data. While the majority of the survey's questions are indeed categorical, answers are not provided on consistent scales. For instance, some question elicit agreement on five-point Likert scales, while other use ten-point scales; some questions are simply binary questions; and again others offer (non)-exclusive categorical answers. To construct the input matrix for the algorithm we thus transform all questions into binary measurements which can be thought as the "smallest common denominator" of answer scales.

In total, we obtain 272 binary variables reflecting all answers to management-related questions in the survey. We convert all types of agreement scales (three-point, five-point, seven-point) into three binary variables: i) an indicator for being to the "left" of the neutral mid-point; ii) an indicator for being at the neutral mid-point; and iii) an indicator for being to the right of the mid-point.[4] Categorical questions are transformed into binaries by generating an indicator for each answer possibility. For instance, a question asks for the number one management priority and offers *cost*, *flexibility*, *innovation*, and *quality* as answers. Our procedure generates four indicator variables which are equal to one if the plant reports the respective number one priority. Finally, there is a set of questions that require the surveyee to report a percentage between zero and 100. We convert the answer into three indicator variables: i) an indicator for the answer being 0 percent; ii) an indicator for the answer being greater than zero but no more 50 percent; iii) an indicator for the answer being larger than 50 percent.

We refer to these 272 binary measurements as the management practices in our survey. Appendix Table 4.B.1 details all the indicators along with the questions they are constructed from, as well as their sample means.

The algorithm requires the input matrix of management practices to only contain complete cases. That is, no management practice ought to be missing in the data. Owing to that restriction, we have to drop 71 plants from the sample. Therefore,

[4] Consider a standard five-point Likert-scale going from strongly disagree, disagree, neither disagree nor agree, agree to strongly agree. "Neiter disagree nor agree" forms the neutral mid-point.

our final sample of plants that we use to estimate management style contains 463 single-plant firms.

4.2.3 Firm performance data

SABI is a database collected by Informa D&B in collaboration with Bureau Van Dijk, both commercial data providers. Informa D&B is a Spanish company that provides online access to a large database of Business, Financial and Marketing information for more than 350 million companies in more than 200 countries. The database contains annual balance sheet information for more than 2 million Spanish firms across all sectors in the Spanish economy.

We searched this extensive set of firms and matched entries to the 1003 manufacturing plants in the survey. We matched our manufacturing plants by firm name, tax ID (*CIF* in Spain), industry and location. This exercise resulted in an unbalanced panel across establishments and years as balance sheet records are not complete. It is important to note that the SABI database does not contain administrative tax data, and therefore not all firms in our sample report their accounting data every year. Furthermore, SABI collects balance sheet data at the *firm-level*, and it would be impossible to assign inputs and outputs to different establishments of a multi-plant firm. This constitutes another reason for why we restrict the sample to single-plant firms.

From the SABI data we primarily employ information on revenue, labor force, and assets to construct productivity.[5] We detail the procedure used to construct a measure of firm productivity in Section 4.4.1. In particular, we measure output using sales; capital input using total assets; and labor input using the number of employees. Appendix Table 4.A.4 provides summary statistics for the input variables used in the TFP estimation.

4.3 Estimating latent management styles

This section describes our use of machine learning to estimate latent management styles using the survey data described in Section 4.2.1. We proceed by first outlining the unsupervised algorithm we use to that effect. Next, we describe the results and analyze correlates of those results.

[5] The number of employees is also elicited in the firm survey. The correlation between both measures is = 0.7.

4.3.1 Latent Dirichlet Allocation

We first briefly describe the algorithm, and the estimation specifications we employ to generate the low-dimensional measure of management style. We then turn to describing the results.

Estimation setup

The goal of the empirical analysis is to retrieve a low-dimensional representation of management practices from the high-dimensional survey data. We argue that there are underlying management *styles* which generate differences in observed management *practices* across firms. In order to construct (econometrically: estimate) these unobserved latent styles from firms' observed behavior, we employ *Latent Dirichlet Allocation* (LDA), an unsupervised learning algorithm which was originally conceived to find *topics* in text data (Blei, Ng, and Jordan, 2003; Erosheva, Fienberg, and Joutard, 2007). Yet, it lends itself to the analysis of categorical data more generally. The seminal analysis of CEO's time allocation by Bandiera, Prat, Hansen, and Sadun (2020) introduced this type of analysis into organizational economics.

LDA is a Bayesian hierarchical factor model. The intuition behind it is most easily explained by using the analogy to text data. Each observation is a snippet of text (in our case, a firm with observed practices). Each snippet of text is a mixture of different topics (analogously, each firm's management is a mixture of styles). In turn, each topic is a mixture of all *words, practices in our case* that appear in the entirety of observed text. Put differently, each topic is a probability distribution across all words, where words that are more strongly associated with a topic carry a higher loading. The analogue in the present situation is that a management style is a probability distribution across all observed practices. Thus, we apply LDA to model latent management styles as distributions over all observed practices, and to model firms' observed configurations of management practices as a mixture of these styles.[6]

The crucial input in the analysis is the number of latent styles to be estimated, which is set by the researcher. We specify two latent styles of management based on the following three reasons. First, unlike traditional cluster analysis, e.g., k-means, LDA does not deterministically assign observations to clusters. Thus, a

[6] From a technical perspective, we estimate the models using Gibbs sampling, a Markov chain Monte Carlo algorithm (Griffiths and Steyvers, 2004). For the Gibbs sampler we specify a *burn in* period of 5,000 iterations; we then implement 10,000 iterations with a thinning parameter of 2,000.

specification with two "pure" styles is able to capture heterogeneity beyond assigning membership to exactly one cluster. Rather, each firm is represented by one weight for each pure style. Second, two latent factors simplify interpretability. As Blei (2012) points out, the ease of interpretation should be taken into account when choosing the parameters of unsupervised learning. Finally, the cross-validation exercise in Appendix Figure 4.A.2 suggests that model fit does not improve markedly when we estimate more latent styles. The at best marginal increase in model fit we obtain through more clusters is unlikely to balance the loss of interpretability.

LDA is a Bayesian technique and, as such, requires priors on both of the Dirichlet distributions. We follow Bandiera, Prat, Hansen, and Sadun (2020) in setting both priors. We place a neutral, uniform prior on the firm-over-style distribution (prior = 1) which would place firms' initial mixture of styles at 50:50. The prior on the style-over-practice distribution promotes sparsity (prior = 0.1). This reflects our conception that styles load heavily on a few rather than a lot of practices since there are likely to be few practices emblematic of a style.

Setting a non-zero prior ensures a non-zero posterior. Thus, the probability distributions we estimate have strictly positive loadings for each element. By virtue of being *probability* distributions, the loadings sum to one and, individually, are strictly smaller than one.

Finally, note that LDA is an *unsupervised* learning algorithm and the estimation procedure does not force the resulting clusters to explain firm performance in any way. In contrast, supervised methods, such as classification trees, regularized regression or neural networks, are usually employed with the goal of using a set of variables to predict the values of a response variable. However, we would like to first understand what groups of management practices firms choose by finding a low-dimensional representation of these practices.

Estimation results

First, we obtain distributions over all practices for both styles. In Appendix Figure 4.A.3 we summarize these distributions but explicitly abstain from attaching any labels to the output. Styles are non-ordinal and hence, for now, we refer to the styles neutrally as *Style 1* and *Style 2*. Panel (a) plots all practices' ordered according to their Style 1 loading. The figure demonstrates that the procedure is indeed able to identify two distinct latent constructs. Practices with lower loadings in Style 1—indicative of a lesser role in style 1—tend to load highly on style 2.

There are also practices that carry high loadings in both styles. This suggests the presence of practices that are employed in conjunction with practices that in turn are emblematic of a specific style. In Panel (b) of Appendix Figure 4.A.3 we plot the practices whose loadings quotient across styles is largest. On the far right—practices with a relatively higher loading in Style 1—the algorithm identifies the absence of evaluation systems as well as a narrow focus on ability and personal interviews in the recruitment process. In Style 2, human resource department decision making and the importance of evaluations for promotions is emphasized. Note that this analysis does not take into account the importance of those feature in the styles, i.e., a kind of baseline probability. Thus, two practices with relatively low loadings in both styles may feature in this description. We return to the practices with highest single style loadings in more detail below.

Second, we can illustrate firms' style distributions. Recall that the two styles' weights are positive and sum to one. Therefore, a firm's style distribution is fully characterized by either style share. We focus on the share of Style 2, which we also refer to as Style 2 intensity. Panel (a) of Appendix Figure 4.A.4 plots the count of firms across the Style 2 continuum. The distribution is bi-modal, and there is a mass of firms that load highly on Style 1. A second mass point is between 0.5 and 0.6, pointing to firms that tend to be rather balanced mixtures of both pure styles. In the analysis, we provide results based on a continuous measure of Style 2 intensity as well as based on indicator variables for tertiles.

4.3.2 Characterizing firms' management styles

Since latent management styles are not ordinal, any labels we may want to attach to these styles are necessarily subjective. We pursue two approaches in order to understand what these latent constructs actually capture.

First, we begin to understand what those styles mean by comparing firms of a certain configuration to a separate set of firms whose management we can characterize *a priori*—that is, without relying on LDA. To this effect, we consider multi-plant firms. These firms can benefit from economies of scale, and may be forced to delegate decision rights across space, leading them to employ more structured management practices (Bloom, Sadun, and van Reenen, 2012b,a). Thus, we seek to describe the management styles of single-plant firms by comparing them to multi-plant firms based on style 2 intensity. An additional advantage of this approach is that it does not require a subjective evaluation of the style-over-practice distribution.[7]

We operationalize this comparison by first pooling the surveys of single-plant and multi-plant firms, and then estimate management styles in this joint sample using the LDA procedure exactly as described above.[8] This estimation returns style shares for each firm in the pooled sample, and we plot and compare the Style 2 intensity of three types of firms defined as follow:[9] i) multi-plant firms (which do not appear in the single-plant sample); ii) single-plant firms whose observed intensity of Style 2 in the *single-plant sample estimation* is (weakly) smaller than 0.5, i.e., those that we would describe as rather Style 1 firms; and iii) single-plant firms with an observed intensity of above 0.5, i.e, those that we would describe as rather Style 2 firms.

Figure 4.1 plots the result of this exercise. First, we note that the distribution of MPFs puts most mass above 0.5. Secondly, SPFs with a higher Style 2 intensity from the single-plant sample also put most mass above 0.5 in the joint estimation. Finally, SPFs with low Style 2 intensity in the single-plant sample below 0.5 behave the opposite way. In line with prior findings in the literature, this would suggest that single-plant Style 2 firms employ a more structured management style as their management style looks more like the one of multi-plant firms.

Second, we analyze those practices that carry the highest loadings in each style. Table 4.2 reports the five practices with the highest loading in each style. Style 2 exhibits practices that suggest structured management, emphasizing the role of dedicated human resource departments. In Appendix Figure 4.A.3(b) we plot those practices whose loadings' quotient in both styles is largest; that is, those with the highest relative loadings in both styles, respectively. This corroborates the notion that Style 2 is exemplified by structured practices, while Style 2 mirrors informal

[7] The second approach to understanding the pure style is by evaluating the style-over-practice distributions which we do below. This is more prone to researchers' imposing their conceptions of what styles *ought* to mean. By comparing styles without attaching labels, we attempt to generate an unbiased understanding of what pure styles represent.

[8] In order to carry out this exercise, we drop 20 practice indicators from the multi-plant survey as they are about autonomy from the superior organization and hence only relevant for MPFs. There is no guarantee that the two resulting pure management styles are comparable to the results obtained from using only the single-plant firms. The estimation in the joint sample proceeds exactly as the one in the single-plant sample. Equivalent Dirichlet priors are employed, and the MCMC parameters are kept constant.

[9] Equivalently, we could have plotted the Style 1 share as well. This would have not affected the conclusions we draw in the following paragraph. These styles are unrelated to the styles estimated in the single-plant sample only. Estimating styles in the joint sample only serves to help understanding the meaning of styles in the single-plant sample.

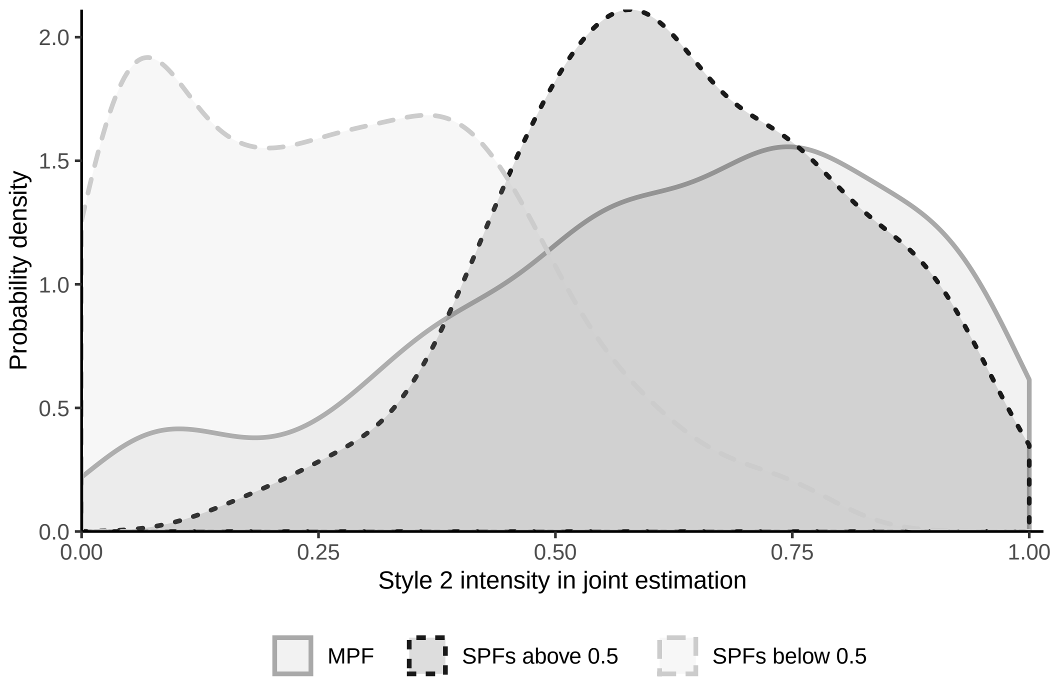


Figure 4.1: Understanding styles by comparing single- and multi-plant firms. *Notes.* In this figure, we apply the LDA procedure described above to estimate management styles in a pooled sample of single- and multi-plant firms (n=871). We then plot the probability density of the corresponding Style 2 share separately for i) those single-plant firms that exhibited a Style 2 intensity of (weakly) below 0.5 when styles are estimated in the *single-plant sample only*, ii) those single-plant firms with a corresponding intensity of above 0.5, and iii) all multi-plant firms.

practices. While we would like to emphasize that any label is subjective, we still conclude that Style 2 captures a more *structured* approach to management, and Style 1 represents a more *informal* approach.

4.3.3 Correlates of management styles

In this section we explore survey data correlates of firms that exhibit high Style 2 intensities and show that management styles are not trivially explained by observables. Recall from the previous discussion that firms with higher Style 2 intensities implement management practices that look more like those of multi-plant firms, stressing more structured forms of management. We denote firm i's Style 2 intensity by γ_i^2 and estimate:

$$\gamma_i^2 = \beta_0 + X_i\beta + \varepsilon_i \tag{4.1}$$

Rank	Style 1	Style 2
1	Recruitment with personal interviews	Dedicated HR department
2	Firm uses no evaluation system	HR part of management team
3	White-collar recruitment through interviews	HR executed administrative tasks
4	% white-collar in management < 50%	% white-collar in intermediate management < 50%
5	% of jobs characterized as manual > 50%	HR reports to plant-director

Table 4.2: Five practices with highest loading in each style. *Notes.* This table lists the five practices with the highest loadings in each style. These are obtained by sorting the respective style-over-practice distribution by practices in descending order of their loading.

X_i captures firm characteristics, such as size, export dependency, or a firm's position along the value chain.[10] We provide both, results from univariate and multivariate specifications. The latter takes into account the correlation structure across firm characteristics. Inference is based on standard errors clustered at the three-digit industry level (at most 78 clusters).

In Appendix Table 4.A.1 we provide summary statistics for those variables which we study in this analysis. The average firm has 116 employees although the distribution is highly skewed to the right (skewness ≈ 5). Further, the average firm has sales of about €28,639,000 worth of goods and services (also skewed to the right; skewness ≈ 7). Firms report sales selectively and only 289 firms report sales in the survey.[11] The modal firm produces a consumer good, while the remaining firms are equally split between intermediate and capital goods. Two thirds of firms are in shared ownership, while a quarter is a limited liability company.

The results suggest that both, the number of employees and sales, are positively correlated with Style 2 intensity. In Appendix Figure 4.A.5 we zoom in on the (univariate) relationship between Style 2 intensity and firms' number of employees. A positive correlation is clearly visible; however, across the support of firms' number of employees, there is variation in Style 2 intensity that cannot be explained be firm size. The positive association of Style 2 intensity and export dependency is to be viewed similarly. Firms that produce consumer goods tend to have lower Style 2 intensity, even after controlling for firm size. On average, a firm producing

[10] Almunia, Antras, Lopez-Rodriguez, and Morales (2018) document that firms at different positions in the value chain had different experiences (and margins of adjustment) during the Great Recession. Hence we control for this position in our analysis

[11] Reporting sales in the survey is not systematically correlated with Style 2 intensity. A linear regression of an indicator for having reported sales on Style 2 intensity results in a coefficient of 0.004 (SE = .09). Controlling for firm size does not alter this conclusion; in fact, firm size measured by the number of employees is not correlated with the incidence of reporting sales either.

	Dependent variable: Style 2 intensity							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log # employees	.17						.2	.17
	(.017)						(.024)	(.017)
Log sales ['000 EUR]		.037					-.01	
		(.01)					(.0088)	
Year plant opened			-.00055				.00075	.00038
			(.00047)				(.00051)	(.00046)
% for export				.0013			.00077	.00057
				(.00045)			(.00044)	(.00038)
Produces consumer good					-.08		-.049	-.088
					(.025)		(.026)	(.024)
Produces intermediate good					-.03		.013	-.023
					(.031)		(.034)	(.031)
Shared ownership						-.0096	-.0061	-.028
						(.052)	(.057)	(.039)
Limited liability						-.077	-.043	-.052
						(.055)	(.056)	(.045)
Adj R-sq	.17	.04	.00064	.02	.013	.01	.22	.2
N. of cases	463	289	456	438	458	463	284	430

Table 4.3: Correlates of Style 2 intensity. *Notes.* This table reports results from OLS regressions where the dependent variable is a firm's Style 2 intensity, a variable between zero and one. "Log" refers to the natural logarithm. "% for export" is a firm's self-reported share of output that is exported abroad. "Produces consumer/intermediate good" are indicator variables equal to one when the firm produces the respective output category, and zero otherwise. The omitted category for this class of indicators is producing a "capital" good. "Shared ownership" and "limited liability" are indicators equal to one when a firm is organized according to the respective ownership structure. The omitted category for this class of indicators is "other" ownership structures. Standard errors clustered at the three-digit industry level are reported in parentheses.

consumer goods has about eight to nine percentage points lower Style 2 intensity. Appendix Figure 4.A.6 zooms in on this aspect, and graphically displays lower Style 2 intensity in the consumption good sector. While medians differ, there is ample common support across different locations in the value chain. Finally, there is no discernible effect of ownership status on Style 2 intensity.

The results in Table 4.3 are obtained from regressions without region nor sector fixed effects. Explanatory power only marginally increases if those fixed effects are included. When region or sector fixed effects are added to the regression, the adjusted R^2 in the analogous specification to column 7 increases to 0.24 or 0.23, respectively.

When they are jointly included, the adjusted R^2 remains at 0.24.[12] An interesting implication is that there is significant variation in styles within economic sectors.

Thus, overall firm characteristics as elicited in the survey can explain about one quarter of variation in management Style 2 intensity. We note a significant positive association between firm size (employees, sales) and Style 2 but these characteristics do not exhaustively explain variation in Style 2 intensity.

4.4 Management style and firm performance before the Great Recession

This section establishes that the management styles we estimated in the previous section correlate with firms' performance in the period before the Great Recession. We measure firm performance using the SABI data we describe in Section 4.2.3, which was collected independently of the firm survey data. This mimics the approach by Bloom and van Reenen (2007) who refer to this as the *two-step procedure* because it first estimates firm-level Total Factor Productivity (TFP), and then projects it onto the space of management styles.

4.4.1 Estimating firms' TFP

We measure firms' performance using TFP which can be interpreted as a firm's technology to combine labor and capital into output. First, we postulate that firms produce output Y using labor (L), capital (K), and a production technology α according to a standard Cobb-Douglas production function $Y = \alpha L^{\beta_1} K^{\beta_2}$. β denote the production elasticity with respect to labor and capital, respectively. By taking the natural logarithm we obtain the following equation where i indexes a firm and t indexes years:

$$y_{it} = \alpha_i + \beta_1 L_{it} + \beta_2 K_{it} + \varepsilon_{it}. \tag{4.2}$$

We use sales in Euro to proxy output, total assets to measure capital input, and the number of employees to measure labor input, and estimate Equation (4.2) using

[12] In Appendix Figure 4.A.7 we show a Style 2 breakdown by sectors and region. The boxplot in panel (a) shows that there is common support across all sectors; that is, the median and interquartile range of Style 2 intensity is comparable across sectors. The map in panel (b) shows regional heterogeneity but comparable means across most regions.

OLS. The underlying, *unbalanced*, panel covers the years 2001-2006.[13]

We obtain a firm's TFP using the predicted value of α_i from Equation (4.2). Appendix Figure 4.A.8 shows the distribution of the estimated $\hat{\alpha}_i$ which is slightly skewed to the right. More importantly, we observe several extreme values indicating relatively (un)productive firms. We account for these in the regression by 95% winsorizing TFP—indicated by the vertical lines in Appendix Figure 4.A.8.

4.4.2 Results

In this section, we provide evidence that the management style we estimated using firm survey data correlates with firms' TFP. To this effect, we estimate

$$\widehat{\alpha_{i,s,r}} = \beta_0 + \beta_1\gamma^2_{i,r,s} + \omega_r + \omega_s + X_{r,s,t}\beta + \varepsilon_{i,r,s} \tag{4.3}$$

where i indexes a firm located in region r which is active in sector s. $\gamma^2_{i,r,s}$ denotes a firm's management Style 2 intensity, which is a value between 0 and 1. Higher values indicate a stronger Style 2 intensity. The ωs absorb time-invariant variation induced by regions and sectors. In $X_{i,r,s}$, we control for firms' location along the value chain by including indicators for producing consumer goods or equipment—producing capital goods is the omitted category. We cluster standard errors at the three-digit industry level.

We provide the results of estimating Equation (4.3) in Table 4.4. Columns 1 and 4 provide simple univariate correlation of Style 2 intensity with firms' TFP. In columns 2 and 5, we add region and sector fixed effects. In columns 3 and 6, we additionally control for value chain location. Columns 1-3 show results based on 95% winsorizing the dependent variable; the remaining columns use non-winsorized outcomes.

Across all columns, there is a significant correlation between Style 2 intensity and firms' TFP. The estimates' magnitude does not change when correlates and fixed effects are included. Unsurprisingly, standard errors are smaller when winsorized data is used on the left-hand side. Style 2 intensity is able to explain about two percent of variation in the dependent variable. The full specification explains 12% of variation. We note that firms producing intermediate goods are more productive

[13] A total of 446 firms enter the productivity estimation, and the average firm appears 5.5 out of 6 times. 331 firms appear in each year. 11 firms only appear once. We estimate the output elasticities of labor and capital to be 0.3 and 0.49, respectively. Below we discuss a robustness check only focusing on full-panel firms. For these firms we estimate elasticities of 0.46 and 0.44, respectively.

	Firm productivity 2001 to 2006 95% winsorized			Firm productivity 2001 to 2006 not winsorized		
	(1)	(2)	(3)	(4)	(5)	(6)
Mgt style 2	.28	.26	.26	.27	.25	.25
	(.079)	(.082)	(.082)	(.091)	(.098)	(.098)
1[consumer good]			.073			.1
			(.066)			(.081)
1[intermediate good]			.13			.16
			(.063)			(.072)
Sector FE	No	Yes	Yes	No	Yes	Yes
Region FE	No	Yes	Yes	No	Yes	Yes
Adj R-squared	.023	.13	.14	.017	.11	.12
N. of cases	385	385	379	385	385	379

Table 4.4: Management style and firms' TFP before the crisis. *Notes.* This table reports the results of estimating Equation (4.4) using OLS. The dependent variable is a firm's estimated TFP; 95% winsorized in columns 1-3 and non-winsorized in columns 4-6. "Mgt Style 2" is a firm's Style 2 intensity. "1[consumer good]" and "1[intermediate good]" are indicators for firms that are located in the respective location along the value chain. The omitted category is firms producing capital goods. Columns 2,3,5 and 6 contain sector and region fixed effects. Standard errors clustered at the three-digit industry level are reported in parentheses.

than firms producing consumer goods; which in turn are more productive than firms producing capital goods. As pointed out above, the full panel structure of production inputs and output is not available for all years. In Appendix Table 4.A.2 we provide results for the same specifications only using those firms where the full panel is available. The estimates are highly comparable to those in Table 4.4.

The magnitude of the correlation of about 0.25 corresponds to a one unit change in Style 2 intensity. To put the magnitude in perspective consider a one standard deviation change in Style 2 intensity ($\sigma(\gamma^2) = 0.25$); this corresponds to a 0.0625 change in TFP. This in turn is equivalent to an effect of 13% of a standard deviation in TFP ($\sigma(\hat{alpha}) = 0.47$). Alternatively, the inter-quartile-range in Style 2 of 0.37

results in a 0.0925 change in TFP; or 20% of a standard deviation in TFP.[14]

To sum up, we find a positive association between Style 2 intensity and productivity prior to the Great Recession. That is, more structured management correlates positively with firms' TFP. The Spanish economy was booming prior to 2006 and firms may have been able to benefit from leveraging economies of scale. A structured management style appears to allow firms to more effectively exploit this beneficial economic environment.

4.5 Management Style and firm performance during the Great Recession

In this section we shine a light on how management style intensity correlates with firms' performance during the Great Recession (2007-2010) that followed the Great Financial Crisis that struck in 2007. Spain's experience in the aftermath of the crisis was markedly different from the, say US or Germany, where after a severe contraction in the short-run, growth rates quickly recovered. As illustrated in Appendix Figure 4.A.9, Spain's economy contracted initially at a comparable rate but its resurgence was slower with GDP starting to grow only in 2011.

4.5.1 Setup

Firms' TFP is derived from estimating a specification akin to Equation (4.2) but now using data for the years 2007-2010. We summarize TFP for the period 2007-2010 in Figure 4.A.10. Panel (a) shows a histogram; the distribution looks comparable to the pre-crisis distribution but points to a number of outliers on the right of the distribution. In the analysis, we account for these again by showing estimates based on 95 percent winsorization. In panel (b), we plot the change in TFP between the two periods (2007-2010 vs 2001-2006) relative to the pre-period (2001-2006). The figure suggests a negative relationship which could indicate regression to the mean—highly productive firms in the pre-period see a mechanical decline in the post-period. We account for this by controlling for the pre-period level of TFP in the regressions.

[14] In Appendix Table 4.A.3 we provide additional results in which we bin management Style 2 intensity into tertiles. In the estimation sample of columns 1,2,4 and 5 of Appendix Table 4.A.3, 130 firms' management Style 2 intensity is smaller or equal to $\frac{1}{3}$; 182 firms' Style 2 intensity is larger than $\frac{1}{3}$ but no larger than $\frac{2}{3}$; finally, 73 firms' Style 2 intensity is larger than $\frac{2}{3}$. We show that the firms in the middle tercile of Style 2 intensity are marginally more productive than firms in the bottom tercile. Firms in the top tercile are significantly more productive than firms in the bottom tercile. Finally, we provide p-values for the comparison of firms in the middle and top tercile; we are unable to statistically reject that the effects are in fact equal.

In the second step we relate firms' productivity in the years 2007-2010 to their management Style 2 intensity, a set of time-invariant controls, and their pre-crisis level of TFP. In a set of robustness checks, we also provide estimates for the effect of Style 2 intensity on the *difference* in TFP across both periods (before vs during the crisis). As in Section 4.4.2, standard errors are clustered at the industry level. In additional results, we show how a censoring approach to missing information due to potentially endogenous firm exit affects the estimates of firm performance during the crisis.

Finally, we turn our focus to channels through which Style 2 intensity affects firm performance during the crisis. While we are unable to authoritatively point to a specific mechanism, we provide a set of results that suggest that a higher Style 2 intensity leads firms to hold fewer non-fixed assets, and to turn over employees at a lower rate—holding constant a wide set of firm characteristics.

4.5.2 Results

In Table 4.5 we provide estimates for the conditional correlation of Style 2 intensity and firm performance during the Great Recession 2007-2010. In panel (a) columns 1-3, we see a statistically significant (at 5 percent) negative coefficient, suggesting that firms with higher style intensity fared worse during the crisis. In columns 4-6, we provide estimates without winsorizing the dependent variable, and see that estimates are comparable in magnitude but less precisely estimated. We consider columns 1-3 to be our preferred specification as Figure 4.A.10(a) points to the presence of extreme values in TFP.

In panel (b) of Table 4.5, we provide results for the case of splitting Style 2 intensity into tertiles. The point estimates suggest—and the p-values in the table's legend confirm—that the effect is predominantly driven by firms in the top tertile, that is those with the highest Style 2 intensity. The effect of -0.12 in column 3 implies that, on average, the TFP of firms in the top tercile of Style 2 intensity is about a third of standard deviation lower ($\sigma_{TFP_{'07-'09'}} = 0.42$) than those of firms in the bottom tertile.

One may be worried that this reflects regression to the mean, in the sense that firms that did better before the crisis do relatively worse now, and vice versa. Indeed, Figure 4.A.10(b) suggests a comparable relationship in the bivariate reduced form. We account for this phenomenon by controlling for firms' TFP in the year 2001-2006—"pre-crisis TFP" in the tables. Thus we are able to interpret the effect of Style

2 intensity on TFP during crisis holding constant pre-crisis TFP. Put differently, in a scenario of two firms with equivalent pre-crisis TFP, the firm with higher Style 2 intensity does worse during the crisis on average.[15]

Another concern one may have is that the most unproductive, high Style 1 intensity, firms had to exit the market during the Great Recession. Thus, what we observe in this period is the set of all firms with high Style 2 intensity, and the subset of relatively more productive firms with high Style 1 intensity.[16] We address this concern in two ways and show that it cannot explain our results. The approach we take is similar in spirit to Blundell, Gosling, Ichimura, and Meghir (2007) in that we try and account for sample selection by simulating a worst-case scenario.

First, we include all firms which we observe before the crisis—Table 4.4—and *impute* the productivity level for those we do not see during the crisis. For the imputation we use the productivity level of the worst-performing firm in the period 2007-2010. That is, for those 44 firms we observe before but not during the crisis, we pretend that they are as productive as the least productive firm we do observe. To account for the fact that we are "adding" extreme observations to the data, we now run quantile regressions and estimate the conditional median. We obtain standard errors by drawing 1,000 three-digit-industry-clustered bootstrap samples.

We present the finding from this exercise in columns 1-3 of Table 4.A.6. In the most saturated specification, we estimate the conditional median to be -0.19 units lower for a point increase in Style 2 intensity. The effect is statistically significant at five percent.

Second, we pretend that the data is in fact censored and that we cannot observe the least productive firms because they had to exit. Thus, for all firms we do not observe in the period 2007-2010 (but do observe before), we impute the fifth percentile of the TFP distribution 2007-2010. In the second step, we estimate a Tobit-model with that fifth percentile being the left-censoring limit. Since the Tobit-model is not rank-based model, and as such sensitive to outliers, we use the fifth percentile rather

[15] In Table 4.A.5 we show additional results in which we use a firms' difference in productivity across the two periods. The results are qualitatively similar, and show that the difference in productivity levels is more negative if Style 2 intensity is higher. That is, holding productivity before the crisis constant, a higher Style 2 intensity results in a more negative difference across the two periods.

[16] The data weakly suggests that Style 2 intensity is indeed negatively, but statistically insignificantly, related to firm exit during the crisis. That is, conditional on sector, region, and value chain location fixed effects, we estimate a negative coefficient of -0.047 (SE = 0.056) of Style 2 intensity on firm survival in a linear probability model. The marginal effect at the mean from a logit regression is comparable.

	Firm productivity 2007 to 2010 95% winsorized			Firm productivity 2007 to 2010 not winsorized		
	(1)	(2)	(3)	(4)	(5)	(6)
Mgt style 2	-.14	-.13	-.13	-.1	-.092	-.09
	(.069)	(.064)	(.066)	(.081)	(.077)	(.077)
Pre-crisis TFP	.64	.6	.62	.62	.58	.6
	(.061)	(.066)	(.061)	(.09)	(.088)	(.084)
1[consumer good]			-.018			-.012
			(.057)			(.062)
1[intermediate good]			.026			.024
			(.053)			(.056)
Sector FE	No	Yes	Yes	No	Yes	Yes
Region FE	No	Yes	Yes	No	Yes	Yes
Adj R-squared	.41	.46	.49	.4	.45	.48
N. of cases	341	341	336	341	341	336

(a) Continuous measure of Style 2 intensity.

	Firm productivity 2007 to 2010 95% winsorized			Firm productivity 2007 to 2010 not winsorized		
	(1)	(2)	(3)	(4)	(5)	(6)
1[sytle 2 > $\frac{2}{3}$]	-.11	-.11	-.12	-.099	-.099	-.1
	(.048)	(.042)	(.043)	(.052)	(.048)	(.048)
1[$\frac{1}{3}$ < sytle 2 ≤ $\frac{2}{3}$]	-.033	-.0023	-.0011	-.0083	.025	.027
	(.042)	(.039)	(.041)	(.047)	(.044)	(.045)
Pre-crisis TFP	.64	.6	.62	.62	.58	.6
	(.06)	(.064)	(.059)	(.088)	(.085)	(.081)
1[consumer good]			-.015			-.0091
			(.057)			(.062)
1[intermediate good]			.034			.033
			(.054)			(.057)
Sector FE	No	Yes	Yes	No	Yes	Yes
Region FE	No	Yes	Yes	No	Yes	Yes
p: mid vs top tertile	.082	.0076	.0056	.056	.0064	.0052
Adj R-squared	.41	.46	.5	.41	.46	.49
N. of cases	341	341	336	341	341	336

(b) Style 2 intensity in terciles.

Table 4.5: Style 2 intensity and firm performance during the crisis. *Notes.* This table reports the relationship between management Style 2 intensity and firm performance during the Great Recession. In panel (a), firms' Style 2 intensity is measured by means of the continuous variable. In panel (b), Style 2 intensity is binned in terciles and the first tercile—the 33 percent of firms with the lowest Style 2 intensity—are the omitted category. The dependent variable is a firm's estimated TFP using data from 2007 to 2010; 95% winsorized in columns 1-3 and non-winsorized in columns 4-6. "Mgt Style 2" is a firm's Style 2 intensity. "1[consumer good]" and "1[intermediate good]" are indicators for firms that are located in the respective location along the value chain. The omitted category is firms producing capital goods. We additionally control for firms' TFP using data from 2001-2006 (the dependent variable in Section 4.4.2; this variable is also winsorized in columns 1-3 but not columns 4-6. Columns 2,3,5 and 6 contain sector and region fixed effects. Standard errors clustered at the three-digit industry level are reported in parentheses.

than the minimum for imputation. In this setting we report (analytic) standard errors again clustered at the three-digit industry level.

The results in columns 4-6 of Table 4.A.6 display the result of this exercise. Estimates retain a negative sign but are smaller in magnitude than those of Table 4.5 and do achieve statistical significance at conventional levels. In sum, we interpret the preponderance of negative estimates for Style 2 intensity as rather strong evidence that firms characterized by higher Style 2 intensity suffer more during the Great Recession *ceteris paribus*.

4.5.3 Mechanisms

At this point it is useful to recap the results presented thus far. First, we illustrated a novel approach to measuring management from high-dimensional survey data. Based on comparing single-plant and multi-plant firms, and the style-over-practices distributions, we argued that Style 2 reflects a more structured organizational design. Second, we reported how this measure of management style significantly correlates with firm performance in the period 2001 to 2006. We show that a higher Style 2 intensity positively affects firm productivity, and speculate that this may have allowed firms to exploit economies of scale during a period of economic expansion in Spain. Finally, in Section 4.5.2 we show that this correlation reverses its sign during the Great Recession 2007-2010. Firms with management more intensely geared towards Style 2 perform worse during the crisis, ceteris paribus.

In this section, we attempt to disentangle the ways and means that could help us understand this sign reversal. We investigate whether a higher Style 2 intensity hampers firms' ability to tackle the challenges of the Great Recession. We analyze two indicators. First, SABI data allows us to distinguish between fixed and non-fixed assets, and we analyze firms' holdings of non-fixed (i.e., rather liquid) assets before the crisis. Second, we analyze changes in the workforce as less rigidly organized firms may better able to adjust the workforce in the short term.

Table 4.6 shows the results of this exercise. Columns 1 and 2 show indeed that higher Style 2 intensity correlates with relatively lower holdings of non-fixed assets in 2006. Put differently, more Style 2 correlates with relatively more fixed assets, even after controlling for sector and region fixed effects. In columns 3 and 4, we show that Style 2 intensity weakly correlates with lower absolute employee turnover during the crisis. The dependent variable is the difference between the average number of employees of 2007-2010 to 2006. The estimates suggest that—holding

	Fraction non-fixed assets		Δ # employees	
	(1)	(2)	(3)	(4)
Mgt style 2	-.063	-.074	-6.3	-9.9
	(.033)	(.032)	(6.4)	(6.2)
Total # employees 2006			-.085	-.063
			(.031)	(.032)
Total assets 2006	.11	.16		
	(.39)	(.35)		
1[consumer good]		-.017		-5.6
		(.028)		(6.6)
1[intermediate good]		-.029		-.093
		(.027)		(5.5)
Sector FE	No	Yes	No	Yes
Region FE	No	Yes	No	Yes
Mean DV	.64	.64	-7.3	-7.3
Adj R-squared	.0027	.12	.069	.079
N. of cases	372	366	354	349

Table 4.6: Management style and ease of adjustment. *Notes.* This table shows conditional correlations of management Style 2 with pre-crisis holdings of non-fixed assets and employee turnover during the crisis. The dependent variable in columns 1 and 2 is the difference of total and fixed assets divided by total assets. All quantities are measured in 2006 and were 95 percent winsorized prior to entering the ratio. The dependent variable is the average number of employees from 2007-2010 minus the number of employees in 2006. "Mgt Style 2" measures firms' management Style 2 intensity from zero to 1. "1[consumer good]" and "1[intermediate good]" are indicators for firms that are located in the respective location along the value chain. The omitted category is firms producing capital goods. Columns 2 and 4 contain sector and region fixed effects. Standard errors clustered at the three-digit industry level are reported in parentheses.

constant employment in 2006—a higher Style 2 intensity correlates negatively with employee turnover. The table legend indicates that the average firm had to lay off about seven workers during the Great Recession. A one standard deviation change in Style 2 intensity (0.25) means that about 2.5 fewer employees were laid off.[17]

4.6 Conclusion

In this paper, we employ an unsupervised learning algorithm to measure clusters of management practices in Spanish firm survey data collected in 2006, i.e., just prior to the onset of the Great Financial Crisis. This allows us to classify every firm in our sample as a mixture of two "pure" styles: one rather informal and one rather structured style. The fact that our algorithm retrieves internally consistent clusters of practices is in line with there being complementarities that lead to sets of practices being adopted jointly.

The styles are meaningful in that they are not substantively determined by observable firm characteristics. Firm characteristics can explain only about ≈ 30 percent of variation in management styles. More importantly, they are correlated with firm performance despite the fact that the unsupervised learning algorithm does not force clusters to explain performance (as a supervised one would do). Specifically, we find positive correlations of a more structured management style with performance prior to the financial crisis. This correlation turns negative during the financial crisis after 2007.

Taking these results at face value, and in line with recent studies by Aghion, Bloom, Lucking, Sadun, and VanReenen (2020) and McElheran, Ohlmacher, and Yang (2020), we conclude that while structured management may fit stable economic conditions, in times of crisis more flexible and informal styles may strive. In terms of exploring mechanisms supporting this interpretation, we are somewhat restricted by our data. However, we document patterns that are consistent with structured management being an impediment to firms' short-term adjustment along two margins. First, we document that more structured firms adjust their workforce to a lesser degree during the crisis and that, prior to the crisis, more structured firms hold relatively more fixed assets than firms with a more informal management style.

[17] This estimates are robust to a number of different specification which we do not report here. Overall, specification that only control sector and region fixed effects lie in between the reported results in terms of magnitude and significance. The results in columns 1-2 become stronger and more statistically significant when we use the raw data for fixed and total assets rather than 95 percent winsorized values. The results in column 3-4 do not change when we control for the natural logarithm instead of the raw value of number of employees in 2006. The same is true for columns 1-2 and the logarithm of total assets in 2006. Finally, taking the difference of the number of employees in 2010 (rather than the average during the crisis) and 2006 produces slightly larger point estimates in columns 3-4.

Finally, we also see the present study as a proof of concept. We, as a profession, have access to a large amount of qualitative and diverse survey data on firm organization and employment practices. Unsupervised learning algorithms, such as LDA, offer a principled way to exploit the entirety of these data and hence a cost effective way to further our understanding of management practices and their intricate relationship to firm performance.

Appendices

The following pages present the appendices to all four chapters in this dissertation. The appendices contain additional tables and figures which I reference in the main text, more detail on certain procedures, or a combination thereof.

First, there are two appendices to Chapter 1. The first provides additional results and tables—especially summary statistics—and offers the reader greater insight into the underlying data structure. The second appendix contains tables and figures which speak to the results' robustness to alternative specifications.

Second, Chapter 2 is accompanied by three appendices. Again, the first reports further summary statistics and additional results. The second appendix describes in detail how the data on natural disasters was matched to the donations data from Betterplace, the online donations platform. Finally, there is an appendix which describes regularized regression—the lasso estimator—and how the procedure was applied to better understand correlates of giving and compare the predictive power of different approaches.

Third, there are two short appendices to Chapter 3. The first details all the variables which enter the construction of the indices used in the main analysis. The second shows the slides that were used during the marketing intervention and how they emphasize the randomly chosen marketing theme.

Fourth, there are two appendices to Chapter 4. The first again provides further results and summary statistics. The second appendix reports all 272 dummy variables which entered the estimation of the management styles via Latent Dirichlet Allocation.

Appendix A to Chapter 1: Additional tables and figures

Statistic	Mean (1)	St. Dev. (2)	Median (3)	Pctl(25) (4)	Pctl(75) (5)	Min (6)	Max (7)
2013	0.056	0.025	0.049	0.038	0.068	0.017	0.159
2014	0.056	0.023	0.050	0.040	0.070	0.011	0.144
2015	0.058	0.022	0.053	0.042	0.070	0.017	0.173
2016	0.060	0.021	0.056	0.045	0.071	0.012	0.142
2017	0.066	0.022	0.063	0.049	0.080	0.021	0.167
2018	0.068	0.024	0.064	0.052	0.082	0.023	0.160

Table 1.A.1: Dropout rates across time. *Notes.* This table reports summary statistics of school dropout rates from 2013 to 2018. It is based on data from 323 counties in Germany. “Pctl(25)” and “Pctl(75)” are the 25^{th} and 75^{th} percentile, respectively.

Statistic	Mean (1)	St. Dev. (2)	Median (3)	Pctl(25) (4)	Pctl(75) (5)	Min (6)	Max (7)
Unemployment rate	6.0	2.7	5.6	3.8	7.5	1.4	14.5
Vacancies helper	17.0	4.2	16.7	14.3	19.0	6.9	44.1
Vacancies skilled	66.8	3.8	66.8	64.7	69.3	45.2	77.4
Asylum seekers	7.1	4.6	6.2	5.1	7.7	1.8	69.4
In-migration	55.5	20.0	51.5	43.9	62.3	24.5	193.0
Out-migration	48.4	18.6	43.8	37.4	53.2	22.0	185.3
Sector 1 employment	1.2	1.3	.7	.3	1.7	.01	10.3
Sector 2 employment	32.9	10.6	32.3	25.5	41.1	7.6	62.8
State transfers	426.9	186.8	422.7	305.9	532.8	0.0	1158.7
Social security recipients	8.2	4.6	7.6	4.4	10.8	1.2	23.5
Local GDP	34.6	14.9	30.8	25.9	37.6	15.6	149.2

Table 1.A.2: Summary statistics of time-varying covariates. *Notes.* This table outlines the regional economic indicators that are used as covariates in the main specifications of this paper. It shows averages across the years 2013 to 2017. The unemployment rate gives the number of unemployed divided by the total labor force. “Vacancies helper” is the share of county vacancies which are defined as being routine and simple, whereas “Vacancy skilled” comprises vacancies that require special skills and training. Inward and outward migration is measured as the number of immigrants (emigrants) in a given year per 1,000 inhabitants in the same year. Employment in the first (agriculture, forestry and fishing) and second sector (manufacturing, mining, construction etc.) is measured as percentage of total employment in the county in the same year. “Transfers” are the total payments made from the state to the county in Euro 1000. “Social security recipients” is the share of the population below 65 that is eligible for social security payments. “Local GDP” is the value of all goods and services produced in a county divided by the total number of inhabitants. “Pctl(25)” and “Pctl(75)” are the 25^{th} and 75^{th} percentile, respectively.

	Dependent variable: dropout rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2015	.0076	.0096	.0068	.0068	.0089	.0059
	(.0031)	(.0036)	(.0036)	(.0031)	(.0036)	(.0037)
Post 2015 x Kaitz ratio Qua 2	.0008	.0007	.0009	.0010	.0008	.0011
	(.0012)	(.0012)	(.0012)	(.0012)	(.0012)	(.0012)
Post 2015 x Kaitz ratio Qua 3	.0033	.0029	.0033	.0032	.0029	.0035
	(.0017)	(.0016)	(.0016)	(.0016)	(.0016)	(.0015)
Post 2015 x Kaitz ratio Qua 4	.0037	.0037	.0044	.0059	.0060	.0065
	(.0023)	(.0021)	(.0021)	(.0024)	(.0023)	(.0023)
Unemployment t-1		.0024	.0010		.0025	.0012
		(.0013)	(.0011)		(.0014)	(.0011)
Vacancies helper t-1		.0002	.0002		.0002	.0003
		(.0001)	(.0001)		(.0001)	(.0001)
Vacancy skilled t-1		−.00004	−0.000000		−.00003	.00002
		(.0001)	(.0002)		(.0001)	(.0001)
Asylum seekers t-1		−.00002	−.00002		−.00002	−.00002
		(.00003)	(.00003)		(.00003)	(.00003)
In-migration t-1		.0001	.0001		.0001	.0001
		(.0001)	(.0001)		(.0001)	(.0001)
Out-migration t-1		−.0001	−.0001		−.0001	−.0001
		(.0001)	(.0001)		(.0001)	(.0001)
Sector 1 t-1			.0008			.0009
			(.0023)			(.0022)
Sector 2 t-1			.00002			.00002
			(.0003)			(.0003)
Transfers t-1			.00001			.00001
			(.000005)			(.000005)
Social security t-1			.0027			.0027
			(.0011)			(.0011)
Gross local product t-1			−.0002			−.0002
			(.0001)			(.0001)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income	Median income	Median income
Fixed effects	County	County	County	County	County	County
State X year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,394	2,390	2,382	2,394	2,390	2,382

Table 1.A.3: Minimum wage bite quartile and dropout rates. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. This interaction effect estimates are reproduced in Figure 1.4. “Post 2015” is an indicator for years 2015, 2016, 2017 and 2018. Counties are grouped into quartiles based on their quartile of minimum wage bite. Indicators for quartiles are interacted with the “Post 2015” indicator. The Kaitz-ratio is constructed from gross earnings (columns 1-3) or median income (columns 4-6). The specifications in columns one and four do not contain any time-varying covariates. The specifications in columns two and four contains the set of controls that can be found in columns two and four of Table 1.2. The set of controls in columns three and six is equivalent to the respective controls in the same columns of Table 1.2. Refer to the table notes of Table 1.2 for definitions of the variables. Each specification includes county and state-by-year fixed effects. Standard errors clustered at the county level are reported in parentheses.

	Dependent variable: dropout rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2015	.0081	.0102	.0076	.0076	.0095	.0069
	(.0030)	(.0035)	(.0035)	(.0031)	(.0035)	(.0037)
Post 2015 x 1[above median Kaitz-ratio]	.0030	.0027	.0031	.0031	.0031	.0035
	(.0015)	(.0014)	(.0014)	(.0014)	(.0013)	(.0013)
Unemployment t-1		.0024	.0010		.0024	.0011
		(.0013)	(.0011)		(.0013)	(.0011)
Vacancies helper t-1		.0002	.0003		.0002	.0003
		(.0001)	(.0001)		(.0001)	(.0001)
Vacancy skilled t-1		−.00004	.000001		−.00003	.00001
		(.0001)	(.0002)		(.0001)	(.0001)
Asylum seekers t-1		−.00002	−.00002		−.00002	−.00002
		(.00003)	(.00003)		(.00003)	(.00003)
In-migration t-1		.0001	.0001		.0001	.0001
		(.0001)	(.0001)		(.0001)	(.0001)
Out-migration t-1		−.0001	−.0001		−.0001	−.0001
		(.0001)	(.0001)		(.0001)	(.0001)
Sector 1 t-1			.0006			.0006
			(.0022)			(.0022)
Sector 2 t-1			.00002			.00002
			(.0003)			(.0003)
Transfers t-1			.00001			.00001
			(.000005)			(.000005)
Social security t-1			.0026			.0027
			(.0011)			(.0011)
Gross local product t-1			−.0002			−.0002
			(.0001)			(.0001)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income	Median income	Median income
Fixed effects	County	County	County	County	County	County
State X year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,394	2,390	2,382	2,394	2,390	2,382

Table 1.A.4: Above median minimum wage bite and dropout rates. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. Post 2015 is an indicator for years 2015, 2016,2017 and 2018. "Post 2015 x 1[above median Kaitz]" is an interaction term of the post 2015 indicator and a binary variable equal to one if counties' Kaitz-ratio was larger than the median Kaitz-ratio in 2014. The Kaitz-ratio is constructed from gross earnings (columns 1-3) or median income (columns 4-6). The specifications in columns one and four do not contain any time-varying covariates. The specifications in columns two and four contains the set of controls that can be found in columns two and four of Table 1.2. The set of controls in columns three and six is equivalent to the respective controls in the same columns of Table 1.2. Refer to the table notes of Table 1.2 for definitions of the variables. Each specification includes county and state-by-year fixed effects. The 95 percent confidence interval based on standard errors clustered at the county level is reported in parentheses.

	Dependent variable: dropout rate							
	Male				Female			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post 2015	−.0045 (.0088)	−.0105 (.0088)	.00004 (.0087)	−.0065 (.0088)	−.0034 (.0068)	−.0021 (.0070)	−.0021 (.0068)	−.0013 (.0069)
Post 2015 x Kaitz ratio	.0333 (.0138)	.0405 (.0133)	.0312 (.0166)	.0414 (.0162)	.0148 (.0100)	.0137 (.0095)	.0154 (.0122)	.0153 (.0116)
Umemployment t-1		.0007 (.0016)		.0008 (.0016)		.0017 (.0011)		.0017 (.0011)
Vacancies helper t-1		.0002 (.0002)		.0002 (.0002)		.0003 (.0001)		.0003 (.0001)
Vacancy skilled t-1		−.0001 (.0002)		−.0001 (.0002)		.0001 (.0001)		.0001 (.0001)
Asylum seekers t-1		−.00005 (.00004)		−.00004 (.00004)		.00001 (.00002)		.00001 (.00002)
In-migration t-1		.0001 (.0001)		.0001 (.0001)		.000001 (.0001)		.000003 (.0001)
Out-migration t-1		−.00001 (.0001)		−.00002 (.0001)		−.0001 (.0001)		−.0001 (.0001)
Sector 1 t-1		−.00003 (.0032)		.0003 (.0032)		.0023 (.0026)		.0024 (.0026)
Sector 2 t-1		.00003 (.0004)		.00003 (.0004)		−.00001 (.0002)		−.00001 (.0002)
Transfers t-1		.00001 (.00001)		.00001 (.00001)		.00001 (.000005)		.00001 (.000005)
Social security t-1		.0041 (.0015)		.0041 (.0015)		.0012 (.0009)		.0012 (.0009)
Gross local product t-1		−.000001 (.0001)		−.00003 (.0001)		−.0002 (.0001)		−.0002 (.0001)
Wage implied by:	Gross earnings	Gross earnings	Median income	Median income	Gross earnings	Gross earnings	Median income	Median income
Time-varying controls	None	Extended	None	Extended	None	Extended	None	Extended
Fixed effects	County	County	County	County	County	County	County	County
State x year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,393	2,381	2,393	2,381	2,393	2,381	2,393	2,381

Table 1.A.5: Minimum wage bite and dropout rates by gender. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates by gender. dropout rates are calculated for males (columns one to four) and females (columns five to eight) separately. “Post 2015” is an indicator for years 2015, 2016, 2017 and 2018. “Post 2015 x Kaitz-ratio” is an interaction term of this indicator and the Kaitz-ratio at the county level. The Kaitz-ratio is constructed from gross earning (columns 1,2,5 and 6) or median income (columns 3,4,7 and 8). The specifications in columns 1,3,5 and 7 do not contain any time-varying covariates. The specifications in columns 2,4,6 and 8 contain the set of controls that can be found in columns 3 and 6 of Table 1.2. Refer to the table notes of Table 1.2 for definitions of the variables. Each specification includes county and state-by-year fixed effects. The 95 percent confidence interval based on standard errors clustered at the county level is reported in parentheses.

	Dependent variable: dropout rate							
	Male				Female			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post 2015	.0102	.0080	.0102	.0080	.0034	.0040	.0034	.0040
	(.0037)	(.0044)	(.0037)	(.0044)	(.0039)	(.0044)	(.0039)	(.0044)
Post 2015 x Kaitz ratio Qua 2	.0010	.0010	.0010	.0010	.0012	.0014	.0012	.0014
	(.0017)	(.0017)	(.0017)	(.0017)	(.0012)	(.0012)	(.0012)	(.0012)
Post 2015 x Kaitz ratio Qua 3	.0046	.0053	.0046	.0053	.0017	.0017	.0017	.0017
	(.0023)	(.0022)	(.0023)	(.0022)	(.0014)	(.0013)	(.0014)	(.0013)
Post 2015 x Kaitz ratio Qua 4	.0079	.0089	.0079	.0089	.0034	.0035	.0034	.0035
	(.0033)	(.0032)	(.0033)	(.0032)	(.0022)	(.0021)	(.0022)	(.0021)
Umemployment t-1		.0008		.0008		.0017		.0017
		(.0016)		(.0016)		(.0012)		(.0012)
Vacancies helper t-1		.0002		.0002		.0003		.0003
		(.0002)		(.0002)		(.0001)		(.0001)
Vacancy skilled t-1		−.0001		−.0001		.0001		.0001
		(.0002)		(.0002)		(.0001)		(.0001)
Asylum seekers t-1		−.00004		−.00004		.00001		.00001
		(.00004)		(.00004)		(.00002)		(.00002)
In-migration t-1		.0001		.0001		.000003		.000003
		(.0001)		(.0001)		(.0001)		(.0001)
Out-migration t-1		−.00002		−.00002		−.0001		−.0001
		(.0001)		(.0001)		(.0001)		(.0001)
Sector 1 t-1		−.0004		−.0004		.0022		.0022
		(.0032)		(.0032)		(.0026)		(.0026)
Sector 2 t-1		.00004		.00004		−.00001		−.00001
		(.0004)		(.0004)		(.0002)		(.0002)
Transfers t-1		.00001		.00001		.00001		.00001
		(.00001)		(.00001)		(.000005)		(.000005)
Social security t-1		.0040		.0040		.0012		.0012
		(.0015)		(.0015)		(.0009)		(.0009)
Gross local product t-1		−.0001		−.0001		−.0002		−.0002
		(.0001)		(.0001)		(.0001)		(.0001)
Wage implied by:	Gross earnings	Gross earnings	Median income	Median income	Gross earnings	Gross earnings	Median income	Median income
Time-varying controls	None	Extended	None	Extended	None	Extended	None	Extended
Fixed effects	County	County	County	County	County	County	County	County
State x year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,393	2,381	2,393	2,381	2,393	2,381	2,393	2,381

Table 1.A.6: Minimum wage bite quartile and dropout rates by gender. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates by gender. This interaction effect estimates are reproduced in Figure 1.4. "Post 2015" is an indicator for years 2015, 2016, 2017 and 2018. Counties are grouped into quartiles based on their quartile of minimum wage bite. Indicators for quartiles are interacted with the "Post 2015" indicator. The Kaitz-ratio is constructed from gross earnings (columns 1-3) or median income (columns 4-6). The specifications in columns one and four do not contain any time-varying covariates. The specifications in columns two and four contains the set of controls that can be found in columns two and four of Table 1.2. The set of controls in columns three and six is equivalent to the respective controls in the same columns of Table 1.2. Refer to the table notes of Table 1.2 for definitions of the variables. Each specification includes county and state-by-year fixed effects. The 95 percent confidence interval based on standard errors clustered at the county level is reported in parentheses.

	Dependent variable: dropout rate							
	Male		Female		Male		Female	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post 2015	−.0020	−.0067	−.0048	−.0032	.0001	−.0056	−.0051	−.0041
	(.0104)	(.0103)	(.0077)	(.0079)	(.0112)	(.0111)	(.0083)	(.0083)
Post 2015 X rural	−.0016	−.0061	−.0015	−.0041	.0069	.0041	.0017	−.0002
	(.0131)	(.0129)	(.0088)	(.0085)	(.0122)	(.0116)	(.0072)	(.0070)
Post 2015 x Kaitz ratio	.0270	.0317	.0188	.0171	.0286	.0373	.0234	.0231
	(.0185)	(.0177)	(.0124)	(.0122)	(.0238)	(.0230)	(.0164)	(.0157)
Post 2015 x Kaitz ratio x rural	.0053	.0132	.0004	.0045	−.0104	−.0052	−.0062	−.0028
	(.0238)	(.0234)	(.0159)	(.0155)	(.0262)	(.0248)	(.0152)	(.0147)
Umemployment t-1		.0007		.0018		.0007		.0018
		(.0016)		(.0011)		(.0016)		(.0011)
Vacancies helper t-1		.0002		.0004		.0002		.0004
		(.0002)		(.0001)		(.0002)		(.0001)
Vacancy skilled t-1		−.0001		.0001		−.0001		.0001
		(.0002)		(.0001)		(.0002)		(.0001)
Asylum seekers t-1		−.00005		.00001		−.00005		.00001
		(.00004)		(.00002)		(.00004)		(.00002)
In-migration t-1		.0001		.00001		.0001		.00001
		(.0001)		(.0001)		(.0001)		(.0001)
Out-migration t-1		−.00001		−.0001		−.00001		−.0001
		(.0001)		(.0001)		(.0001)		(.0001)
Sector 1 t-1		−.00001		.0025		−.00002		.0026
		(.0032)		(.0026)		(.0032)		(.0026)
Sector 2 t-1		.00003		−.00002		.00004		−.00001
		(.0004)		(.0002)		(.0004)		(.0002)
Transfers t-1		.00001		.00001		.00001		.00001
		(.00001)		(.000005)		(.00001)		(.000005)
Social security t-1		.0042		.0012		.0041		.0012
		(.0015)		(.0009)		(.0015)		(.0009)
Gross local product t-1		−.00001		−.0002		−.00004		−.0002
		(.0001)		(.0001)		(.0001)		(.0001)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Gross earnings	Median income	Median income	Median income	Median income
Time-varying controls	None	Extended	None	Extended	None	Extended	None	Extended
Fixed effects	County	County	County	County	County	County	County	County
State X year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,393	2,381	2,393	2,381	2,393	2,381	2,393	2,381

Table 1.A.7: Minimum wage bite and gender-specific dropout rates by rural vs urban.
Notes. This table reports estimates of the minimum wage introduction on dropout rates by gender; male students' dropout rates in columns 1 through 4, and females' columns 5 through 8. "Post 2015" is an indicator for years 2015, 2016 and 2017. "Post 2015 x Kaitz-ratio" is an interaction term of this indicator and the Kaitz-ratio at the county level. "Rural" is an indicator equal to one if the county is classfied as urban, and 0 otherwise. The Kaitz-ratio is constructed from gross earning (columns 1 through 4) or median income (columns 5 through 8). The specifications in columns 1,3,5 and 7 do not contain any time-varying covariates. Precise definitions of control variables can be found in the notes to Table 1.2.Each specification includes county and state-by-year fixed effects. Standard errors clustered at the county level are reported in parentheses.

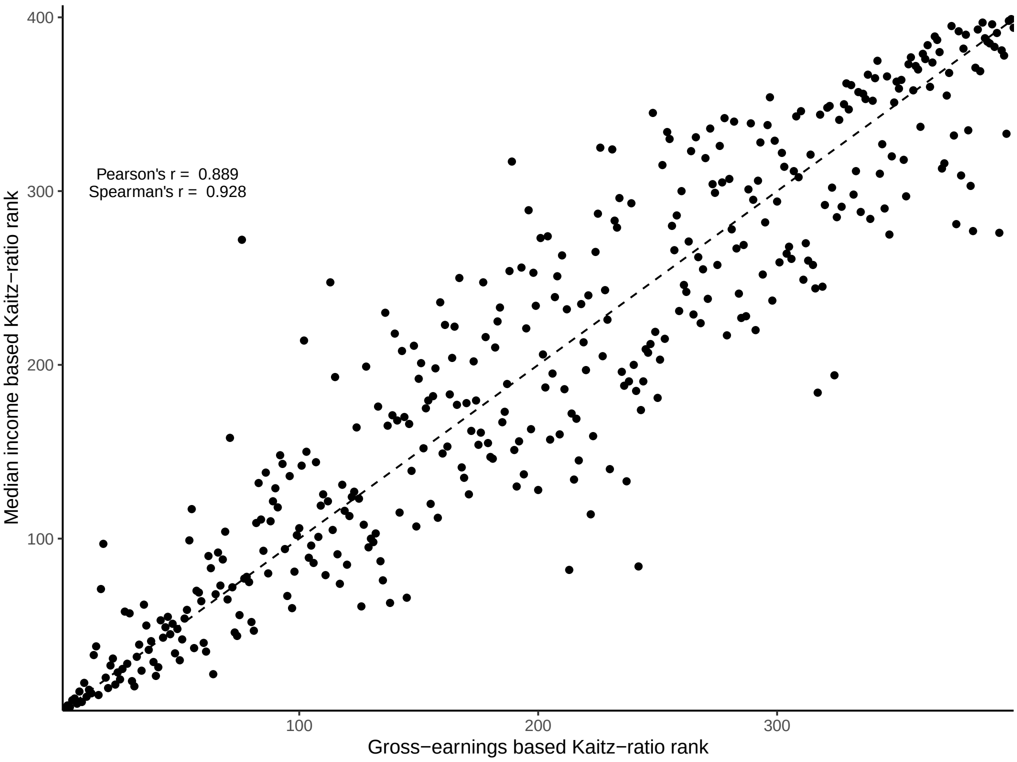


Figure 1.A.1: Comparison of minimum wage bite measures. *Notes.* This figure plots the ranks of counties based on their Kaitz-ratios implies by gross earnings (x-axis) and median income (y-axis). The dashed line indicates the 45^{deg} which would suggest perfect rank dependence. Higher ranks indicate higher values of the Kaitz-ratio which implies lower hourly wages in 2014.

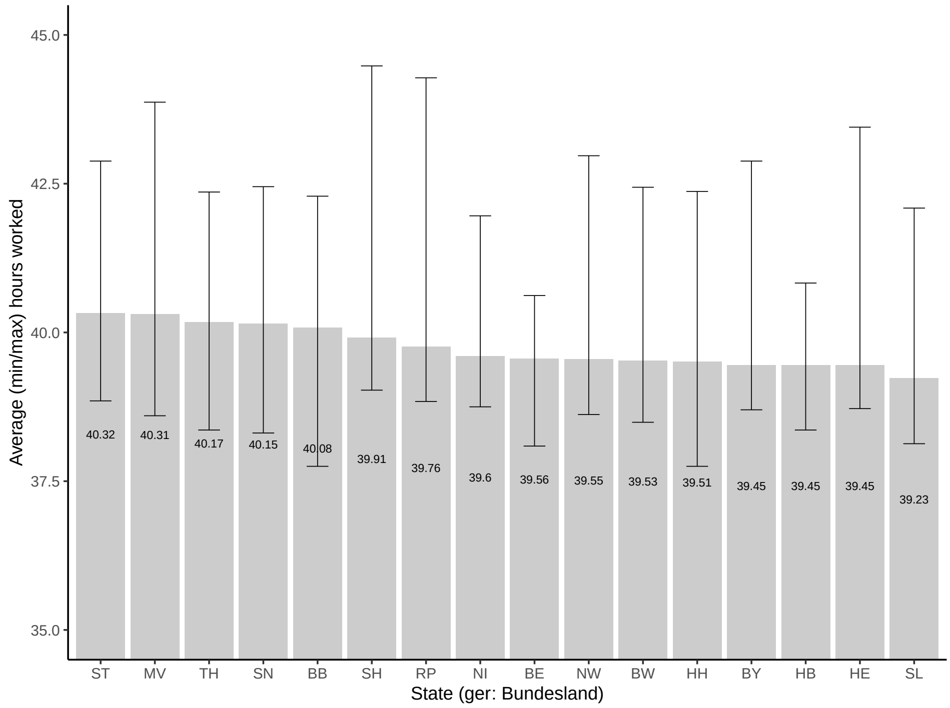


Figure 1.A.2: Hours worked across states in 2014. *Notes.* This figure plots average labor hours across German states (ger: Bundeslaender) for April 2014 from the labor earnings survey (ger: *Verdienststrukturerhebung*). The grey bars indicate the average hours worked in a given week for an employee in a given state. The precise number if the label inside the bar. The error bars report the minimum and maximum hours worked across sectors in that particular state. Maximum hours are typically worked in mining and agriculture while minimum hours are in the energy sector and manufacturing. There is a total of 19 sectors.

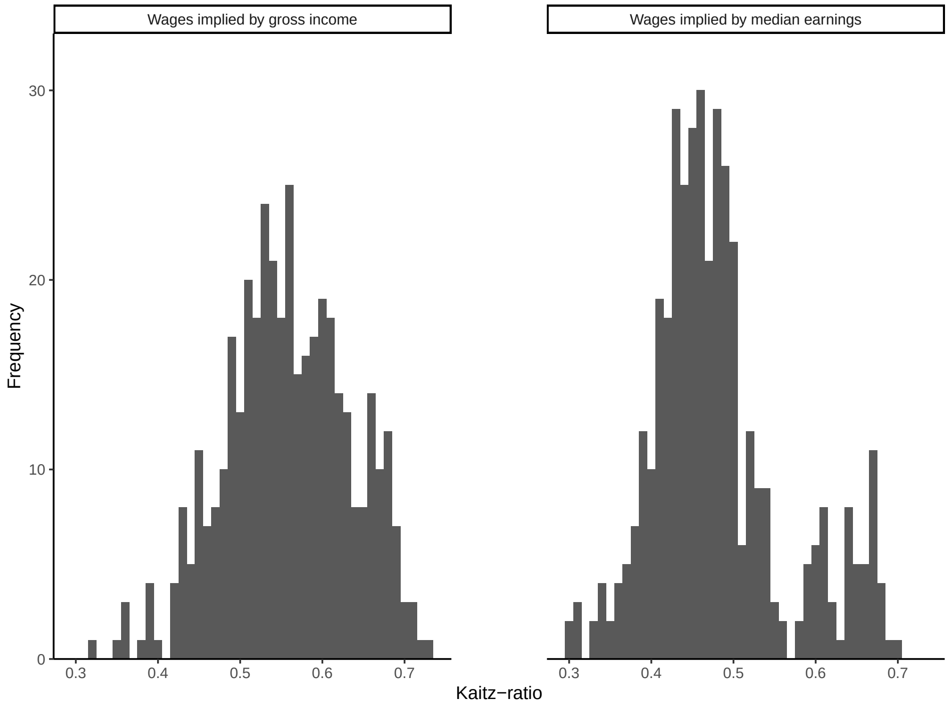


Figure 1.A.3: Kaitz-ratios in 2014. *Notes.* This figure plots distributions of observed Kaitz-ratios of all German states in 2014 (see Table 1.1). The distributions are based on wages implied by gross income (left) or median earnings (right).

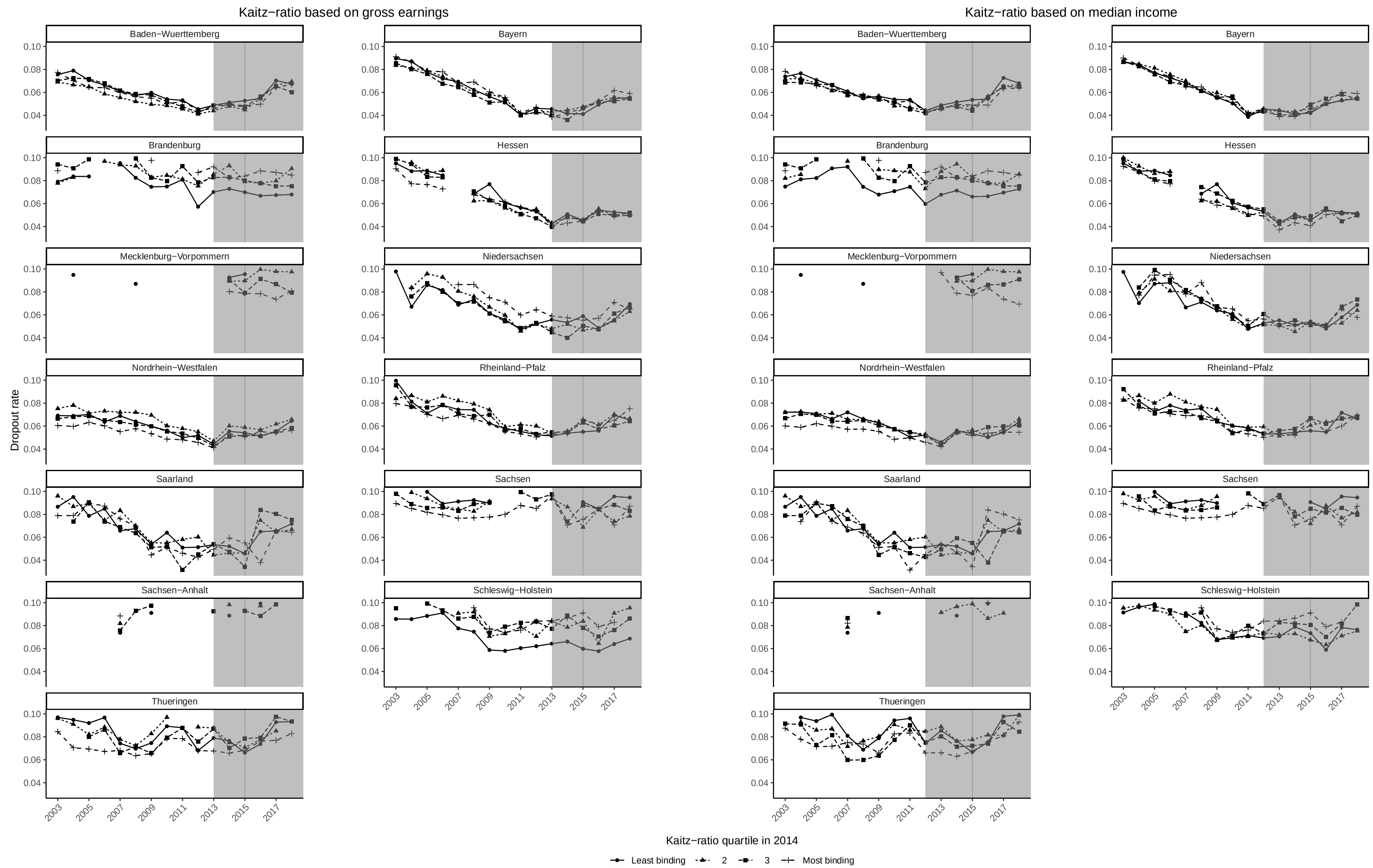


Figure 1.A.4: Trends in dropout rates by minimum wage bite across states. *Notes.* This figure shows the Kaitz-ratios for the years 2003 to 2017 separately for each German state. The sample observation period is shaded grey and the introduction of the minimum wage in 2015 is marked by a vertical line in each graph. Each state's counties are placed into quartiles depending on the value of the prevailing Kaitz-ratio in 2014. The counties with the least binding minimum wage are in the first quartile (solid line with dots) while the counties with the most binding minimum wage are displayed by a dashed line with crosses. The left (right) section shows trends based on Kaitz-ratio constructed from gross earnings (median income).

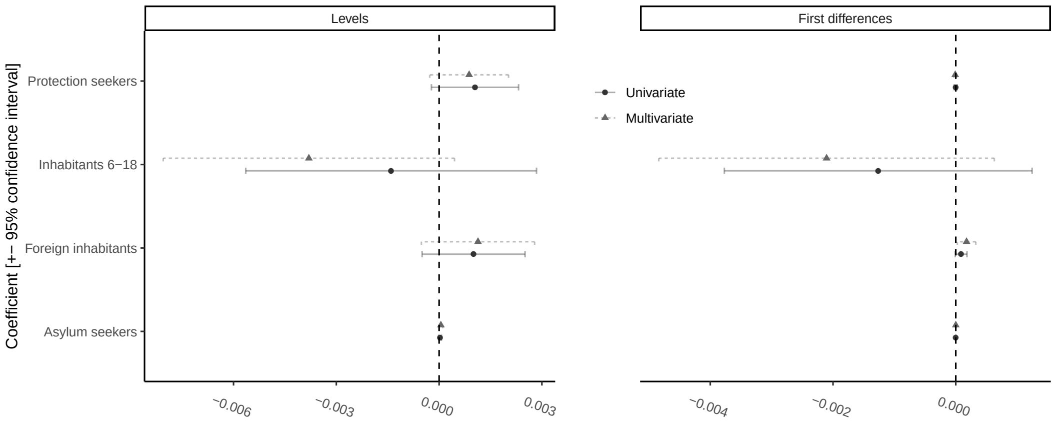


Figure 1.A.5: Immigration correlates of dropout rates. *Notes.* This figure plots coefficients from a regression in which the dependent variable is a county's dropout rate in year t, and the independent variable of interest is either a county's share of inhabitants aged 6-18, its share of foreigners, its share of those seeking protection or its share of those granted asylum. Point estimates labeled "univariate" are obtained from each variable separately; "multivariate" means that all variables are included at once. Each regression includes state by year and county fixed effects. The models depicted in the left-hand panel are estimated in levels while those on the right hand were estimated in first differences. The error bars show 95 percent confidence intervals based on standard errors clustered at the county level.

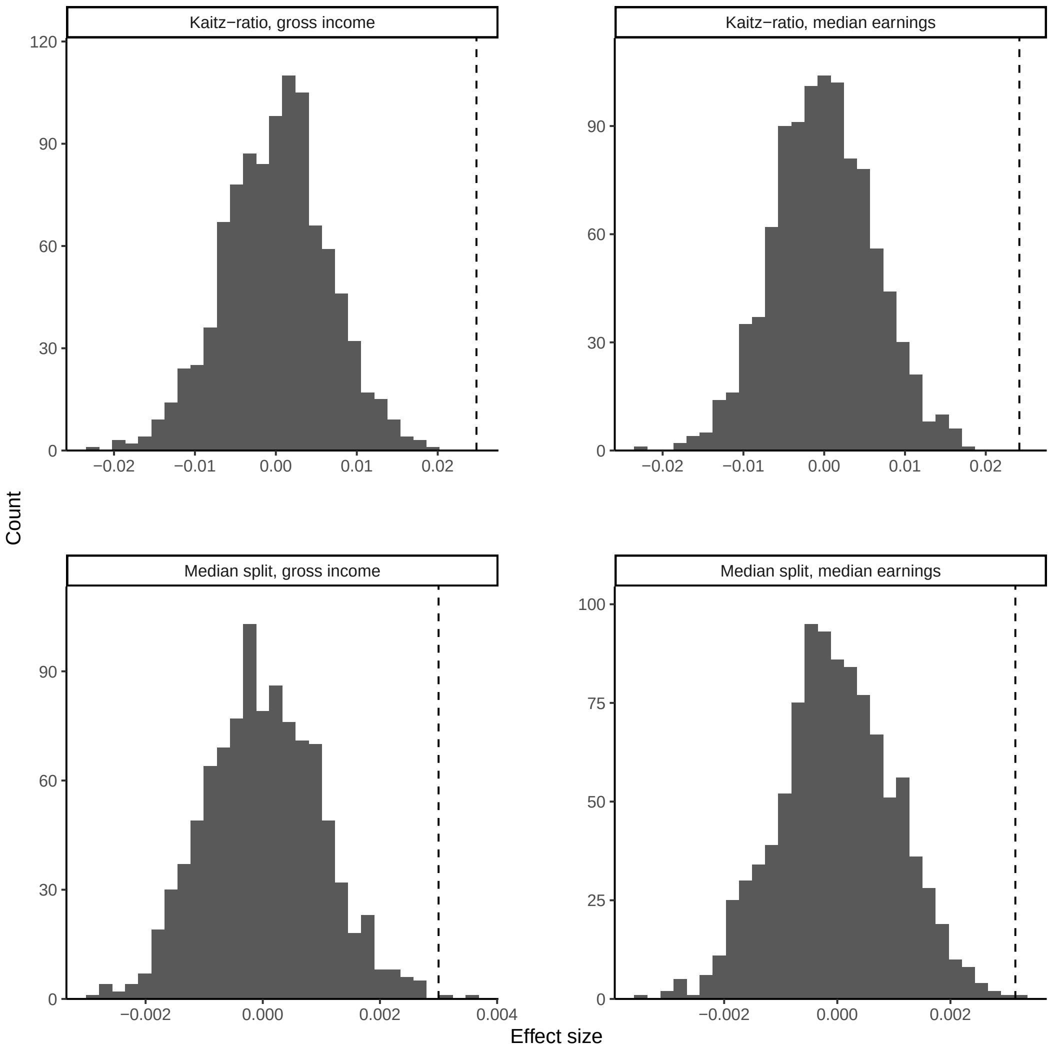


Figure 1.A.6: Randomization distributions of key quantities. *Notes.* This figure plots the randomization distributions for four "treatment effects": The two upper graphs show distributions in which treatment is a county's Kaitz-ratio in 2014, as implied by either gross income or median earnings. In the lower graphs, treatment is based on a median split on those quantities. The vertical, dashed lines indicate the observed effect in each of those four scenarios.

Appendix B to Chapter 1: Robustness checks

	Dependent variable: dropout rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2015	−.0072	−.0083	−.0097	−.0022	−.0036	−.0055
	(.0051)	(.0049)	(.0047)	(.0042)	(.0041)	(.0040)
Post 2015 x Kaitz ratio	.0064	.0084	.0125	−.0030	−.0001	.0056
	(.0092)	(.0089)	(.0085)	(.0089)	(.0087)	(.0084)
Unemployment t-1		.0023	−.0002		.0022	−.0002
		(.0010)	(.0010)		(.0010)	(.0010)
Vacancies helper t-1		.0001	.0002		.0001	.0002
		(.0001)	(.0001)		(.0001)	(.0001)
Vacancy skilled t-1		−.0001	−.0001		−.0001	−.0001
		(.0002)	(.0002)		(.0002)	(.0002)
Asylum seekers t-1		−.00003	−.00001		−.00003	−.00001
		(.00002)	(.00002)		(.00002)	(.00002)
In-migration t-1		−.00005	−.000001		−.00004	.000004
		(.00004)	(.00005)		(.00004)	(.00005)
Out-migration t-1		.0001	.00003		.0001	.00002
		(.00005)	(.0001)		(.00005)	(.0001)
Sector 1 t-1			−.0030			−.0030
			(.0027)			(.0027)
Sector 2 t-1			−.0001			−.0001
			(.0003)			(.0003)
Transfers t-1			.000002			.000002
			(.000005)			(.000005)
Social security t-1			.0037			.0037
			(.0009)			(.0009)
Gross local product t-1			−.0001			−.0001
			(.0001)			(.0001)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income	Median income	Median income
Fixed effects	County	County	County	County	County	County
State-linear trend	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,394	2,390	2,382	2,394	2,390	2,382

Table 1.B.1: Continuous bite and state-specific linear trends. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. Variable definitions follow exactly from Table 1.2. In contrast to the main specification, these specifications include state-specific linear time trends for the period 2013-2018. Each specification further includes county fixed effects. Standard errors clustered at the county level are reported in parentheses.

	Dependent variable: dropout rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2015	−.0045	−.0045	−.0040	−.0048	−.0047	−.0043
	(.0011)	(.0011)	(.0011)	(.0012)	(.0012)	(.0011)
Post 2015 x Kaitz ratio Qua 2	.0011	.0011	.0012	.0010	.0009	.0011
	(.0012)	(.0012)	(.0012)	(.0012)	(.0012)	(.0012)
Post 2015 x Kaitz ratio Qua 3	.0019	.0018	.0023	.0027	.0025	.0030
	(.0016)	(.0016)	(.0015)	(.0015)	(.0015)	(.0015)
Post 2015 x Kaitz ratio Qua 4	.0002	.0007	.0016	.0007	.0011	.0020
	(.0019)	(.0018)	(.0018)	(.0019)	(.0019)	(.0018)
Unemployment t-1		.0022	−.0003		.0021	−.0003
		(.0010)	(.0010)		(.0010)	(.0010)
Vacancies helper t-1		.0001	.0002		.0001	.0002
		(.0001)	(.0001)		(.0001)	(.0001)
Vacancy skilled t-1		−.0001	−.0001		−.0001	−.0001
		(.0002)	(.0002)		(.0002)	(.0002)
Asylum seekers t-1		−.00003	−.00001		−.00003	−.00001
		(.00002)	(.00002)		(.00002)	(.00002)
In-migration t-1		−.00004	.000001		−.00004	.000002
		(.00004)	(.00005)		(.00004)	(.00005)
Out-migration t-1		.0001	.00002		.0001	.00002
		(.0001)	(.0001)		(.0001)	(.0001)
Sector 1 t-1			−.0034			−.0034
			(.0027)			(.0028)
Sector 2 t-1			−.0001			−.0001
			(.0003)			(.0003)
Transfers t-1			.000002			.000002
			(.000005)			(.000005)
Social security t-1			.0037			.0037
			(.0009)			(.0009)
Gross local product t-1			−.0001			−.0001
			(.0001)			(.0001)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income	Median income	Median income
Fixed effects	County	County	County	County	County	County
State-linear trend	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,394	2,390	2,382	2,394	2,390	2,382

Table 1.B.2: Quartile bite and state-specific linear trends. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. Variable definitions follow exactly from Table 1.2. In contrast to the main specification, these specifications include state-specific linear time trends for the period 2013-2018. Each specification further includes county fixed effects. Standard errors clustered at the county level are reported in parentheses.

	Dependent variable: dropout rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2015	−.0155	−.0174	−.0287	−.0126	−.0139	−.0258
	(.0066)	(.0066)	(.0062)	(.0064)	(.0065)	(.0061)
Post 2015 x Kaitz ratio	.0248	.0249	.0382	.0242	.0230	.0405
	(.0108)	(.0107)	(.0100)	(.0128)	(.0127)	(.0121)
Unemployment t-1		−.0026	−.0049		−.0027	−.0050
		(.0007)	(.0008)		(.0007)	(.0008)
Vacancies helper t-1		.0004	.0003		.0004	.0003
		(.0002)	(.0002)		(.0002)	(.0002)
Vacancy skilled t-1		−.0001	−.00002		−.0001	−.00004
		(.0002)	(.0002)		(.0002)	(.0002)
Asylum seekers t-1		−.00003	−.00001		−.00003	−.00001
		(.00002)	(.00002)		(.00002)	(.00002)
In-migration t-1		−.0001	−.00001		−.0001	−.00001
		(.00005)	(.00005)		(.00005)	(.00005)
Out-migration t-1		.0002	.0001		.0002	.0001
		(.0001)	(.0001)		(.0001)	(.0001)
Sector 1 t-1			−.0021			−.0018
			(.0026)			(.0025)
Sector 2 t-1			−.0004			−.0004
			(.0002)			(.0002)
Transfers t-1			.00001			.00001
			(.000005)			(.000005)
Social security t-1			.0051			.0051
			(.0009)			(.0009)
Gross local product t-1			.0004			.0004
			(.0002)			(.0002)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income	Median income	Median income
Fixed effects	County	County	County	County	County	County
State X post FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,394	2,390	2,382	2,394	2,390	2,382

Table 1.B.3: Continuous bite and state post-2015 fixed effects. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. Variable definitions follow exactly from Table 1.2. In contrast to the main specification, these specifications include state post-2015 fixed effects; that is, a dummy for each interaction of a state and the "post 2015" indicator. Each specification further includes county fixed effects. Standard errors clustered at the county level are reported in parentheses.

	Dependent variable: dropout rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2015	−.0037	−.0054	−.0104	−.0045	−.0063	−.0113
	(.0022)	(.0023)	(.0024)	(.0021)	(.0023)	(.0024)
Post 2015 x Kaitz ratio Qua 2	.0008	.0007	.0015	.0010	.0012	.0022
	(.0012)	(.0012)	(.0012)	(.0012)	(.0012)	(.0012)
Post 2015 x Kaitz ratio Qua 3	.0033	.0032	.0047	.0032	.0033	.0050
	(.0017)	(.0017)	(.0016)	(.0016)	(.0016)	(.0015)
Post 2015 x Kaitz ratio Qua 4	.0037	.0036	.0060	.0059	.0056	.0075
	(.0022)	(.0022)	(.0021)	(.0024)	(.0024)	(.0024)
Unemployment t-1		−.0027	−.0051		−.0027	−.0050
		(.0007)	(.0008)		(.0007)	(.0008)
Vacancies helper t-1		.0004	.0003		.0004	.0003
		(.0002)	(.0001)		(.0002)	(.0001)
Vacancy skilled t-1		−.0001	−.00003		−.0001	−.00002
		(.0002)	(.0002)		(.0002)	(.0002)
Asylum seekers t-1		−.00003	−.000004		−.00003	−.00001
		(.00002)	(.00002)		(.00002)	(.00002)
In-migration t-1		−.0001	−.00001		−.0001	−.00001
		(.00005)	(.00005)		(.00005)	(.00005)
Out-migration t-1		.0002	.00005		.0002	.00005
		(.0001)	(.0001)		(.0001)	(.0001)
Sector 1 t-1			−.0024			−.0024
			(.0026)			(.0026)
Sector 2 t-1			−.0004			−.0004
			(.0002)			(.0002)
Transfers t-1			.00001			.00001
			(.000005)			(.000005)
Social security t-1			.0052			.0051
			(.0009)			(.0009)
Gross local product t-1			.0004			.0004
			(.0002)			(.0002)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income	Median income	Median income
Fixed effects	County	County	County	County	County	County
State X post FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,394	2,390	2,382	2,394	2,390	2,382

Table 1.B.4: Quartile bite and state post-2015 fixed effects. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. Variable definitions follow exactly from Table 1.2. In contrast to the main specification, these specifications include state post-2015 fixed effects; that is, a dummy for each interaction of a state and the "post 2015" indicator. Each specification further includes county fixed effects. Standard errors clustered at the county level are reported in parentheses.

	Dependent variable: dropout rate			
	(1)	(2)	(3)	(4)
Post 2015	−.0144	−.0213	−.0124	−.0200
	(.0067)	(.0068)	(.0065)	(.0068)
Post 2015 x Kaitz ratio	.0308	.0380	.0334	.0437
	(.0101)	(.0103)	(.0121)	(.0125)
Umemployment	.0029	.0013	.0029	.0014
	(.0013)	(.0013)	(.0013)	(.0013)
Vacancies helper	.0001	.0001	.0001	.0001
	(.0002)	(.0002)	(.0002)	(.0002)
Vacancy skilled	−.00001	.00002	−.00002	.000003
	(.0002)	(.0002)	(.0002)	(.0002)
Asylum seekers	.00003	.00003	.00003	.00003
	(.00002)	(.00002)	(.00002)	(.00002)
In-migraton	.0001	.0001	.0001	.0001
	(.0001)	(.0001)	(.0001)	(.0001)
Out-migration	−.0001	−.0001	−.0001	−.0001
	(.0001)	(.0001)	(.0001)	(.0001)
Sector 1		.0036		.0041
		(.0026)		(.0026)
Sector 2		−.0002		−.0002
		(.0002)		(.0002)
Transfers		.000003		.000003
		(.000004)		(.000004)
Social security		.0026		.0027
		(.0010)		(.0010)
Gross local product		.0001		.0001
		(.0001)		(.0001)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income
Fixed effects	County	County	County	County
State X year FE	Yes	Yes	Yes	Yes
Observations	1,991	1,983	1,991	1,983

Table 1.B.5: Continuous bite and with contemporaneous time-varying covariates. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. Variable definitions follow exactly from Table 1.2. In contrast to the main specification, these specifications include the contemporaneous level of the control variables. This also accounts for the loss of 323 observations compared to Table 1.2 as the year 2018 cannot be considered due to incomplete schooling records. Each specification further includes county fixed effects. Standard errors clustered at the county level are reported in parentheses.

	Dependent variable: dropout rate			
	(1)	(2)	(3)	(4)
Post 2015	.00005	−.0031	−.0009	−.0039
	(.0032)	(.0033)	(.0033)	(.0033)
Post 2015 x Kaitz ratio Qua 2	.0016	.0021	.0012	.0016
	(.0013)	(.0013)	(.0012)	(.0012)
Post 2015 x Kaitz ratio Qua 3	.0039	.0045	.0042	.0049
	(.0017)	(.0017)	(.0017)	(.0017)
Post 2015 x Kaitz ratio Qua 4	.0049	.0064	.0078	.0084
	(.0021)	(.0022)	(.0024)	(.0024)
Umemployment	.0028	.0012	.0029	.0014
	(.0013)	(.0013)	(.0013)	(.0013)
Vacancies helper	.0001	.0001	.0002	.0002
	(.0002)	(.0002)	(.0002)	(.0002)
Vacancy skilled	−.00001	.00001	.00001	.00004
	(.0002)	(.0002)	(.0002)	(.0002)
Asylum seekers	.00003	.00003	.00003	.00003
	(.00002)	(.00002)	(.00002)	(.00002)
In-migraton	.0001	.0001	.0001	.0001
	(.0001)	(.0001)	(.0001)	(.0001)
Out-migration	−.0001	−.0001	−.0001	−.0001
	(.0001)	(.0001)	(.0001)	(.0001)
Sector 1		.0034		.0032
		(.0027)		(.0026)
Sector 2		−.0002		−.0002
		(.0002)		(.0002)
Transfers		.000003		.000003
		(.000004)		(.000004)
Social security		.0026		.0025
		(.0010)		(.0010)
Gross local product		.00005		.0001
		(.0001)		(.0001)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income
Fixed effects	County	County	County	County
State X year FE	Yes	Yes	Yes	Yes
Observations	1,991	1,983	1,991	1,983

Table 1.B.6: Quartile bite and with contemporaneous time-varying covariates. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. Variable definitions follow exactly from Table 1.2. In contrast to the main specification, these specifications include the contemporaneous level of the control variables. This also accounts for the loss of 323 observations compared to Table 1.2 as the year 2018 cannot be considered due to incomplete schooling records. Each specification further includes county fixed effects. Standard errors clustered at the county level are reported in parentheses.

	Dependent variable: dropout rate			
	(1)	(2)	(3)	(4)
Post 2015	−.0020	−.0033	.0009	−.0001
	(.0075)	(.0072)	(.0072)	(.0070)
Post 2015 x Kaitz ratio	.0211	.0210	.0197	.0190
	(.0108)	(.0106)	(.0128)	(.0127)
Umemployment t-2	.0005	−.0003	.0005	−.0002
	(.0013)	(.0012)	(.0013)	(.0012)
Vacancies helper t-2	−.00004	−.00003	−.00005	−.00003
	(.0001)	(.0001)	(.0001)	(.0001)
Vacancy skilled t-2	−.00002	−.00001	−.00003	−.00002
	(.0001)	(.0001)	(.0001)	(.0001)
Asylum seekers t-2	.00002	.00002	.00002	.00002
	(.00002)	(.00002)	(.00002)	(.00002)
In-migration t-2	.0002	.0002	.0002	.0002
	(.0001)	(.0001)	(.0001)	(.0001)
Out-migration t-2	−.0002	−.0002	−.0002	−.0002
	(.0001)	(.0001)	(.0001)	(.0001)
Sector 1 t-2		−.0007		−.0005
		(.0028)		(.0028)
Sector 2 t-2		−.0001		−.0001
		(.0003)		(.0003)
Transfers t-2		.00001		.00001
		(.000005)		(.000005)
Social security t-2		.0016		.0015
		(.0012)		(.0012)
Gross local product t-2		.00001		−.00001
		(.0001)		(.0001)
Wage implied by:	Gross earnings	Gross earnings	Gross earnings	Median income
Fixed effects	County	County	County	County
State X year FE	Yes	Yes	Yes	Yes
Observations	2,391	2,391	2,391	2,391

Table 1.B.7: Continuous bite and with second lag of time-varying covariates. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. Variable definitions follow exactly from Table 1.2. In contrast to the main specification, these specifications include the second lag level of the control variables. Each specification further includes county fixed effects. Standard errors clustered at the county level are reported in parentheses.

	Dependent variable: dropout rate			
	(1)	(2)	(3)	(4)
Post 2015	.0081	.0068	.0074	.0060
	(.0037)	(.0036)	(.0036)	(.0036)
Post 2015 x Kaitz ratio Qua 2	.0005	.0004	.0006	.0006
	(.0012)	(.0012)	(.0012)	(.0012)
Post 2015 x Kaitz ratio Qua 3	.0030	.0030	.0027	.0028
	(.0017)	(.0016)	(.0016)	(.0016)
Post 2015 x Kaitz ratio Qua 4	.0031	.0031	.0055	.0055
	(.0022)	(.0022)	(.0024)	(.0023)
Umemployment t-2	.0005	−.0003	.0006	−.0002
	(.0013)	(.0012)	(.0013)	(.0012)
Vacancies helper t-2	−.0001	−.00004	−.00004	−.00002
	(.0001)	(.0001)	(.0001)	(.0001)
Vacancy skilled t-2	−.00003	−.00001	−.00002	−.00001
	(.0001)	(.0001)	(.0001)	(.0001)
Asylum seekers t-2	.00002	.00002	.00002	.00002
	(.00002)	(.00002)	(.00002)	(.00002)
In-migration t-2	.0002	.0002	.0002	.0002
	(.0001)	(.0001)	(.0001)	(.0001)
Out-migration t-2	−.0002	−.0002	−.0002	−.0002
	(.0001)	(.0001)	(.0001)	(.0001)
Sector 1 t-2		−.0009		−.0009
		(.0028)		(.0028)
Sector 2 t-2		−.0001		−.0001
		(.0003)		(.0003)
Transfers t-2		.00001		.00001
		(.000005)		(.000005)
Social security t-2		.0016		.0016
		(.0012)		(.0012)
Gross local product t-2		−.00003		−.00001
		(.0001)		(.0001)
Observations	2,391	2,391	2,391	2,391

Table 1.B.8: Quartile bite and with second lag of time-varying covariates. *Notes.* This table reports estimates of the minimum wage introduction on dropout rates. Variable definitions follow exactly from Table 1.2. In contrast to the main specification, these specifications include the second lag level of the control variables. Each specification further includes county fixed effects. Standard errors clustered at the county level are reported in parentheses.

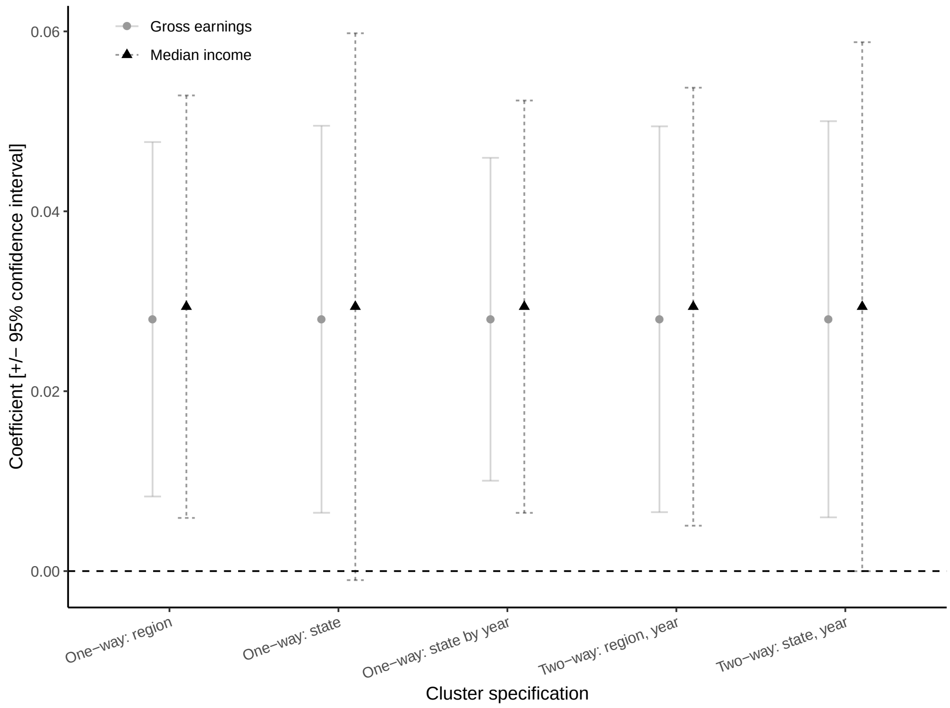


Figure 1.B.1: Alternative cluster specifications. *Notes.* This figure plots the benchmark estimate from columns three and six of Table 1.2 of the interaction term of the post-2015 period and the 2014 minimum wage bite. The error bars indicate 95 percent confidence intervals based on different ways of specifying clusters in the variance-covariance matrix computation.

Appendix A to Chapter 2: Additional tables and figures

Start date (1)	Country (2)	Disaster type (3)	# casualties (4)	# affected (5)	Damage (in \$) (6)	End date (7)	Total relief (8)
25/04/2015	Nepal	Earthquake	8,969	5,642,150	5,174,000	25/04/2015	865,055
/01/2015	India	Drought	0	330,000,000	3,000,000	/12/2016	2,678
08/11/2013	Philippines (the)	Storm	7,354	16,106,870	10,000,000	08/11/2013	564,058
12/06/2013	India	Flood	6,054	504,473	1,100,000	27/06/2013	150
21/03/2014	Liberia	Epidemic	4,810	10,682	0	14/01/2016	110,139
10/03/2014	Sierra Leone	Epidemic	3,956	14,124	0	/08/2015	151,478
29/06/2015	France	Extreme temperature	3,275	0	0	09/08/2015	0
/02/2014	Guinea	Epidemic	2,544	3,814	0	29/12/2015	5,951
23/01/2016	United States of America (the)	Storm	50	85,000,012	550,000	26/01/2016	0
20/05/2015	India	Extreme temperature	2,248	0	0	31/05/2015	0

(a) Top 10 by severity index

Start date (1)	Country (2)	Disaster type (3)	# casualties (4)	# affected (5)	Damage (in \$) (6)	End date (7)	Total relief (8)
25/04/2015	Nepal	Earthquake	8,969	5,642,150	5,174,000	25/04/2015	865,055
08/11/2013	Philippines (the)	Storm	7,354	16,106,870	10,000,000	08/11/2013	564,058
10/03/2014	Sierra Leone	Epidemic	3,956	14,124	0	/08/2015	151,478
21/03/2014	Liberia	Epidemic	4,810	10,682	0	14/01/2016	110,139
27/04/2017	Yemen	Epidemic	11	180	0	03/06/2017	84,484
/08/2015	Somalia	Drought	0	4,700,000	0	/05/2017	60,666
13/05/2014	Bosnia and Herzegovina	Flood	25	1,000,000	436,580	20/05/2014	58,728
/02/2016	South Sudan	Drought	0	3,600,000	0	/11/2016	36,654
16/04/2016	Ecuador	Earthquake	672	389,364	2,000,000	16/04/2016	26,197
28/09/2016	Haiti	Storm	546	2,100,439	2,000,000	07/10/2016	23,620

(b) Top 10 by total disaster relief

Start date (1)	Country (2)	Disaster type (3)	# casualties (4)	# affected (5)	Damage (in \$) (6)	End date (7)	Total relief (8)
07/06/2014	Argentina	Flood	0	0	62,000	30/06/2014	0
/07/2015	Botswana	Drought	0	0	44,000	/12/2015	0
24/12/2013	Switzerland	Storm	0	0	0	24/12/2013	0
16/10/2016	China	Storm	0	0	890,000	19/10/2016	0
07/05/2016	Dominican Republic (the)	Flood	0	0	0	08/05/2016	0
/08/2015	Panama	Drought	0	0	0	/08/2015	0
07/11/2013	Palau	Storm	0	0	0	07/11/2013	0
/07/2016	Paraguay	Drought	0	0	0	/08/2016	0
02/12/2014	United States of America (the)	Flood	0	0	90,000	05/12/2014	0
28/09/2017	Saint Vincent and the Grenadines	Storm	0	0	0	29/09/2017	0

(c) Bottom 10 by severity

Start date (1)	Country (2)	Disaster type (3)	# casualties (4)	# affected (5)	Damage (in \$) (6)	End date (7)	Total relief (8)
25/04/2015	Nepal	Earthquake	8,969	5,642,150	5,174,000	25/04/2015	865,055
08/11/2013	Philippines (the)	Storm	7,354	16,106,870	10,000,000	08/11/2013	564,058
13/05/2014	Bosnia and Herzegovina	Flood	25	1,000,000	436,580	20/05/2014	58,728
16/04/2016	Ecuador	Earthquake	672	389,364	2,000,000	16/04/2016	26,197
28/09/2016	Haiti	Storm	546	2,100,439	2,000,000	07/10/2016	23,620
08/09/2017	Mexico	Earthquake	328	2,069	2,000,000	08/09/2017	16,058
24/08/2016	Italy	Earthquake	297	7,881	5,000,000	24/08/2016	12,859
12/04/2014	Chile	Wildfire	12	11,000	34,000	21/04/2014	7,598
20/02/2016	Fiji	Storm	44	350,000	600,000	21/02/2016	6,381
05/10/2017	Nicaragua	Storm	15	39,200	0	06/10/2017	3,141

(d) Top 10 precise onset events by total disaster relief

Table 2.A.1: Top and bottom 10 disaster-country events. *Notes.* This table reports the ten worst disaster-country events, from 2013-2017, as ranked i) in descending order of severity as measured by the normalized index of deaths and number of affected described in Section 2.5.2, Panel (a), ii) in descending order of total disaster relief donations generated on Betterplace, Panel (b), iii) in ascending order of the severity index, Panel (c), and, iv) in descending order of total disaster relief donations after removing events with unknown start dates and epidemics, heat waves as well droughts, Panel (d). 41 ties (generated by no deaths and no affected) in iii) are broken randomly.

	# fundraising pages (2)	Requested amount in € (3)	# charities (4)	# German charities (5)
All years	19,474	97,860,498	15,583	12,012
2013	2,642	12,160,062	2,082	1,627
2014	3,263	16,084,026	2,699	1,799
2015	4,725	25,772,442	3,861	2,555
2016	4,912	25,045,182	3,882	3,009
2017	3,932	18,798,787	3,059	3,022

(a) Funraising pages

Year	Mean	S.D.	Median	25^{th}	75^{th}	10^{th}	90^{th}	No. Obs.	Total donations
				Percentile					
	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All	52	215	20	9	50	3	100	680,988	35,187,000
2013	26	117	10	5	20	2	50	126,083	3,234,657
2014	44	180	18	6	38	3	100	88,773	3,900,751
2015	55	203	20	10	50	5	100	137,383	7,615,323
2016	63	234	20	10	50	5	100	148,186	9,361,824
2017	61	266	20	10	50	5	100	180,563	11,074,445

(b) Donations

Table 2.A.2: Yearly summary statistics of demand and supply on Betterplace. *Notes.* Panel (a): This table reports summary statistics about the number and characteristics of fundraising pages created on Betterplace by year and overall. It shows the total number; the total amount requested through these fundraising pages; the total number of distinct charities that created the pages; and the total number of these charities that are headquartered in Germany. Panel (b): This table reports the mean; standard deviation; median; the 10^{th}, 25^{th}, 75^{th} and 90^{th} percentile; the number of individual donations; and the total amount of donations.

	Donation demand			Donation supply
	1[Any fundraising] OLS Full sample	1[Any fundraising] OLS Restricted sample	1[Any fundraising] Logit odds ratios Restricted sample	Log total disaster relief OLS Restricted sample
	(1)	(2)	(3)	(4)
Log # casualties	.02	.022	1.7	.11
	(.007)	(.007)	(.25)	(.045)
Log # affected	.011	.011	1.3	.062
	(.0033)	(.003)	(.087)	(.019)
Log # of tweets	.12	.12	3.3	1.2
	(.017)	(.017)	(.63)	(.18)
Log # of active NGOs	.012	.02	1.8	.075
	(.0047)	(.0059)	(.41)	(.033)
Log of trade with GER		-.01		
		(.0025)		
Distance from capital to Berlin [1000km]		-.0049		
		(.0015)		
Freedom of press [/ 100]		-.014		
		(.029)		
Ease of doing business [/ 100]		.03		
		(.077)		
Corruption perception [/ 100]		.051		
		(.048)		
Basic controls	Yes	Yes	Yes	Yes
Quarter-of-year FE	No	Yes	No	Yes
# positives dep var	67	62	62	51
Pse/Adj R-sq	.2	.22	.42	.28
N. of observations	1720	1604	1604	1604

Table 2.A.3: Correlates of disaster relief fundraising activity and giving on Betterplace, additional specifications I. *Notes.* This table presents additional parameter estimates for Equation (2.1). The dependent variable in columns 1, 2 and 3 is an indicator equal to one if at least one fundraising page for a given natural disaster was created on Betterplace, and zero otherwise. In column 4, the dependent variable is the natural logarithm of total disaster relief for a given event. We show odds ratios from logit models in columns 1, 2 and 3 and OLS estimates in column 4. The estimation sample in column 1 is the full set of natural disasters in the observation window; in columns 2,3 and 4, the sample is restricted to those events for which we observe all covariates included in the specifications of columns 2 and 4 of Table 2.4. See Section 2.3.4 for definitions of variables. We report standard errors clustered at the country level in parentheses.

	Donation demand				Donation supply			
	Log # of fundraising pages OLS		Log relief requested OLS		Log # of donations OLS		1[Received any donations] Logit odds ratios	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log # casualties	.036	.037	.12	.13	.058	.054	1.5	1.5
	(.012)	(.012)	(.049)	(.049)	(.03)	(.03)	(.2)	(.19)
Log # affected	.015	.016	.086	.087	.037	.037	1.3	1.4
	(.0044)	(.0042)	(.026)	(.024)	(.011)	(.01)	(.088)	(.1)
Log # of tweets	.24	.25	1.3	1.3	.7	.72	3.9	5.3
	(.053)	(.056)	(.19)	(.19)	(.12)	(.13)	(.72)	(1.2)
Log # of active NGOs	.013	.029	.086	.17	.025	.065	1.6	2.4
	(.0066)	(.0076)	(.038)	(.046)	(.016)	(.019)	(.3)	(.47)
Log of trade with GER		-.02		-.09		-.046		.63
		(.0037)		(.019)		(.0093)		(.069)
Distance from capital to Berlin [1000km]		-.0058		-.032		-.012		.81
		(.0022)		(.011)		(.005)		(.058)
Freedom of press [/ 100]		.054		.14		.19		.36
		(.039)		(.25)		(.11)		(.65)
Ease of doing business [/ 100]		.16		.057		.098		3.4
		(.1)		(.54)		(.28)		(11)
Corruption perception [/ 100]		.099		.71		.39		1.7
		(.071)		(.4)		(.2)		(4.9)
Basic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter-of-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# positives dep var	56	51	56	51	56	51	56	51
Pse/Adj R-sq	.27	.3	.24	.26	.27	.3	.43	.51
N. of observations	1720	1604	1720	1604	1720	1604	1720	1604

Table 2.A.4: Correlates of disaster relief fundraising activity and giving on Betterplace, additional specifications II. *Notes.* This table presents additional parameter estimates for Equation (2.1) using alternative measures of demand and supply. The dependent variable in columns 1 and 2 is the natural logarithm of the number of fundraising pages created for a given disaster; the natural logarithm of total disaster relief requested by charities across all fundraising pages for a given event in columns 3 and 4; the natural logarithm of the number of disaster relief donations to all fundraising pages for a given event in columns 5 and 6; and an indicator for whether there was at least one donation to any of the fundraising pages devoted to a given event in columns 7 and 8. We show OLS estimates in columns 1 through 6, and odds ratios from logit models in columns 7 and 8. See Section 2.3.4 for definitions of variables. We report standard errors clustered at the country level in parentheses.

	1[Any fundraising]	
	Logit odds ratios	LPM
	(1)	(2)
Log # of NGOs active in 2011	2.1	.017
	(.47)	(.0064)
Log # casualties	1.5	.024
	(.22)	(.0068)
Log # affected	1.5	.0068
	(.11)	(.0018)
Log # of tweets	4.5	.12
	(1.2)	(.021)
Log of trade with GER	.69	-.0082
	(.071)	(.0027)
Distance from capital to Berlin [1000km]	.77	-.0049
	(.062)	(.0016)
Freedom of press [/ 100]	.2	-.032
	(.41)	(.035)
Ease of doing business [/ 100]	3.7	.027
	(11)	(.075)
Corruption perception [/ 100]	.65	.021
	(1.9)	(.05)
Basic controls	Yes	Yes
Quarter-of-year FE	Yes	Yes
Pse/Adj R-sq	.5	.21
N. of observations	1346	1346

Table 2.A.5: Disaster relief fundraising activity—first stage estimates. *Notes.* This table presents the first stage of the 2SLS estimations in Table 2.5. Column 1 is a logit specification whereas column 2 is a linear probability model (OLS). The dependent variable—and thus the endogenous regressor of the analysis in Section 2.6—is an indicator for whether there was any fundraising activity on Betterplace for a given disaster event, and zero otherwise. The exogenous regressor—and thus the instrument—is "Log # of active charities in 2011", the natural logarithm of number of charities active in a country in 2011. "χ^2/F first stage" reports the test statistic of instrument relevance in the first stage. See Section 2.3.4 for definitions of variables. We report standard errors clustered at the country level in parentheses.

	Log total donations to disaster relief		Log # of donations to disaster relief	
	(1)	(2)	(3)	(4)
1[Any fundraising]	6.8	6.5	3.5	3.2
	(0.68)	(1.2)	(0.49)	(0.7)
Log # casualties	-0.059	-0.054	-0.043	-0.036
	(0.026)	(0.034)	(0.018)	(0.02)
Log # affected	-0.013	-0.011	-0.008	-0.0058
	(0.0048)	(0.0088)	(0.0028)	(0.005)
Log # of tweets	0.21	0.24	0.11	0.15
	(0.088)	(0.17)	(0.066)	(0.11)
Log of trade with GER	-0.0084	-0.0096	-0.0053	-0.0068
	(0.0076)	(0.0091)	(0.0052)	(0.0061)
Distance from capital to Berlin [1000km]	0.007	0.006	0.0032	0.002
	(0.0066)	(0.0088)	(0.0038)	(0.0049)
Freedom of press [/ 100]	0.3	0.28	0.22	0.2
	(0.13)	(0.12)	(0.085)	(0.08)
Ease of doing business [/ 100]	0.084	0.089	-0.0037	0.0024
	(0.29)	(0.29)	(0.2)	(0.19)
Corruption perception [/ 100]	0.17	0.16	0.17	0.15
	(0.17)	(0.15)	(0.12)	(0.11)
Basic controls	Yes	Yes	Yes	Yes
Quarter-of-year FE	Yes	Yes	Yes	Yes
First stage	Logit	LPM	Logit	LPM
χ^2 / F first stage	10	5.9	10	5.9
Adj R-sq	0.68	0.69	0.53	0.57
N. of observations	1336	1336	1336	1336

Table 2.A.6: Effect of project creation on disaster relief—without top 10 most severe events. *Notes.* This table presents estimates of the effect of a project creation on the natural logarithm of total disaster relief (columns 1-3), and the natural logarithm of the number of donations to all fundraising pages devoted to an event (column 4-6). The ten most severe natural disasters according to the severity index described in Section 2.5.2 are dropped. The dependent variable is the natural logarithm of total disaster relief for a given event in columns 1 and 2, and the natural logarithm of the number of disaster relief donations to all fundraising pages for a given event in columns 3 and 4. We instrument for "1[Any fundraising]" using the natural logarithm of the number of charities present in a country in 2011. Columns 1 and 3 show estimates from the two-step procedure described in Section 2.6.1; columns 2 and 4 show estimates from a standard 2SLS procedure. "χ^2/F first stage" reports the test statistic of instrument relevance in the first stage. See Section 2.3.4 for definitions of variables. We report standard errors clustered at the country level in parentheses.

	Log total donations to disaster relief		Log # of donations to disaster relief	
	(1)	(2)	(3)	(4)
1[Any fundraising]	5.2	5.9	2.2	2.6
	(0.42)	(0.98)	(0.23)	(0.46)
Log # casualties	-0.03	-0.046	-0.021	-0.031
	(0.016)	(0.033)	(0.01)	(0.018)
Log # affected	-0.0052	-0.0095	-0.0023	-0.0047
	(0.0035)	(0.0076)	(0.0016)	(0.0036)
Log # of tweets	0.11	0.061	0.053	0.027
	(0.062)	(0.11)	(0.038)	(0.057)
Log of trade with GER	-0.0027	-0.0014	-0.001	-0.00024
	(0.0061)	(0.0065)	(0.0032)	(0.0033)
Distance from capital to Berlin [1000km]	0.0067	0.0087	0.0032	0.0043
	(0.0053)	(0.0072)	(0.0023)	(0.0032)
Freedom of press [/ 100]	0.094	0.16	0.057	0.097
	(0.1)	(0.12)	(0.044)	(0.054)
Ease of doing business [/ 100]	-0.02	-0.0066	-0.03	-0.022
	(0.21)	(0.22)	(0.094)	(0.1)
Corruption perception [/ 100]	0.11	0.15	0.08	0.1
	(0.12)	(0.12)	(0.067)	(0.07)
Basic controls	Yes	Yes	Yes	Yes
Quarter-of-year FE	Yes	Yes	Yes	Yes
First stage	Logit	LPM	Logit	LPM
χ^2 / F first stage	11	6.1	11	6.1
Adj R-sq	0.7	0.67	0.61	0.55
N. of observations	1336	1336	1336	1336

Table 2.A.7: Effect of project creation on disaster relief donations—without top 10 events with largest donations. *Notes.* This table presents estimates of the effect of a project creation on the natural logarithm of total disaster relief donations (columns 1-3), and the natural logarithm of the number of donations to all fundraising pages devoted to an event (column 4-6). The ten events that received most disaster relief in the sample are dropped. The dependent variable is the natural logarithm of total disaster relief for a given event in columns 1 and 2, and the natural logarithm of the number of disaster relief donations to all fundraising pages for a given event in columns 3 and 4. We instrument for "1[Any fundraising]" using the natural logarithm of the number of charities present in a country in 2011. Columns 1 and 3 show estimates from the two-step procedure described in Section 2.6.1; columns 2 and 4 show estimates from a standard 2SLS procedure. "χ^2/F first stage" reports the test statistic of instrument relevance in the first stage. See Section 2.3.4 for definitions of variables. We report standard errors clustered at the country level in parentheses.

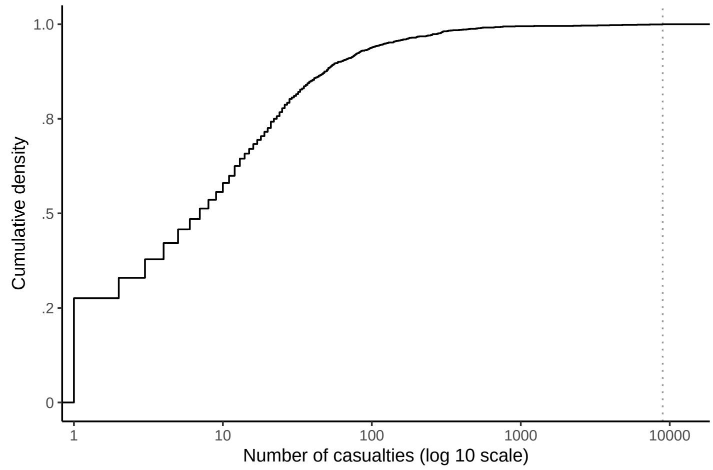
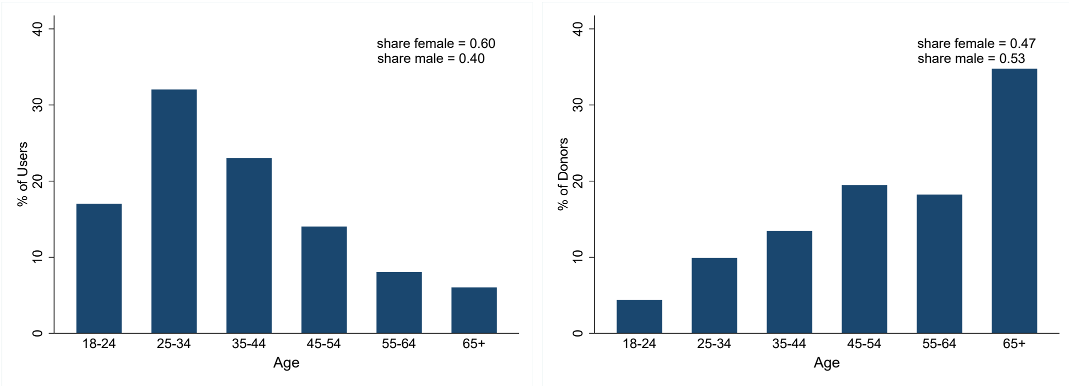
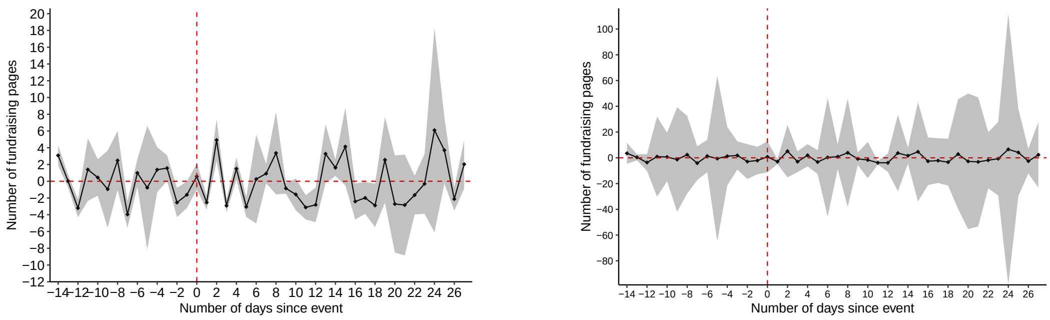


Figure 2.A.1: Casualties from natural disasters. *Notes.* This figure depicts the empirical cumulative density of the number of casualties across the 1,720 events in the sample. The most severe event is marked by the dotted vertical line—the earthquake in Nepal in 2015 which caused 8,969 casualties.



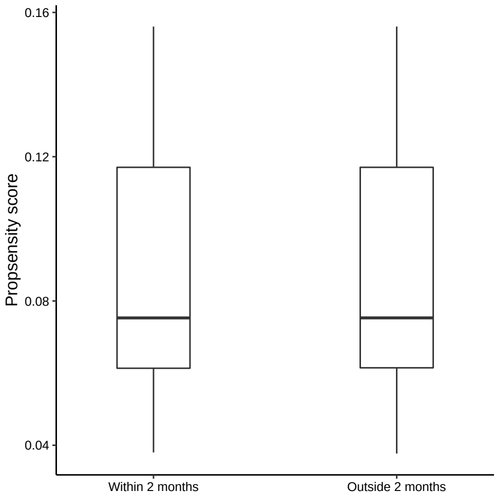
(a) Donors on Betterplace (b) Donors in German Philanthropy Market

Figure 2.A.2: Users on Betterplace vs. donor population in Germany. *Notes.* Panel (a) shows a bar chart with the age distribution of users on Betterplace and the share of donors that are female or male. The data comes from Google analytics and is i) only available for June through October 2018, and ii) only for 51% of users. Panel b) shows a bar chart with the age distribution of donors in Germany and the share of donors that are female and male. This figure is based on data from the 2015 wave of the German Socio-Economic Panel Study (G-SOEP).

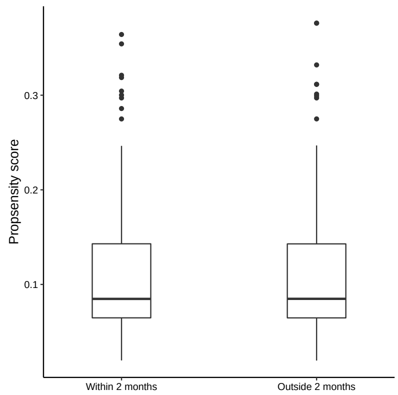


(a) Top 10 events (b) Top 2 events

Figure 2.A.3: Crowding out of demand by the top 10 and top 2 events. *Notes.* Panels (a) and (b) depict the daily fundraising activity, measured as the number of new non-disaster relief fundraising project pages posted on the platform after a disaster event. In Panel (a) we consider the ten events that raised most disaster relief donations on Betterplace; in Panel (b) we consider the top two events. The underlying data structure and the model are described just before and in the beginning of Section 2.5.1. The error bands show the 95% confidence interval.

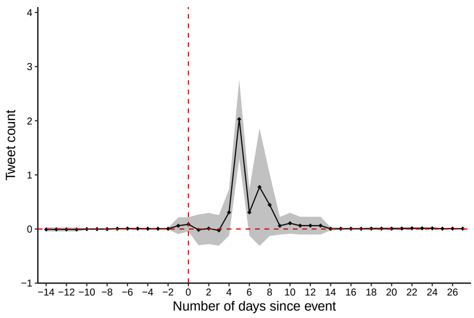


(a) 1:5 matching, severity

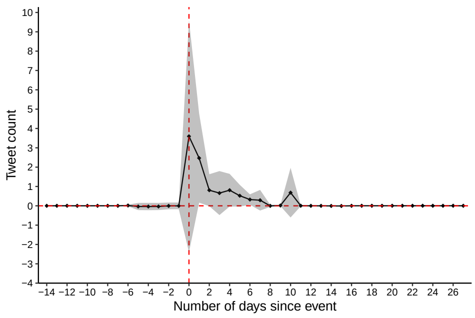


(b) 1:1 matching, full covariates

Figure 2.A.4: Matched event sample balance. *Notes.* This figure shows boxplots of the propensity scores in the matched samples of events. Panel (a) results from the matching procedure matching five controls to each treated event, controlling only for event severity. Panel (b) results from the matching procedure from matching one treated observation to five control observations controlling for the extensive set of covariates used in column 2 of Table 2.4. We detail the matching procedure in Section 2.5.2. In the boxplot, the bold solid bar shows the median of the distributions, while the lower and upper hinges correspond to the $25th$ and 75^{th} percentile. The upper (lower) whisker extends from the hinge to the largest (smallest) value no further than 1.5 times the inter-quartile range from the hinge. Data beyond the end of the whiskers are called "outlying" points and are plotted individually.

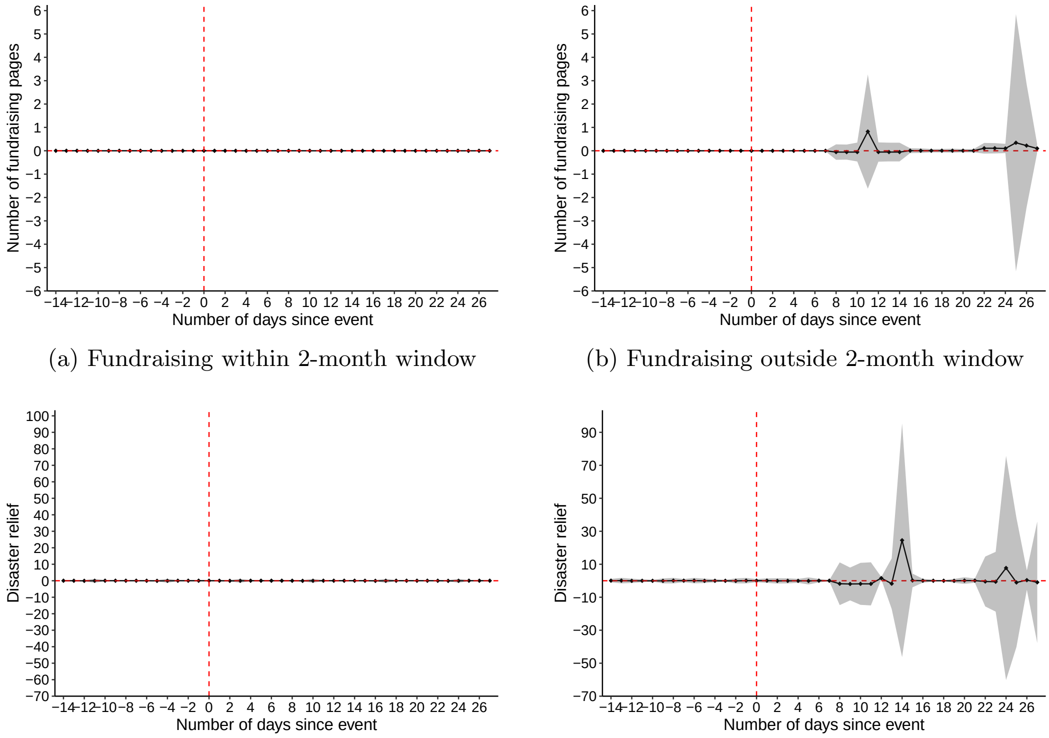


(a) Within two month window



(b) Outside two month window

Figure 2.A.5: Media fatigue. *Notes.* This figures compares media coverage for events within two months—Panel (a)—of Typhoon Haiyan and the Nepal earthquake to that of a matched sample of events outside that two month window—Panel (b). We measure media coverage by the daily number of tweets related to a disaster (see Section 2.3.4). The matched sample contains five control events for each treated event using propensity scores based on events' severity. The procedure is detailed in Section 2.5.2. The underlying data structure and the model are described just before and in the beginning of Section 2.5.1. The error bands show the 95% confidence interval.



(a) Fundraising within 2-month window

(b) Fundraising outside 2-month window

(c) Donations within 2-month window

(d) Donations outside 2-month window

Figure 2.A.6: Disaster fatigue: Fundraising and donations to other disasters (1:1 matched sample) *Notes.* This figures compares fundraising activity for and disaster relief donations to events within two months—Panels (a) and (c)—of Typhoon Haiyan and the Nepal earthquake to that of a matched sample of events outside that two month window—Panels (b) and (d). Panels (a) and (b) show fundraising activity, measured as the number of new disaster relief fundraising project pages posted on the platform in response to disaster events. Panel (b) shows the daily disaster relief donations response, measured as total donations going to disaster relief fundraising projects. The matched sample contains one control event for each treated event using propensity scores based the extensive list of covariates used in column 2 of Table 2.4. The procedure is detailed in Section 2.5.2. The underlying data structure and the model are described just before and in the beginning of Section 2.5.1. The error bands show the95% confidence interval.

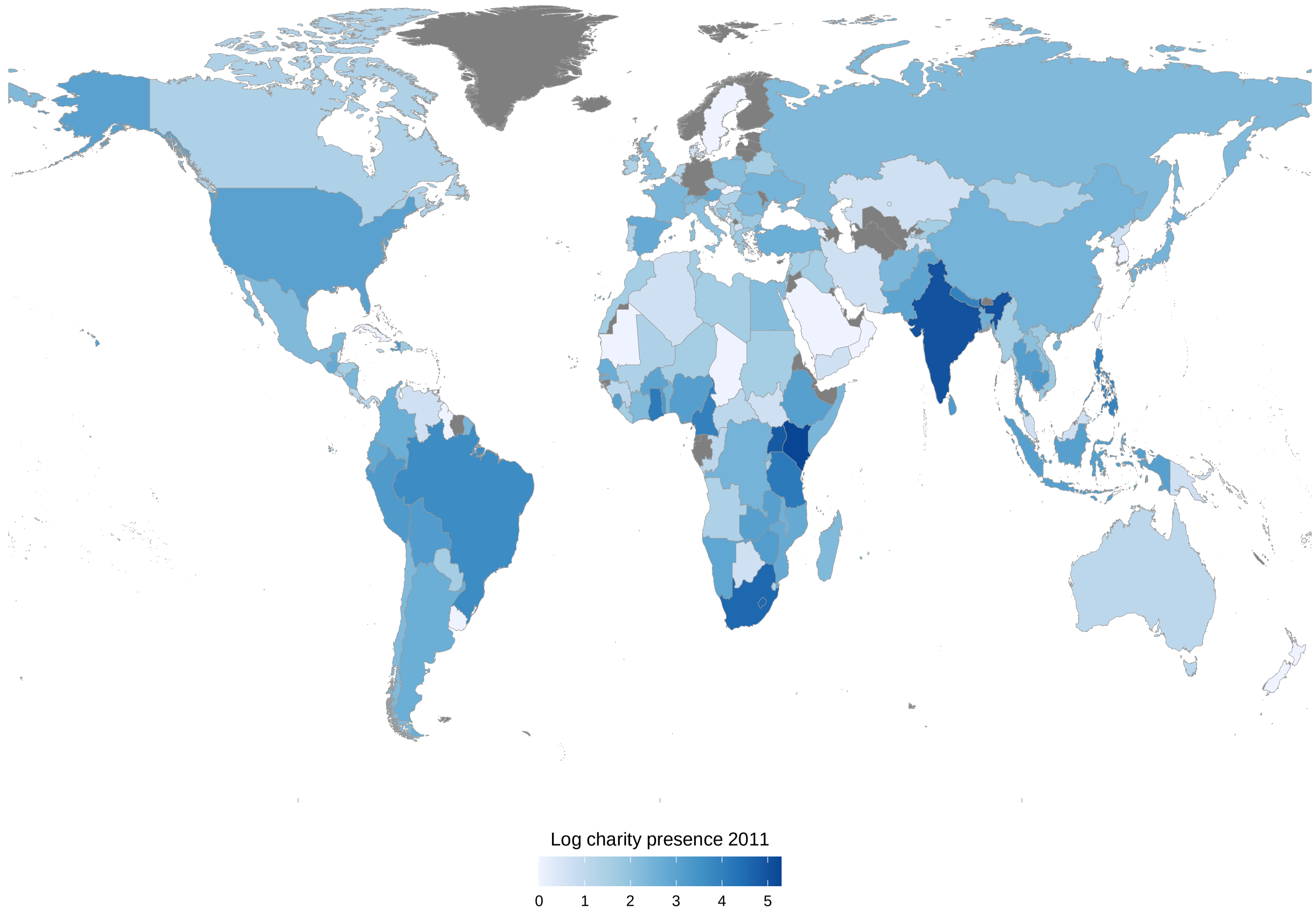


Figure 2.A.7: Charity presence in 2011. *Notes.* This map depicts variation in the natural logarithm of charity presence across countries in 2011, the instrumental variable we employ in Section 2.6.1. Lighter shades of blue indicate lower charity presence; darker shades indicate a higher charity presence.

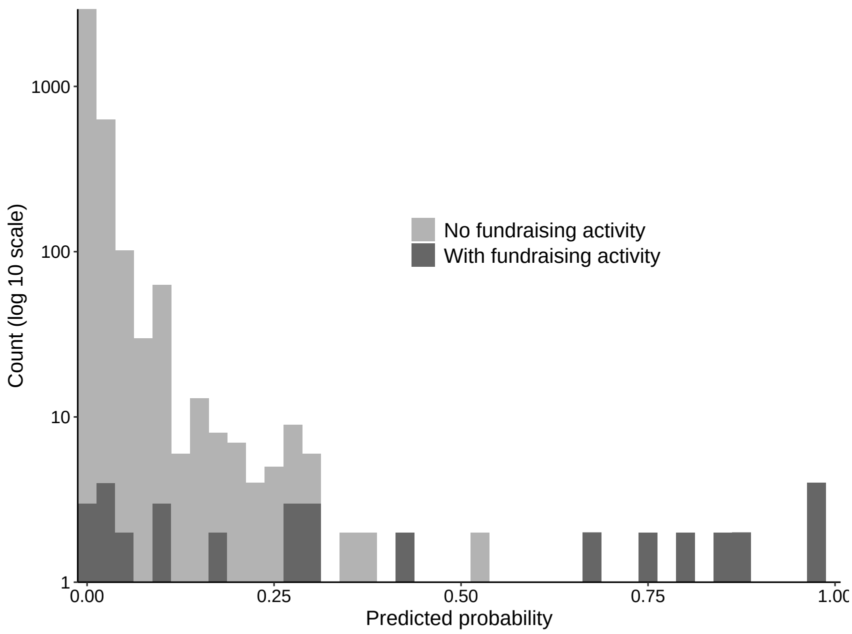


Figure 2.A.8: First stage predictions. *Notes.* This figure shows the predicted probabilities of there being any fundraising activity in the wake of a natural disaster on Betterplace. The estimates are obtained from the logit regression in the first step of the IV procedure described in Section 2.6.1. The histogram is constructed using a fixed bin width of .025. Note that the y-axis is shown in log-10 scale.

Appendix B to Chapter 2: Matching Projects to Disasters

This section provides a detailed description of the procedure used to match fundraising pages from Betterplace to disaster events in the EMDAT database. Upon registration of the project page, the charitable organization provides detailed information on the project. Firstly, each charity describes the goal of the project in the project description, which should entail a detailed description of the project's aims, the project plan, desired project volume, the time frame and the project location. Other data entries that are required are details regarding the charity itself, such as location, and name. Additionally, the charities can create "needs", which are project elements to which donors can donate and which are earmarked for a particular element of the project. This data is entered by the charities and Betterplace stores the most recent data entry. Because funding is earmarked and project managers have to clearly mark changes in their project descriptions, major changes in project aims tend to be rare.[1] Betterplace also stores the creation date of the project, the date the project became visible on the platform, which happens when the project has received its first €250 in donations.

The project location is not always the same as the location of the charitable activity. This is because some projects target multiple countries, or for some projects the location is indicated to be the charity's headquarters although activities take place in another country. For example, a fundraising page could be hosted in Germany but utilize the funds to build a water purification system somewhere in Africa. The majority of donations, over €20 million, go to projects hosted in Germany, followed by Kenya and Nepal, each receiving around €1 million Euro. In terms of donation volume per target country, Germany still receives the bulk of donations but now Syria follows with about €1.75 million. Overall, 59.3% of donations on Betterplace go to projects hosted in Germany, but only 56.9% of these donations are used for causes located in Germany. To match projects to disaster events, we use the country where the project is carried out.

We match all projects which were visible on Betterplace from 2013-2017, and all projects created from 2013-2017. This comprises projects that were created prior to

[1] Charities can change the project description, although these changes have to be clearly marked and donations received before the change have to be used for the purpose previously stated. According to Betterplace, due to these regulations projects rarely change their focus completely, although new elements may be added and old ones removed. Additionally, the costs of creating a new project are low and it is therefore more likely that charities create new projects to support new causes instead of changing old projects.

2013, too. In total there are 24,151 projects, which have all been classified, regardless of whether the project received any donations. The projects were distributed in four equally sized batches (of around 6,000 projects), and this was done twice with random assignment of projects to each batch. Each of the eight batches thus contained a different combination of projects and each project was included in exactly two separate batches. The information provided in these batches was the project description, the location and the charity's name.

To physically match fundraising pages to events, we contracted four individuals. These people were unaffiliated to the university or the research team in any way. The contractors were paid a fixed salary, and conducted the matching process according to the following steps:

Step 1

Each batch is assigned to a different individual. The contractor carefully reads through and classifies projects using a broad set of categories. These categories are natural disaster, conflict, refugee and other. For the main categories "other" and "refugees", a more detailed subcategory had to be selected using a drop-down menu.

Step 2

When the project was related to natural disasters, additional information had to be completed to capture disaster-specific characteristics. This was done in order to later aid the matching process with disaster entries in the EMDAT database. First, when host and target country differed, the country code of the country where the disaster happened had to be recorded (see above). If the project targeted multiple countries an additional indicator variable was recorded. For example, a fundraising page could collect funds for Ebola epidemic relief in the Horn of Africa. Second, the disaster type had to be chosen from a set of predefined disaster types that matched the disaster types in the EMDAT database. If applicable, the name of the disaster was entered, e.g. Typhoon Haiyan in the Philippines.

Step 3

Since all projects were classified by two separate individuals, the classifications were cross-checked. Whenever classifications differed (on the basis of the main category, but also differences in disaster characteristics) the project was carefully read again by a third person to decide upon the final classification. In total there were 2113 projects where classifications did not agree and were classified by a third person.

Step 4

In the final step, the disaster relief projects were matched to entries in the EM-DAT database, based on the location stamp, time stamp, disaster type and name (if applicable). Some projects targeted multiple disasters at the same time. For example, the Nepal earthquake on April 25^{th}, 2015 was followed by another major earthquake on May 12^{th}, 2015; several projects mentioned both in their descriptions. These events are recorded as separate disasters in the EMDAT database and therefore these projects were assigned two disaster matches. In general, when projects targeted multiple disasters at once, they were matched with all respective event entries. Some projects, such as the red cross, were classified as general disaster relief. Individual donating to this type of projects support the operations of these charities without funding being earmarked. Often, these charities also have separate projects to target specific events.

Appendix C to Chapter 2: Using regularization to select predictive covariates

The analysis presented in Table 2.4 is complicated by two characteristics of the data. First, the response data is skewed in the sense that only 3.9% of projects generated a fundraising response in the first place. A predictive model which would predict "no project" for all events would thus tend to be correct in over 96% of cases. Second, the list of covariates presented in Table 2.4 could be further extended to include various sub-indices of the measures of countries' socio-political climate. In the regressions we include variables which, based on previous literature, seem pertinent in explaining heterogeneity.

In this section, we discuss an alternative approach to understanding the heterogeneity in responses to events. We employ regularized (also referred to as penalized regression) which i) may outperform standard OLS in terms of predictive performance as it avoids over-fitting, and, ii) can be employed to select the most predictive covariates. Since we are not interested in making out of sample predictions, the latter rationale motivates our use of regularized regression. In particular, we would like to answer the following question: Given the wide set of possible correlates of supply and demand of charitable giving, which ones are most predictive of there being fundraising activity for an event on Betterplace?

Estimation approach

Specifically, we employ the lasso—least absolute shrinkage and selection operator—estimator (Tibshirani, 1996). In the simple case of a linear model, the lasso estimator can be expressed as the solution to the following optimization problem:

$$min_\beta \sum_{i=1}^{N}(Y_i - X_i\beta)^2 + \lambda \sum_{k=1}^{K} |\beta_k|$$

where i indexes event and K the number of variables, i.e., the number of columns in the matrix X. The first term corresponds to the sum of squared residuals, which is the standard OLS objective function. The second term however penalizes the sum of absolute coefficient estimates using a parameter λ. Given a value of λ, larger coefficients will lead to a higher penalty. Similarly, given coefficient estimates a higher value of λ will cause a stronger penalty. Note that a variable x_k is only "used" to explain variation in Y if $\beta_k \neq 0$. Thus, the lasso balances the increasing explanatory power from including more variables—achieved by reducing the sum of squared residuals—with a higher penalty incurred as a consequence of more variables having non-zero coefficients. The key advantage of the lasso estimator over other regularization techniques is the fact it will shrink estimates to exactly zero, rather than towards zero. Therefore, the vector $\hat{\beta}$ which minimizes the optimization problem above is likely to have a number of entries that are exactly zero. This indicates little explanatory power of the respective variables. This approach can be extended to other estimation methods, and indeed, we will apply it to a standard logit model as well.

One key aspect of lasso estimation is the degree of regularization, i.e., the value of λ. We choose λ through tenfold cross validation. To ascertain that each validation fold contains a non-zero number of "yes" events, we apply the following procedure. First, split the set of events into those with fundraising activity and those without. Second, randomly sample with replacement integers from one through ten within both subsets. Third, recombine those subsets. Now, each event (regardless of project status) is associated with exactly one value of sequence from one through ten which denotes its cross validation fold membership. The cross validation then proceeds by in turn using nine folds to estimate a number of lasso models across a wide range of candidate values for λ, and evaluates their predictive performance using the tenth, held out, fold as the test set. In the end, the value of λ which achieves minimum out of sample error is the one that is chosen.

We consider three models using lasso estimation. We apply a lasso to a linear probability model (outcome is whether or not fundraising activity takes place) and use mean squared error as the cross validation criterion. Moreover, we apply a lasso to the analogous logit model and use the misclassification rate as the cross validation criterion. Finally, we apply a lasso to the linear model where the outcome is the log of total disaster relief, and evaluate using mean squared error.

Results

We first compare the predictive performance of different estimators as to their ability to classify events into whether fundraising activity took place. Specifically, we compare i) simple OLS (LPM), ii) standard logit, iii) lasso OLS, and iv) lasso logit. We consider a prediction to be a "yes" when $Pr(Y = 1|X) > .5$. Note that we include all available covariates in X of which those considered in Table 2.4 are a strict subset. In Table 2.C.1, we present the predictive power of various models vis-a-vis fundraising activity and total amount donated. The results demonstrate that the logit model outperforms the LPM by a considerable margin. Further, lasso estimators perform comparatively well despite their (mechanical and desired) sparsity but are not able to capture the complexity as accurately as the standard logit model. Overall, the models find it difficult to accurately predict "yes fundraising activity" given the long tail of smaller disasters, of which only a tiny fraction received funding. We take these results as reason to consider logit models the preferred first stage specification in the IV analysis of Section 2.6.1.

Recall that these models differ from those in Table 2.4 in that they use even more covariates. For comparison, the model in column 2 of Table 2.4 correctly predicts 24 "yes", and 1534 "no", out of 1604 cases while the model with additional covariates correctly predicts 25 "yes" and also 1534 "no". Hence, the full set of covariates contains elements which help the model make one more correct prediction.[2]

Despite not being able to capture the complexity of disaster relief demand as well as the standard logit model, the lasso results are informative since they point out the most predictive covariates. In Figure 2.C.1, we present the coefficients that the lasso estimator did *not* shrink to zero, i.e. those most helpful ones in predicting a demand response on Betterplace.[3] Three key findings emerge from this analysis. First, the severity of the event is the most powerful predictor a response, and the number of affected matters more than the number of casualties. Second, existing charity presence on the ground matters and is strongly predictive of demand for disaster relief. Third, media attention is also strongly predictive of a demand response. Finally, the analogous exposition for predictive covariates of donation supply indicates that media attention is the most powerful predictor while severity in terms of the number of affected is second most important.

[2] Note that the results displayed in Table 2.1, Table 2.A.4, and Table 2.A.3 would not substantively change if those additional covariates were included. The magnitudes and signs as well as the estimates' precision are hardly affected. These results are not reported and are available on request. The IV analysis in Section 2.6.1 would not be affected in a substantive way either.

[3] The lasso estimation also included the quarter-of-year dummies, which are not included in the exposition. Some of those were not shrunk to zero but are not reported in Figure 2.C.1.

	Target variable: 1[Any fundraising]					
			Loss: MSE		Loss: misclassification error	
	OLS (1)	Logit (2)	Lasso OLS - min (3)	Lasso OLS - 1 SD (4)	Lasso logit - min (5)	Lasso logit - 1 SD (6)
Predicts no (n = 1542)	1596	1570	1597	1604	1576	1599
Fraction correct	1	0.99	1	1	1	1
Predicts yes (n = 62)	8	34	7	0	28	5
Fraction correct	0.13	0.40	0.11	0	0.35	0.080
Overall accuracy	0.97	0.97	0.97	0.96	0.97	0.96
	Target variable: log total amount donated					
	Loss: mean squared error (MSE)					
Mean squared error	3.22		1.26	1.44		

Table 2.C.1: Model performance. *Notes.* This table compares the models estimated in Sections 2.4.3 to those estimated using lasso regression. The top part of the table evaluates models in which the dependent variable is a binary indicator equal to one if there was any fundraising activity in response to a natural disaster on Betterplace, and zero otherwise. The dependent variable in the bottom part is the natural logarithm of total disaster relief for an event on Betterplace. "'Lasso - min" denotes proceedings from lasso models in which the penalty that minimizes the loss function is chosen via cross validation. "Lasso - 1 SD" is the model implied by the penalty term within one standard deviation of the loss function minimizing penalty term, that results in the lowest value of the loss function. In the top half, "Predicts no/yes" counts the number of cases a model predicts no/some fundraising activity takes place after an event. "Fraction correct" is the fraction correctly predicted among those events without activity. "Overall accuracy" is the total fraction of correctly predicted cases using a threshold of .5. "MSE" reports the implied mean squared error for the respective models across cases.

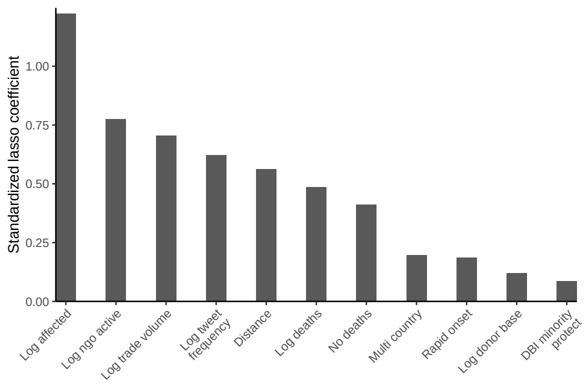


Figure 2.C.1: Non-zero coefficients in lasso models. *Notes.* This figure reports non-zero coefficients from a lasso estimator with the dependent variable being a binary indicator equal to one if there was any fundraising activity in response to a natural disaster on Betterplace, and zero otherwise. Refer to the text of Appendix Section 4.6 for the full set of predictors included. All variables were normalized and the coefficients should not be interpreted numerically. The lasso penalty was chosen via tenfold cross validation. These coefficients are obtained from the "Lasso - min" model described in the notes to Table 2.C.1.

Appendix A to Chapter 3: Construction of outcome indices

Table 3.1 provides an overview of the hypotheses. In the following, we detail which variables are used to construct those indices. Note that we spell out winsorization and transformation in Section 3.4.3, the creation of indexes based on z-scores in Section 3.3.2.

1. Entrepreneurship training study

1.1. Economic outcomes (four hypotheses)

1 *Business creation*
- Business exists (yes/no)
- Average hours contributed by the hour per week

2 *Business success*
- Monthly profits
- Monthly sales

3 *Capital and labor input*
- Value of physical assets
- Value of inventory
- Capital investment over past 3 months
- Number of full-time employees
- Number of part-time employees
- Number of partners in business

4 *Economic self-sufficiency*
- Earnings from self-employment (monthly profits)
- Earnings from wage employment
- Earnings from other sources

1.2 Business and personal input (eight hypotheses)

1 *Business practices*
- Share of business practices employed

2 *Financial professionalization*
- Taken out a loan (yes/no)
- Size of loan
- Business registration
- Local trade licenses
- Knowledge about funding initiatives

- Actual funding from initiatives
- Received equity investment
- Banking account
- Emergency borrowing
- Business banking account
- Hours of consulting services

3 *Marketing*

- Number of marketing channels used

4 *Innovation*

- Introduction of a new product (yes/no)
- Number of new products
- Main new product is a new product line (yes/no)
- Product improvement (yes/no)
- Product new to neighborhood (yes/no)
- Origin of idea (own idea vs. inspired vs. purchased/others idea)
- Process improvement (yes/no)
- Introduced a new method for pricing (yes/no)
- Website with functioning URL (yes/no)

5 *Networks*

- Number of contacts in friends and family
- Number of contacts in "other"
- Scope of potential advice
- Scope of advice used
- Number of business partners

6 *Entrepreneurial mindset*

- Personal initiative
- Aspirations
- Entrepreneurial self-efficacy (general and task-specific, separately)
- Entrepreneurial future

7 *Owner's non-cognitive traits*
- Big-5
- Grit

8 *Preferences*
- Risk preferences
- Subjective risk preferences
- Loss aversion
- Time preferences
- Subjective time preferences

2. Selection study

2.1 Selection into entrepreneurship among those with interest (four hypotheses)

1 *Submitted application*

2 *Cognitive ability*
- Number of correcly solved Raven's matrices

3 *Over-confidence*
- Over-estimation
- Over-placement

4 *Entrepreneurial self-assessment*
- Believes about becoming a successful entrepreneur,
- Subjective rank of entrepreneurial ability,

2.2.1 Economic outcomes (non-experimental) [*identical to 1.1*]

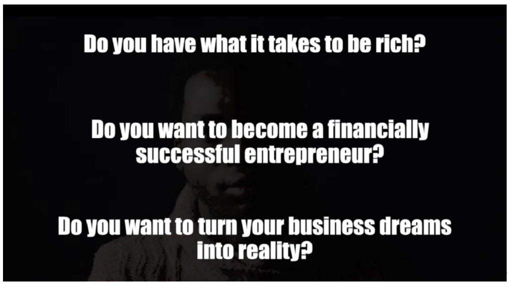
2.2.2 Business and personal input (non-experimental) [*identical to 1.2*]

Appendix B to Chapter 3: Marketing themes

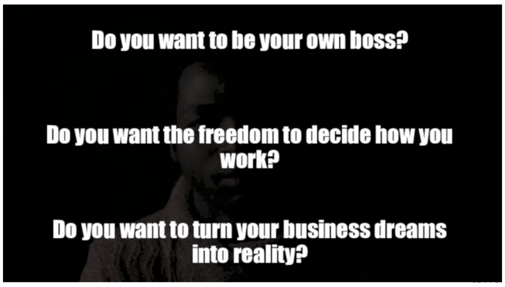
Section 3.2.2 describes how our design allows us to study selection into entrepreneurship. In order to apply to the entrepreneurship training program, students ought to attend information sessions where application forms can be obtained. We randomly vary the content of those information sessions by emphasizing either that entrepreneurship offers the possibility of achieving *financial independence*, or that entrepreneurship offers the *freedom to be creative*. Information sessions take approximately 15-20 minutes and the content is presented by a member of our partner organization. In each session, a presenter went through 12 presentation slides and two videos.

The videos constituted the main source of variation in the presentation. This guaranteed that students across sessions are exposed to the identical content. The first video differed in both visual and audio content. It was was 3 minutes 57 seconds in the profit condition, and 3 minutes 38 seconds in the creative freedom condition. The difference stems from the voice over being longer in the former. The second video only differed in audio content, and took 1 minute 53 seconds in both treatment conditions. Videos were embedded in the presentations to reduce technological complexity. The first video was presented on slide seven, while the second video was presented on the last slide. In between, slide nine presented different content.

In Figure 3.B.1 we show examples of different content across the two treatments. In panels (a) and (b), we show a still frame of the first video's first slide. Two of three statements differ, and the voice over emphasized the differences between the two treatments. Note that *not* entire presentation was kept in this black and white layout. In panels (c) and (d) we show the first frame of the second video. Again, the voice over emphasized the differences. Finally, we show the slide in which the presentations further differed in panels (e) and (f).



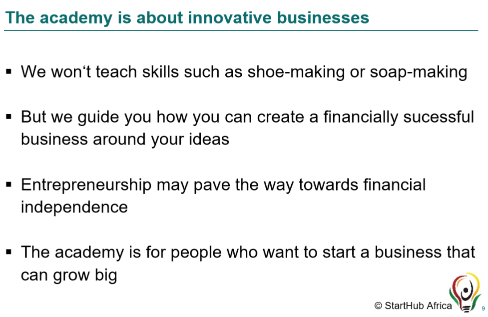
(a) First video, profit



(b) First video, creative freedom

(c) Second video, profit

(d) Second video, creative freedom



(e) Slide nine, profit

(f) Slide nine, creative freedom

Figure 3.B.1: Example for treatment variation in information sessions. *Notes.* The figures illustrates three instances of how the information session presentation differed between profit and creativity treatment. Panels (a), (c) and (e) show slides that emphasize the profit motive of becoming an entrepreneur. Panels (b), (d) and (f) show the corresponding slide in the creative freedom treatment. The images (a) through (d) are obtained from screen shots of videos that were shown during the information sessions. Note that students were able to only see one deck of slides—either the profit or the creativity theme.

Appendix A to Chapter 4: Additional tables and figures.

	Mean (1)	S.D. (2)	Median (3)	25^{th} (4)	75^{th} (5)	N (6)
# employees	116	114	85	56	130	463
Sales ['000 EUR]	28639	74308	10000	5000	21941	289
Year plant opened	1970	25	1976	1961	1986	456
% for export	27	28	15	2	45	438
Produces consumer good	.5	.5	1	0	1	458
Produces intermediate good	.29	.45	0	0	1	458
Produces capital good	.22	.41	0	0	0	458
Shared ownership	.67	.47	1	0	1	463
Limited liability	.27	.44	0	0	1	463
Other ownership	.063	.24	0	0	0	463

Table 4.A.1: Summary statistics of firms survey characteristics. *Notes.* This table reports summary statistics of survey-level variables used in the analysis of correlates of firm's Style 2 intensity of Table 4.3. Column 2 reports the standard deviation, while columns 4 and 5 report the 25^{th} and 75^{th} percentile, respectively. "% for export" is a firm's self-reported share of output that is exported abroad. "Produces consumer/intermediate/capital good" are indicator variables equal to one when the firm produces the respective output category, and zero otherwise. "Shared ownership", "limited liability" and "Other ownership" are indicators equal to one when a firm is organized according to the respective ownership structure.

	Firm productivity 2001 to 2006 95% winsorized			Firm productivity 2001 to 2006 not winsorized		
	(1)	(2)	(3)	(4)	(5)	(6)
Mgt style 2	.3	.24	.25	.3	.23	.24
	(.092)	(.093)	(.093)	(.11)	(.11)	(.11)
1[consumer good]			.071			.13
			(.078)			(.098)
1[intermediate good]			.18			.24
			(.082)			(.095)
Sector FE	No	Yes	Yes	No	Yes	Yes
Region FE	No	Yes	Yes	No	Yes	Yes
Adj R-squared	.029	.16	.17	.021	.13	.15
N. of cases	288	288	285	288	288	285

Table 4.A.2: Management style and firms' TFP before the crisis—full panel structure. *Notes.* This table reports the results of estimating Equation (4.4) using OLS, and limiting the sample to those firms for which the full panel to estimate TFP is available—see Section 4.4.1. "Mgt style 2" is a variable between 0 and 1 and indicates style 2 intensity. "1[consumer good]" and "1[intermediate good]" are indicators for firms that are located in the respective location along the value chain. The omitted category is firms producing capital goods. Columns 2,3,5 and 6 contain sector and region fixed effects. Standard errors clustered at the three-digit industry level are reported in parentheses.

	Firm productivity 2001 to 2006 95% winsorized			Firm productivity 2001 to 2006 not winsorized		
	(1)	(2)	(3)	(4)	(5)	(6)
$1[\text{sytle } 2 > \frac{2}{3}]$	.19	.15	.16	.19	.15	.16
	(.056)	(.053)	(.053)	(.06)	(.059)	(.06)
$1[\frac{1}{3} < \text{sytle } 2 \leq \frac{2}{3}]$	.09	.11	.1	.094	.12	.11
	(.05)	(.049)	(.049)	(.055)	(.056)	(.056)
1[consumer good]			.065			.097
			(.065)			(.08)
1[intermediate good]			.12			.16
			(.064)			(.073)
Sector FE	No	Yes	Yes	No	Yes	Yes
Region FE	No	Yes	Yes	No	Yes	Yes
p: mid vs top tertile	.068	.37	.28	.097	.52	.41
Adj R-squared	.02	.13	.13	.015	.1	.11
N. of cases	385	385	379	385	385	379

Table 4.A.3: Management style tertiles and firms' TFP before the crisis. *Notes.* This table reports the results of estimating Equation (4.4) using OLS. The dependent variable is a firm's estimated TFP; 95% winsorized in columns 1-3 and non-winsorized in columns 4-6. The indicators bin management style intensity into tertiles; the omitted category is the bottom third of style 2 intensity. "1[consumer good]" and "1[intermediate good]" are indicators for firms that are located in the respective location along the value chain. The omitted category is firms producing capital goods. Columns 2,3,5 and 6 contain sector and region fixed effects. Standard errors clustered at the three-digit industry level are reported in parentheses.

	Mean (1)	S.D. (2)	Median (3)	25^{th} (4)	75^{th} (5)	N (6)
Sales 2001-2006	17558	29778	9768	5332	19386	499
Log sales 2001-2006	9.2	.98	9.2	8.6	9.9	499
Total assets 2001-2006	15563	22062	8520	4374	17428	505
Log total assets 2001-2006	9.1	1.1	9.1	8.4	9.8	505
# employees 2001-2006	102	84	77	53	118	452
Log # employees 2001-2006	4.4	.73	4.3	4	4.8	452
Sales 2007-2010	19554	29809	10182	4813	21064	473
Log sales 2007-2010	9.2	1.1	9.2	8.5	10	473
Total assets 2007-2010	21371	39522	10600	4754	23869	479
Log total assets 2007-2010	9.2	1.3	9.3	8.5	10	479
# employees 2007-2010	100	89	73	47	124	439
Log # employees 2007-2010	4.3	.84	4.3	3.9	4.8	439

Table 4.A.4: Summary statistics of TFP inputs. *Notes.* This table provides summary statistics for (time aggregated) inputs to the TFP estimation. The variables are averaged over the corresponding time horizon. “25^{th}” and “75^{th}” denote the respective percentile of the distribution. “Log” refers to the natural logarithm.

	Firm productivity (2007 to 2010)-(2001-2006) 95% winsorized			Firm productivity (2007 to 2010)-(2001-2006) not winsorized		
	(1)	(2)	(3)	(4)	(5)	(6)
Mgt style 2	-.14	-.13	-.13	-.1	-.092	-.09
	(.069)	(.064)	(.066)	(.081)	(.077)	(.077)
Pre-crisis TFP	-.36	-.4	-.38	-.38	-.42	-.4
	(.061)	(.066)	(.061)	(.09)	(.088)	(.084)
1[consumer good]			-.018			-.012
			(.057)			(.062)
1[intermediate good]			.026			.024
			(.053)			(.056)
Sector FE	No	Yes	Yes	No	Yes	Yes
Region FE	No	Yes	Yes	No	Yes	Yes
Adj R-squared	.2	.27	.28	.22	.28	.29
N. of cases	341	341	336	341	341	336

Table 4.A.5: Management style and difference in firm productivity. *Notes.* This table shows results of regressions in which the outcome is the difference between a firm's productivity calculated from 2007-2010 data and its productivity calculated from 2001-2006 data. Both quantities are 95 percent winsorized prior to calculating the difference. "Mgt style 2" is a variable between 0 and 1 and indicates style 2 intensity. "Pre-crisis TFP" is a firm's productivity calculated from 2001-2006 data. "1[consumer good]" and "1[intermediate good]" are indicators for firms that are located in the respective location along the value chain. The omitted category is firms producing capital goods. Standard errors clustered at the three-digit industry level are reported in parentheses.

	Quantile regression			Tobit models		
	Firm productivity 2007 to 2010 95% winsorized			Firm productivity 2007 to 2010 not winsorized		
	(1)	(2)	(3)	(4)	(5)	(6)
Mgt style 2	-.18	-.17	-.19	-.081	-.05	-.051
	(.091)	(.095)	(.097)	(.081)	(.081)	(.083)
Pre-crisis TFP	.75	.67	.73	.55	.5	.52
	(.079)	(.096)	(.094)	(.067)	(.065)	(.065)
1[consumer good]			.13			-.0083
			(.1)			(.083)
1[intermediate good]			.14			.065
			(.098)			(.077)
Sector FE	No	Yes	Yes	No	Yes	Yes
Region FE	No	Yes	Yes	No	Yes	Yes
N. of cases	385	385	379	385	385	379

Table 4.A.6: Management style and firm productivity during the crisis - robustness.
Notes. This table shows results of regressions which use imputed data to account for possibly endogenous firm exit during the Great Recession. The dependent variable in columns 1-3 is i) a firm's observed productivity level using 2007 to 2010 data, or ii) the productivity level of the least productive firm in that period for those firm we do not observe in that period but do observe in the prior period. The dependent variable in columns 4-6 differs in that we use the firm at the fifth percentile to impute missing values (rather than the least productive). We estimate quantile regressions for the median in column 1-3 and bootstrap standard errors. The bootstrap procedure is replicated 1,000 and draws cluster-robust samples. Standard errors in column 4-6 are analytic and clustered at the three-digit industry level, too. "Mgt style 2" is a variable between 0 and 1 and indicates style 2 intensity. "Pre-crisis TFP" is a firm's productivity calculated from 2001-2006 data. "1[consumer good]" and "1[intermediate good]" are indicators for firms that are located in the respective location along the value chain. The omitted category is firms producing capital goods. All standard errors are reported in parentheses.

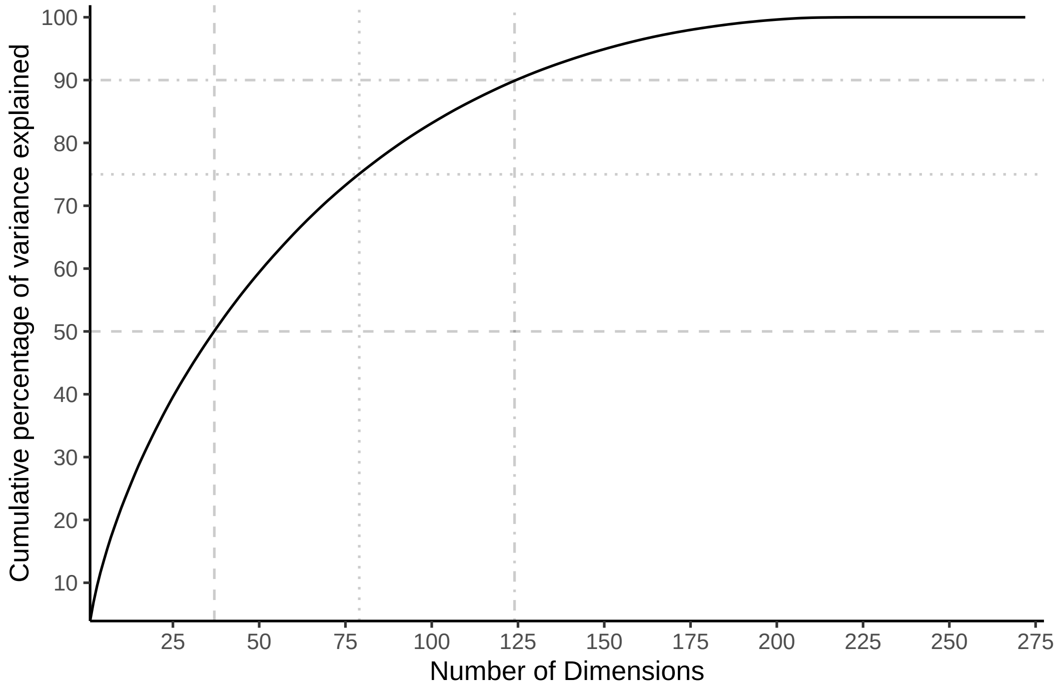


Figure 4.A.1: Cumulative % of variance explained. *Notes.* This figure shows the cumulative percentage of variation explained by the 272 management practice indicators. The results were obtained from running Multiple Correspondence Analysis (MCA) using all indicators. The x-axis contains all indicators ranked from the most to the least explanatory dimension.

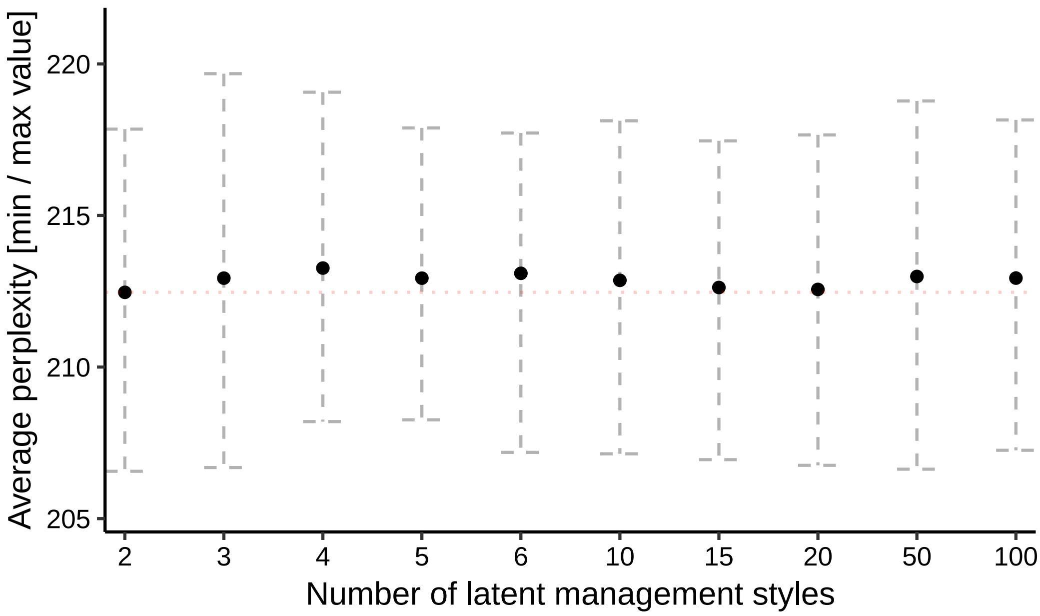
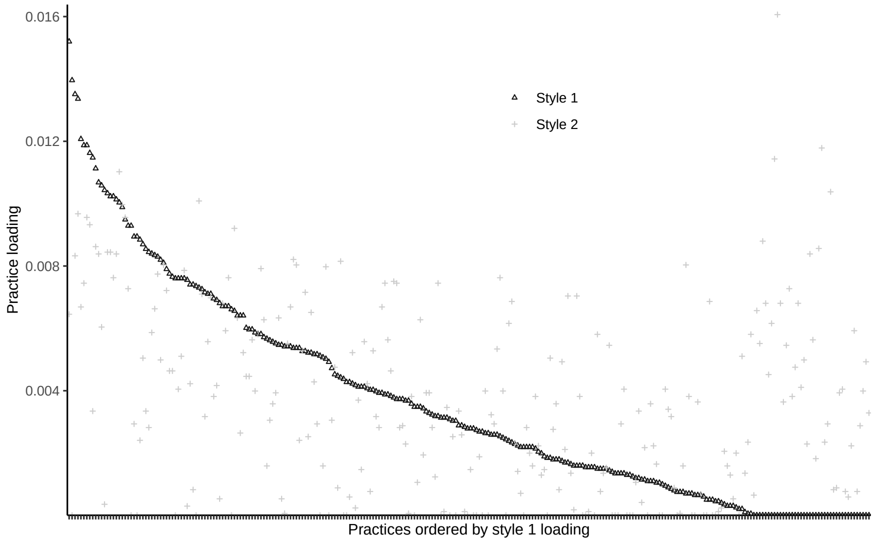
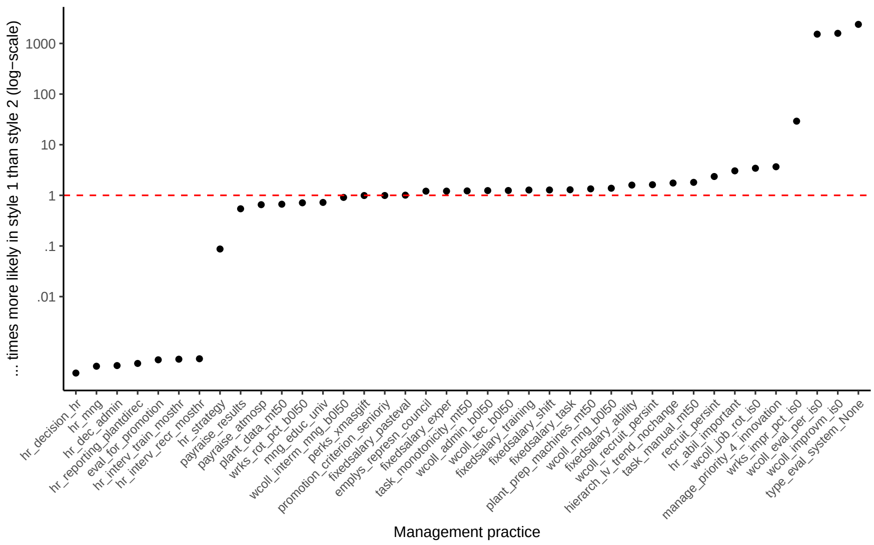


Figure 4.A.2: Cross-validating the number of latent styles. *Notes.* This figure plots the *perplexity* of cross-validated LDA models which vary only in the number of latent styles they estimate. Perplexity is a quantity that measures out-of-sample fit and higher values show better fit. The results are obtained from randomly splitting the sample into ten cross-validation folds. Then nine of those ten folds are used to estimate the model which is then tested on the held-out fold. This procedure is repeated ten times such that each fold is in the training sample exactly nine times, and in the test sample exactly one. The dots show the average perplexity across these ten repetitions for ach number of latent styles. The upper and lower end of the error bars show the maximum and minimum perplexity, respectively. The dashed red line shows the average perplexity obtained with the preferred model with two latent styles. The remaining parameters of the estimation are left unchanged and are described in Section 4.3.1.

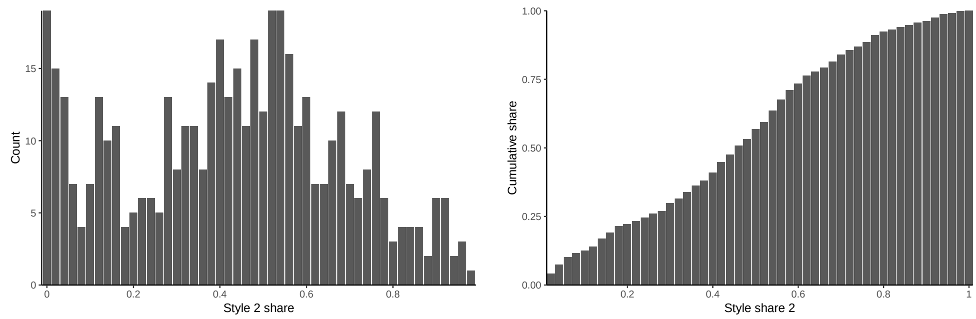


(a) Style distributions.



(b) Differences between styles.

Figure 4.A.3: Style-over-practice distributions. *Notes.* In this figure we visualize differences in practices' loadings across both latent style distributions. The distributions were estimates using the the single-plant sample alone. Each style is a distribution across 272 observed practices with each practice having a positive weight, and with the sum of weights summing to one. In panel (a),the practices are ordered such that the practice with the highest loading on style 1 is the far left of the x-axis. The y-axis shows the respective loadings of practices. In panel (b), we plot the quotient in loadings of the same practice across styles. A high value results from a case in which a practice's loading is higher in style 1 than in style 2, and vice versa. We plot the 20 highest ranks on either side breaking ties using the average.



(a) Probability density. (b) Cumulative density.

Figure 4.A.4: Firms' style 2 intensities. *Notes.* This figure plots the observed Style 2 intensities for all single-plant firms. These intensities were estimated using the single-plant sample alone. Panel (a) presents a histogram in which the unit interval was binned into 50 equidistant intervals. Panel (b) plots the cumulative density across those same 50 intervals.

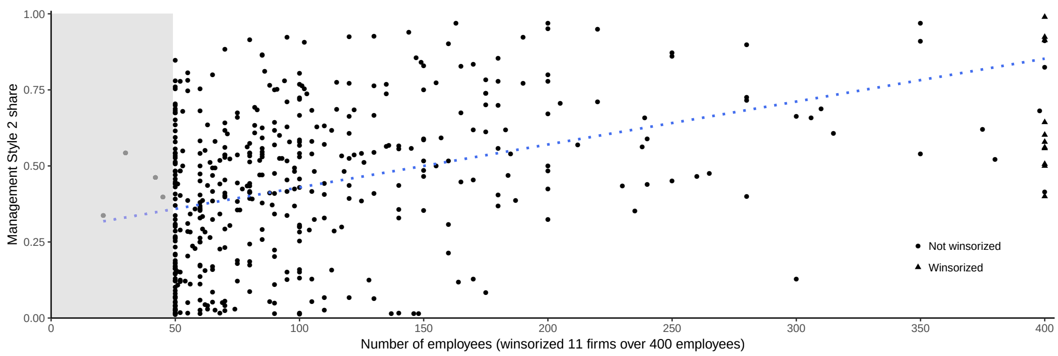


Figure 4.A.5: Style 2 intensity and firms' number of employees. *Notes.* This figure plots the simple univariate relationship between a firm's Style 2 intensity, and its self-reported number of employees from the survey. 11 firms with over 400 employees were winsorized for visual ease; they are represented with triangles rather than circles. The dotted blue line shows the line of linear best fit. Grey dots on the far left of the figure indicate firms that report less than 50 employees.

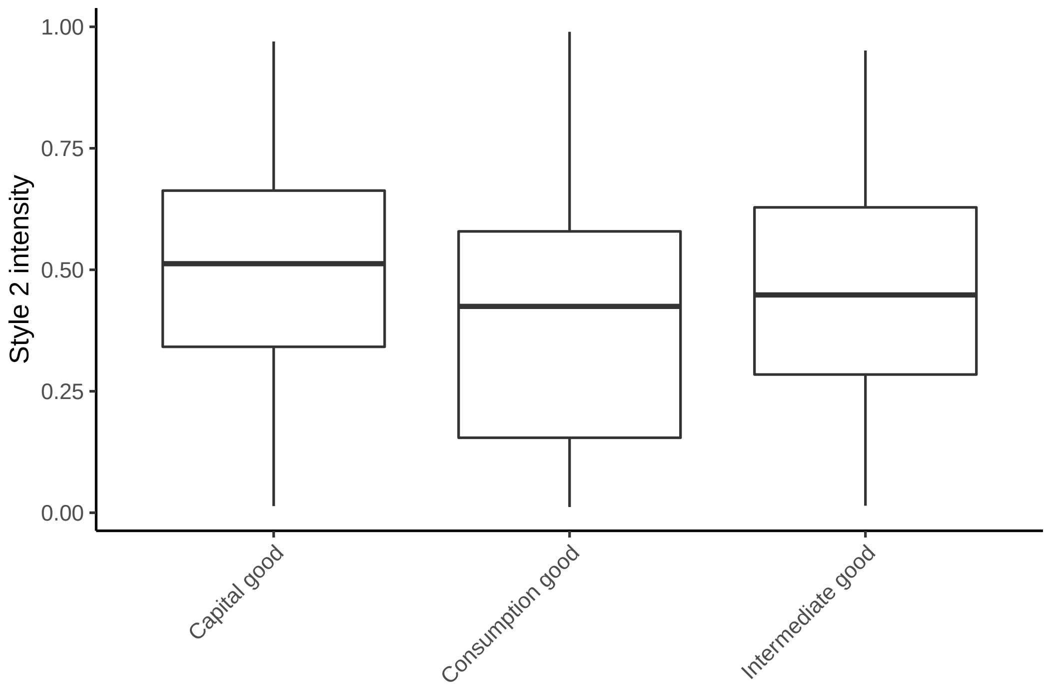
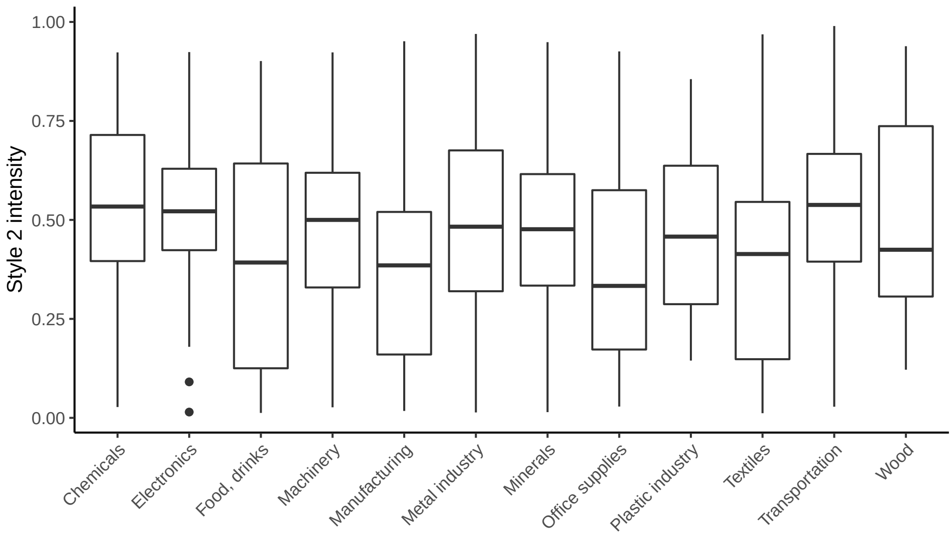
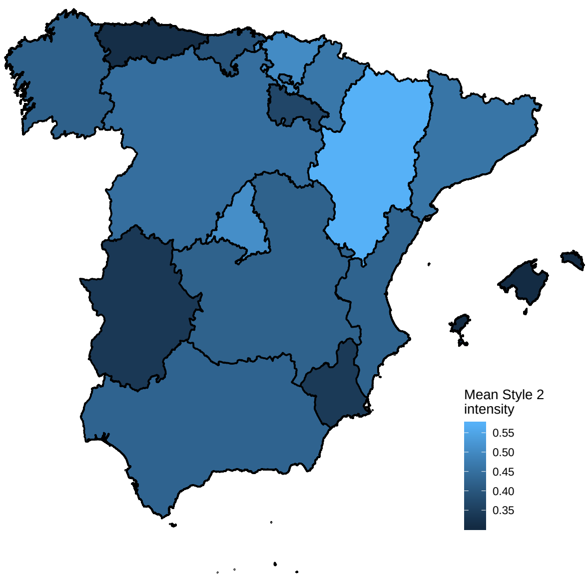


Figure 4.A.6: Style 2 intensity and firms' position in the value chain. *Notes.* This figure shows a box-and-whisker plot of Style 2 intensity relative to firms' position in the value chain. Firms indicate to be producing one of "consumption", "intermediate" or "capital" good in the survey. The horizontal bar within a *box* represents the median; the upper and lower hinge report the largest and small value within 1.5 times the interquartile range, respectively. Dotted values report values beyond the hinges but smaller than three times the interquartile range.



(a) Sectors.



(b) Regions.

Figure 4.A.7: Sectoral and regional variation in style 2 intensity. *Notes.* This figure plots the observed Style 2 intensities across sectors and regions. Panel (a) shows a box-and-whisker plot of Style 2 intensity relative to firms' sector of operation. Firms self-report in which sector they are active. The horizontal bar within a *box* represents the median; the upper and lower hinge report the largest and small value within 1.5 times the interquartile range, respectively. Dotted values report values beyond the hinges but smaller than three times the interquartile range. Panel (b) shows a map of Spanish region with color intensity reflecting firms' average Style 2 intensity.

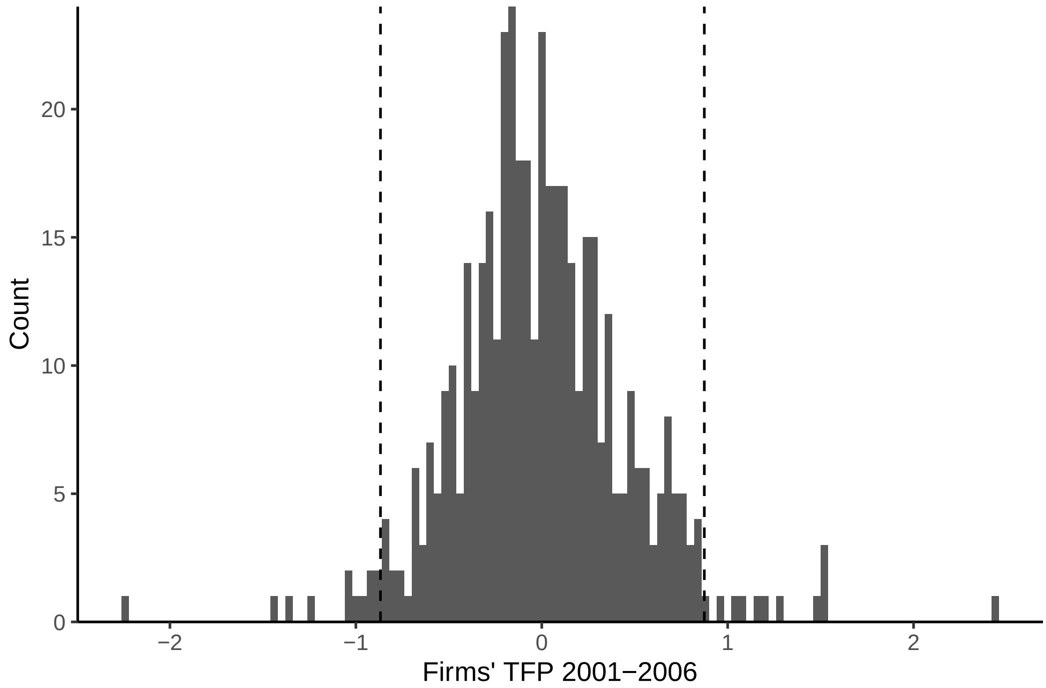


Figure 4.A.8: Firms's total factor productivity 2001-2006. *Notes.* This figure shows a histogram of firms' total factor productivity before the Great Recession using data from 2001-2006. We plot the predicted value of α obtained from estimating Equation (4.2). The histogram is constructed using a constant binwidth of 0.04. The vertical lines mark the 2.5^{th} and the 97.5^{th} percentile of the distribution. We use these values to winsorize the distribution in some specifications.

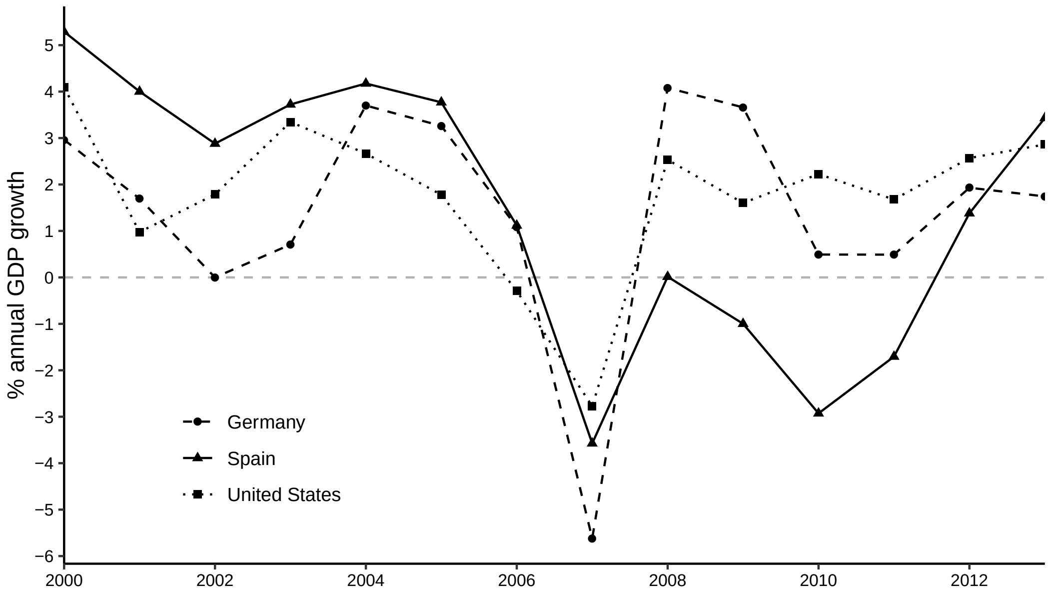
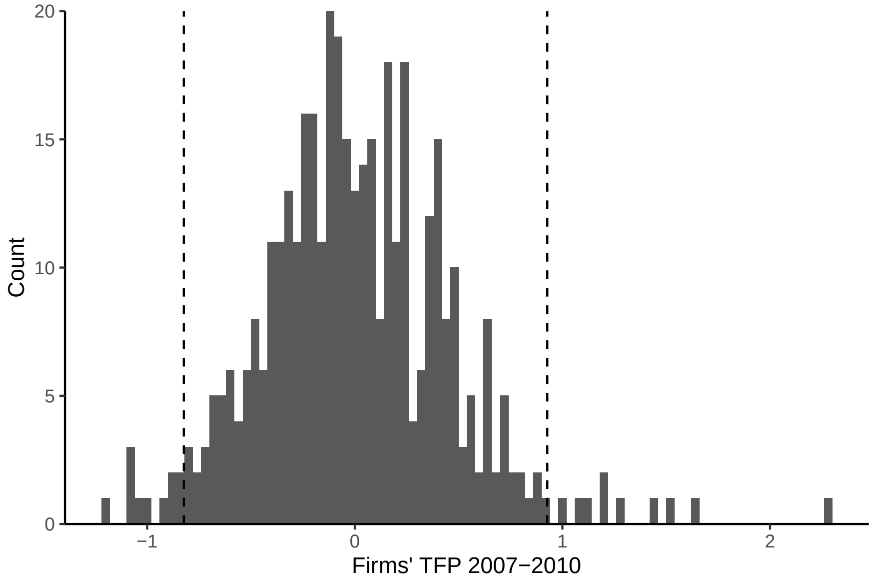
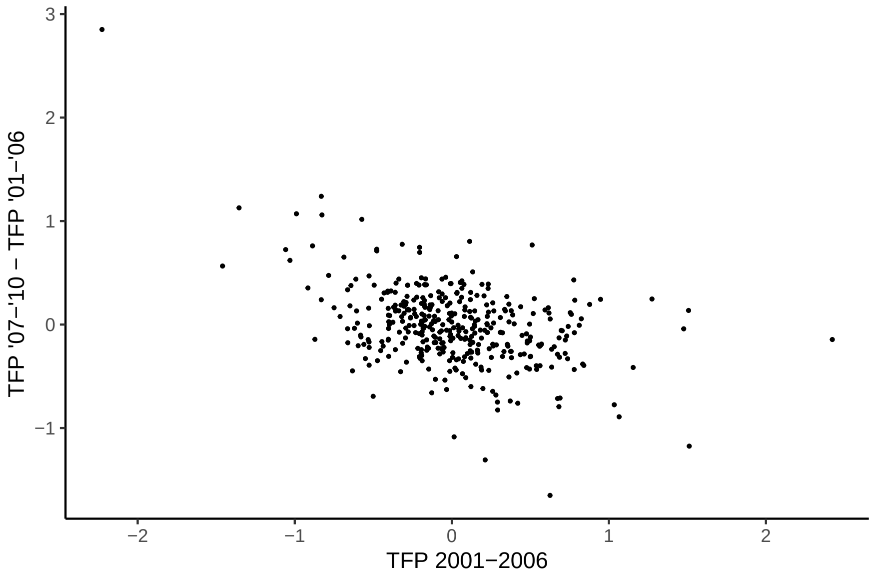


Figure 4.A.9: The Great Recession's impact on GDP growth for select countries.
Notes. This figure plot year-to-year GDP growth (in percent) for Germany, Spain and the United States. The data on which this figure is based is taken from the World Bank's World Development Indicators.



(a) TFP 2007-2010.



(b) Δ TFP relative to pre-crisis levels.

Figure 4.A.10: TFP during the Great Recession.. *Notes.* This figure presents firms' estimated TFP using data from 2007-2010. In panel (a) we show a simple histogram of firms' TFP using a binwidth of 0.04. Panel (b) is a scatter plot where we plot firms' TFP before the crisis (2001-2006 data) on the x-axis, and TFP during the crisis (panel (a) quantity) on the y-axis.

4.B. Overview of management practices.

Practice indicator	#	Question text	Answer	Mean
manage_priority_1_cost	A.10	How important are these factor to manage the plant?	First priority: Cost	0.22
manage_priority_1_flexibility	A.10	How important are these factor to manage the plant?	First priority: Flexibility	0.14
manage_priority_1_innovation	A.10	How important are these factor to manage the plant?	First priority: Innovation	0.13
manage_priority_1_quality	A.10	How important are these factor to manage the plant?	First priority: Quality	0.51
manage_priority_2_cost	A.10	How important are these factor to manage the plant?	Second Priority: Cost	0.30
manage_priority_2_flexibility	A.10	How important are these factor to manage the plant?	Second Priority: Flexibility	0.24
manage_priority_2_innovation	A.10	How important are these factor to manage the plant?	Second Priority: Innovation	0.17
manage_priority_2_quality	A.10	How important are these factor to manage the plant?	Second Priority: Quality	0.28

continued on next page ...

... continued from previous page

Practice indicator	#	Question text	Answer	Mean
manage_priority_3_cost	A.10	How important are these factor to manage the plant?	Third Priority: Cost	0.27
manage_priority_3_flexibility	A.10	How important are these factor to manage the plant?	Third Priority: Flexibility	0.33
manage_priority_3_innovation	A.10	How important are these factor to manage the plant?	Third Priority: Innovation	0.22
manage_priority_3_quality	A.10	How important are these factor to manage the plant?	Third Priority: Quality	0.17
manage_priority_4_cost	A.10	How important are these factor to manage the plant?	Forth Priority: Cost	0.21
manage_priority_4_flexibility	A.10	How important are these factor to manage the plant?	Forth Priority: Flexibility	0.29
manage_priority_4_innovation	A.10	How important are these factor to manage the plant?	Forth Priority: Innovation	0.47

continued on next page ...

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
manage_priority_4_quality	A.10	How important are these factor to manage the plant?	Forth Priority: Quality	0.03
num_certification_is1	A.18-20	Is plant certified with ISO 9000? + Some other certification? + . ISO 14000?	1 Certification?	0.38
num_certification_mt1	A.18-20	Is plant certified with ISO 9000? + Some other certification? + . ISO 14000?	More than 1 Certification?	0.33
num_certification_is0	A.18-20	Is plant certified with ISO 9000? + Some other certification? + . ISO 14000?	0 certifications?	0.29
recruit_personality	B.5	What of these tools are used in recruitment?	Personality	0.14
recruit_iq	B.5	What of these tools are used in recruitment?	IQ	0.07
recruit_genknowl	B.5	What of these tools are used in recruitment?	General Knowledge test	0.21
recruit_persint	B.5	What of these tools are used in recruitment?	Personal Interview	0.90
recruit_groupdyn	B.5	What of these tools are used in recruitment?	Group Dynamics	0.03

continued on next page . . .

... continued from previous page

Practice indicator	#	Question text	Answer	Mean
recruit_outsourced	B.5	What of these tools are used in recruitment?	Outsourced	0.03
hire_prim_age	B.6	Which of these factors does this plant take into account when hiring?	Primary: Age	0.05
hire_prim_education	B.6	Which of these factors does this plant take into account when hiring?	Primary: Education	0.16
hire_prim_experience	B.6	Which of these factors does this plant take into account when hiring?	Primary: Experience	0.54
hire_prim_personality	B.6	Which of these factors does this plant take into account when hiring?	Primary: Personality	0.05
hire_prim_qualification	B.6	Which of these factors does this plant take into account when hiring?	Primary: Qualification	0.12
hire_prim_teamwork	B.6	Which of these factors does this plant take into account when hiring?	Primary: Teamwork	0.06
hire_second_age	B.6	Which of these factors does this plant take into account when hiring?	Secondary: Age	0.12
hire_second_education	B.6	Which of these factors does this plant take into account when hiring?	Secondary: Education	0.24

continued on next page ...

... continued from previous page

Practice indicator	#	Question text	Answer	Mean
hire_second_experience	B.6	Which of these factors does this plant take into account when hiring?	Secondary: Experience	0.14
hire_second_personality	B.6	Which of these factors does this plant take into account when hiring?	Secondary: Personality	0.09
hire_second_qualification	B.6	Which of these factors does this plant take into account when hiring?	Secondary: Qualification	0.23
hire_second_teamwork	B.6	Which of these factors does this plant take into account when hiring?	Secondary: Teamwork	0.15
emplys_train_outside_amed	B.7	Percentage of workers got training outside of the plant and paid by the firm in 2005.	Percentage ¿ 50%	0.51
managers_fromwithin_all	B.9	How many supervisors and middle managers in the plant have previously been plain workers in the plant?	All	0.27
managers_fromwithin_bot_p20	B.9	How many supervisors and middle managers in the plant have previously been plain workers in the plant?	Bottom 20 %	0.12
managers_fromwithin_none	B.9	How many supervisors and middle managers in the plant have previously been plain workers in the plant?	None	0.03

continued on next page ...

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
managers_fromwithin_p21p40	B.9	How many supervisors and middle managers in the plant have previously been plain workers in the plant?	21 % - 40 %	0.11
managers_fromwithin_p41p60	B.9	How many supervisors and middle managers in the plant have previously been plain workers in the plant?	41 % - 60 %	0.09
managers_fromwithin_p61p80	B.9	How many supervisors and middle managers in the plant have previously been plain workers in the plant?	61 % - 80 %	0.17
managers_fromwithin_top_p20	B.9	How many supervisors and middle managers in the plant have previously been plain workers in the plant?	Top 20 %	0.19
vacant_spots_how_no_pref	B.10	How do you fill in vacant spots in the plant? 4 options.	No preference	0.07
vacant_spots_how_only_extern	B.10	How do you fill in vacant spots in the plant? 4 options.	Only external candidates	0.02
vacant_spots_how_only_internal	B.10	How do you fill in vacant spots in the plant? 4 options.	Only internal candidates	0.50

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
vacant_spots_how_pref_extern	B.10	How do you fill in vacant spots in the plant? 4 options.	Prefer external	0.02
vacant_spots_how_pref_internal	B.10	How do you fill in vacant spots in the plant? 4 options.	Prefer internal	0.37
promotion_criterion_equal	B.11	When promoting workers, rank seniority and merit.	Equally	0.19
promotion_criterion_merrit	B.11	When promoting workers, rank seniority and merit.	Merit	0.02
promotion_criterion_seniority	B.11	When promoting workers, rank seniority and merit.	Seniority	0.76
fin_discl_wrks_no	B.12	Do you publicly and periodically report financial status of the plant to workers?	No	0.33
fin_discl_wrks_reps	B.12	Do you publicly and periodically report financial status of the plant to workers?	Periodically?	0.39
fin_discl_wrks_yes	B.12	Do you publicly and periodically report financial status of the plant to workers?	Yes	0.28
emplys_represn_council	B.13	Are plant workers represented somehow?	Council	0.75
emplys_represn_delgates	B.13	Are plant workers represented somehow?	Delegates	0.12

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
emplys_represn_none	B.13	Are plant workers represented somehow?	No representation	0.11
emplys_represn_other	B.13	Are plant workers represented somehow?	Other form of representation	0.02
labor_agreement_collect_branch	B.14	Describe labor conditions in the plant? Type of labor agreement in place.	Sectoral agreement	0.52
labor_agreement_collect_firm	B.14	Describe labor conditions in the plant? Type of labor agreement in place.	Firm level agreement	0.38
labor_agreement_other	B.14	Describe labor conditions in the plant? Type of labor agreement in place.	Other	0.09
union_influence_high	B.15	Describe union influence on worker behavior.	High influence	0.29
union_influence_low	B.15	Describe union influence on worker behavior.	Low influence	0.33
union_influence_medium	B.15	Describe union influence on worker behavior.	Medium Influence	0.18
union_influence_veryhigh	B.15	Describe union influence on worker behavior.	Very high Influence	0.03

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
union_influence_verylow	B.15	Describe union influence on worker behavior.	Very low influence	0.12
lowprod_tol_below6	B.16	Tolerance towards worker of continuous low productivity.	Tolerance below 6	0.43
workers_incentivepay_mt0	C.1	Does any manufacturing worker receive variable pay/incentives?	More than 0	0.44
share_variablepay_11to20	C.2	Of those receiving variable pay, what percentage of their pay is variable?	11 - 20%	0.21
share_variablepay_1to10	C.2	Of those receiving variable pay, what percentage of their pay is variable?	1 - 10%	0.15
share_variablepay_21to30	C.2	Of those receiving variable pay, what percentage of their pay is variable?	21 - 30%	0.05
share_variablepay_30plus	C.2	Of those receiving variable pay, what percentage of their pay is variable?	31%+	0.06
share_variablepay_none	C.2	Of those receiving variable pay, what percentage of their pay is variable?	None	0.49

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
incentivepay_indivperf	C.3	What type of incentives are used, what percentage of workers receive these, and what percentage of their pay comes from this incentive?	Individual performance	0.32
incentivepay_firmperf	C.3	What type of incentives are used, what percentage of workers receive these, and what percentage of their pay comes from this incentive?	Firm performance	0.08
incentivepay_teamperf	C.3	What type of incentives are used, what percentage of workers receive these, and what percentage of their pay comes from this incentive?	Team performance	0.17
fixedsalary_task	C.4	What determines the fixed part of the workers compensation?	Type of task	0.80
fixedsalary_training	C.4	What determines the fixed part of the workers compensation?	Training	0.77
fixedsalary_tenure	C.4	What determines the fixed part of the workers compensation?	Tenure	0.61

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
fixedsalary_pasteval	C.4	What determines the fixed part of the workers compensation?	Past evaluations	0.65
fixedsalary_exper	C.4	What determines the fixed part of the workers compensation?	Experience	0.76
fixedsalary_ability	C.4	What determines the fixed part of the workers compensation?	Ability	0.79
fixedsalary_shift	C.4	What determines the fixed part of the workers compensation?	Shift	0.67
fixedsalary_personal	C.4	What determines the fixed part of the workers compensation?	Personal circumstances	0.45
payraise_inflation	C.6	What determines wage increases?	Inflation	0.57
payraise_recruit	C.6	What determines wage increases?	Recruiting and retention	0.44
payraise_results	C.6	What determines wage increases?	Firm results	0.49
payraise_atmosp	C.6	What determines wage increases?	Keeping good environment	0.54
payraise_compete	C.6	What determines wage increases?	Salaries of competing firms	0.38

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
payraise_law	C.6	What determines wage increases?	Law/labour agreements	0.61
payraise_hq	C.6	What determines wage increases?	Headquarter	0.25
workers_buyequity	C.7	Can workers buy equity on the firm?	Buy Equities of the firm or not	0.05
perks_discounts	C.9	Do you use these perks in your plant?	Discount for the final product	0.32
perks_family	C.9	Do you use these perks in your plant?	Family-based help	0.26
perks_xmasgift	C.9	Do you use these perks in your plant?	Christmas gift	0.80
perks_pension	C.9	Do you use these perks in your plant?	Pension	0.09
perks_lifeinsur	C.9	Do you use these perks in your plant?	Life insurance	0.26
perks_healthinsur	C.9	Do you use these perks in your plant?	Health insurance	0.11
type_eval_system_None	C.11	Does the firm use formal or informal evaluation systems? Both?	None	0.61

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
type_eval_system_both	C.11	Does the firm use formal or informal evaluation systems? Both?	Both	0.22
type_eval_system_objective	C.11	Does the firm use formal or informal evaluation systems? Both?	Objective / formal	0.15
type_eval_system_subjective	C.11	Does the firm use formal or informal evaluation systems? Both?	Subjective / informal	0.03
eval_frequency_semester_more	C.13	How often?	More than semester	0.17
eval_frequency_trimester	C.13	How often?	Trimester	0.23
wrk_eval_sup	C.14	Who evaluates the workers?	Supervisor?	0.18
wrk_eval_mng	C.14	Who evaluates the workers?	Manager	0.15
wrk_eval_hr	C.14	Who evaluates the workers?	HR	0.11
eval_for_salary	C.15	Evaluation results affect the workers salary increases, on-the-job training, promotion, firing?	Salary	0.25
eval_for_onjobtrain	C.15	Evaluation results affect the workers' salary increases, on-the-job training, promotion, firing?	On job training	0.20

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
eval_for_promotion	C.15	Evaluation results affect the workers' salary increases, on-the-job training, promotion, firing?	Promotion	0.32
eval_for_firing	C.15	Evaluation results affect the workers' salary increases, on-the-job training, promotion, firing?	Firing	0.24
hierarch_lv_trend_diminishing	D.1	What's the trend in the number of hierarchical levels in the plant?	Down	0.19
hierarch_lv_trend_increasing	D.1	What's the trend in the number of hierarchical levels in the plant?	Up	0.13
hierarch_lv_trend_nochange	D.1	What's the trend in the number of hierarchical levels in the plant?	Same	0.68
hierarchy_lev_12	D.2	How many hierarchical levels between supervisor and plant manager?	12 levels?	0.19
hierarchy_lev_3	D.2	How many hierarchical levels between supervisor and plant manager?	3 levels?	0.36
hierarchy_lev_4	D.2	How many hierarchical levels between supervisor and plant manager?	4 levels?	0.27

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
hierarchy_lev_5p	D.2	How many hierarchical levels between supervisor and plant manager?	5 levels?	0.18
wrksperspv_hl12_amed	D.3	What is the number of workers under one same supervisor?	12?	0.11
wrksperspv_hl3_amed	D.3	What is the number of workers under one same supervisor?	3?	0.18
wrksperspv_hl4_amed	D.3	What is the number of workers under one same supervisor?	4?	0.15
wrksperspv_hl5p_amed	D.3	What is the number of workers under one same supervisor?	5?	0.09
spv_coord_vimp	D.4a	What characterizes the job of a supervisor?	Coordination	0.65
spv_prod_vimp	D.4a	What characterizes the job of a supervisor?	Production	0.38
spv_deal_vimp	D.4a	What characterizes the job of a supervisor?	Problem solving	0.48
spv_spv_vimp	D.4a	What characterizes the job of a supervisor?	Supervision	0.47
spv_quality_vimp	D.4a	What characterizes the job of a supervisor?	Quality	0.47
spv_comm_act_vimp	D.4a	What characterizes the job of a supervisor?	Information flow	0.38

continued on next page . . .

... continued from previous page

Practice indicator	#	Question text	Answer	Mean
spv_comm_lev_vimp	D.4a	What characterizes the job of a supervisor?	Upstream communication	0.44
degr_spvision_high	D.5	How would you describe the degree of control/supervision of plant workers?	High amount of supervision	0.40
degr_spvision_low	D.5	How would you describe the degree of control/supervision of plant workers?	Low amount of supervision	0.06
degr_spvision_medium	D.5	How would you describe the degree of control/supervision of plant workers?	Medium amount of supervision	0.54
wrks_rot_pct_is0	D.6	Percentage of workers that rotate jobs, work in teams, contribute to improvement in processes?	Rotation: 0%	0.21
wrks_rot_pct_b0l50	D.6	Percentage of workers that rotate jobs, work in teams, contribute to improvement in processes?	Rotation: Between 0 and 50%	0.62
wrks_rot_pct_mt50	D.6	Percentage of workers that rotate jobs, work in teams, contribute to improvement in processes?	Rotation: More than 50%	0.17

continued on next page ...

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
wrks_team_pct_is0	D.6	Percentage of workers that rotate jobs, work in teams, contribute to improvement in processes?	Work in teams: 0	0.32
wrks_team_pct_b0l50	D.6	Percentage of workers that rotate jobs, work in teams, contribute to improvement in processes?	Work in teams: between 0 and 50%	0.42
wrks_team_pct_mt50	D.6	Percentage of workers that rotate jobs, work in teams, contribute to improvement in processes?	Work in teams: more than 50%	0.26
wrks_impr_pct_is0	D.6	Percentage of workers that rotate jobs, work in teams, contribute to improvement in processes?	Contribute to improvement in processes: 0	0.47
wrks_impr_pct_b0l50	D.6	Percentage of workers that rotate jobs, work in teams, contribute to improvement in processes?	Contribute to improvement in processes: between 0 and 50%	0.41

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
wrks_impr_pct_mt50	D.6	Percentage of workers that rotate jobs, work in teams, contribute to improvement in processes?	Contribute to improvement in processes: more than 50%	0.12
plant_prep_machines_is0	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Prepare machines they use: 0	0.14
plant_prep_machines_b0l50	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Prepare machines they use: between 0 and 50%	0.13
plant_prep_machines_mt50	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Prepare machines they use:	0.73

continued on next page . . .

... continued from previous page

Practice indicator	#	Question text	Answer	Mean
plant_maintenance_is0	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Do maintenance: 0	0.21
plant_maintenance_b0l50	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Do maintenance: between 0 and 50%	0.22
plant_maintenance_mt50	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Do maintenance: More than 50%	0.57
plant_data_is0	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Analyse Data: 0	0.22

continued on next page ...

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
plant_data_b0l50	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Analyse Data: between 0 and 50%	0.25
plant_data_mt50	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Analyse Data: More than 50%	0.53
plant_work_orga_is0	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Organize their workload autonomously: 0	0.24
plant_work_orga_b0l50	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Organize their workload autonomously: between 0 and 50%	0.30

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
plant_work_orga_mt50	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Organize their workload autonomously: More than 50%	0.47
plant_pace_is0	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Set their own pace: 0	0.20
plant_pace_b0l50	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Set their own pace: between 0 and 50%	0.22
plant_pace_mt50	D.7	To what extent plant workers prepare machines they use, do maintenance, analyze data, organize their workload autonomously, set their own pace?	Set their own pace: More than 50%	0.58

continued on next page . . .

... continued from previous page

Practice indicator	#	Question text	Answer	Mean
task_monotonicity_is0	D.8	Jobs of plant workers are monotone, complex, manual?	Monotone: 0	0.07
task_monotonicity_b0l50	D.8	Jobs of plant workers are monotone, complex, manual?	Monotone: between 0 and 50%	0.17
task_monotonicity_mt50	D.8	Jobs of plant workers are monotone, complex, manual?	Monotone: More than 50%	0.76
task_tec_complexity_is0	D.8	Jobs of plant workers are monotone, complex, manual?	Complex: 0	0.12
task_tec_complexity_b0l50	D.8	Jobs of plant workers are monotone, complex, manual?	Complex: between 0 and 50%	0.34
task_tec_complexity_mt50	D.8	Jobs of plant workers are monotone, complex, manual?	Complex: More than 50%	0.54
task_manual_is0	D.8	Jobs of plant workers are monotone, complex, manual?	Manual: 0	0.03

continued on next page ...

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
task_manual_b0l50	D.8	Jobs of plant workers are monotone, complex, manual?	Manual: between 0 and 50%	0.20
task_manual_mt50	D.8	Jobs of plant workers are monotone, complex, manual?	Manual: More than 50%	0.77
hr_absent_important	E.8	Rate the importance of HR goals?	Reduce absenteeism: Important	0.37
hr_absent_medium	E.8	Rate the importance of HR goals?	Reduce absenteeism: Medium importance	0.49
hr_absent_unimportant	E.8	Rate the importance of HR goals?	Reduce absenteeism: Unimportant	0.13
hr_moti_important	E.8	Rate the importance of HR goals?	Motivate employees: Important	0.47

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
hr_moti_medium	E.8	Rate the importance of HR goals?	Motivate employees: Medium importance	0.47
hr_moti_unimportant	E.8	Rate the importance of HR goals?	Motivate employees: Unimportant	0.06
hr_costs_important	E.8	Rate the importance of HR goals?	Reduce labor cost: Important	0.48
hr_costs_medium	E.8	Rate the importance of HR goals?	Reduce labor cost: Medium importance	0.48
hr_costs_unimportant	E.8	Rate the importance of HR goals?	Reduce labor cost: Unimportant	0.04
hr_climate_important	E.8	Rate the importance of HR goals?	Improve morale: Important	0.51

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
hr_climate_medium	E.8	Rate the importance of HR goals?	Improve morale: Medium importance	0.44
hr_climate_unimportant	E.8	Rate the importance of HR goals?	Improve morale: Unimportant	0.06
hr_retention_important	E.8	Rate the importance of HR goals?	Retention: Important	0.43
hr_retention_medium	E.8	Rate the importance of HR goals?	Retention: Medium importance	0.51
hr_retention_unimportant	E.8	Rate the importance of HR goals?	Retention: Unimportant	0.06
hr_recruit_important	E.8	Rate the importance of HR goals?	Recruitment: Important	0.42
hr_recruit_medium	E.8	Rate the importance of HR goals?	Recruitment: Medium importance	0.51

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
hr_recruit_unimportant	E.8	Rate the importance of HR goals?	Recruitment: Unimportant	0.07
hr_red_wrks_important	E.8	Rate the importance of HR goals?	Reduce number of workers: Important	0.26
hr_red_wrks_medium	E.8	Rate the importance of HR goals?	Reduce number of workers: Medium importance	0.18
hr_red_wrks_unimportant	E.8	Rate the importance of HR goals?	Reduce number of workers: Unimportant	0.56
hr_abil_important	E.8	Rate the importance of HR goals?	Improve training and ability: Important	0.50

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
hr_abil_medium	E.8	Rate the importance of HR goals?	Improve training and ability: Medium importance	0.42
hr_abil_unimportant	E.8	Rate the importance of HR goals?	Improve training and ability: Unimportant	0.08
hr_strategy	F.1	Is there a strategic plan in the plant detailing HR goals?	There is a strategic plan	0.33
hr_decision_admin	F.3	Where are HR decisions made?	Administration	0.15
hr_decision_genmgt	F.3	Where are HR decisions made?	General management	0.19
hr_decision_hr	F.3	Where are HR decisions made?	HR	0.59
hr_decision_other	F.3	Where are HR decisions made?	Other	0.01
hr_decision_prod	F.3	Where are HR decisions made?	Production?	0.06
hr_dec_admin	F.4	Does this department do other clerical tasks?	Yes?	0.42

continued on next page . . .

... continued from previous page

Practice indicator	#	Question text	Answer	Mean
hr_mng	F.5	HR department is part of managing team?	Yes?	0.43
hr_reporting_hrmgr	F.6	Who does the HR department report to?	HR manager	0.09
hr_reporting_othermgr	F.6	Who does the HR department report to?	Other manager	0.11
hr_reporting_plantdirec	F.6	Who does the HR department report to?	Plant director	0.38
hr_interv_recr_equal	F.7	Who intervenes in the following HR decisions?	Recruitment: Equal	0.20
hr_interv_recr_higherups	F.7	Who intervenes in the following HR decisions?	Recruitment: Highre-ups	0.08
hr_interv_recr_mosthr	F.7	Who intervenes in the following HR decisions?	Recruitment: Mostly HR	0.31
hr_interv_empl_equal	F.7	Who intervenes in the following HR decisions?	Retention: Equal	0.25
hr_interv_empl_higherups	F.7	Who intervenes in the following HR decisions?	Retention: Higher-ups	0.13
hr_interv_empl_mosthr	F.7	Who intervenes in the following HR decisions?	Retention: Mostly HR	0.20
hr_interv_prom_equal	F.7	Who intervenes in the following HR decisions?	Promotion: Equal	0.25

continued on next page ...

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
hr_interv_prom_higherups	F.7	Who intervenes in the following HR decisions?	Promotion: Higher-ups	0.15
hr_interv_prom_mosthr	F.7	Who intervenes in the following HR decisions?	Promotion: Mostly HR	0.18
hr_interv_eval_equal	F.7	Who intervenes in the following HR decisions?	Evaluation: Equal	0.27
hr_interv_eval_higherups	F.7	Who intervenes in the following HR decisions?	Evaluation: Higher-ups	0.14
hr_interv_eval_mosthr	F.7	Who intervenes in the following HR decisions?	Evaluation: Mostly HR	0.17
hr_interv_train_equal	F.7	Who intervenes in the following HR decisions?	Training: Equal	0.21
hr_interv_train_higherups	F.7	Who intervenes in the following HR decisions?	Training: Higher-ups	0.07
hr_interv_train_mosthr	F.7	Who intervenes in the following HR decisions?	Training: Mostly HR	0.32
wcoll_recruit_personality	G.1	What tools are used for recruitment and selection of white-collar employees?	Personality	0.22

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
wcoll_recruit_iq	G.1	What tools are used for recruitment and selection of white-collar employees?	IQ	0.16
wcoll_recruit_genknowl	G.1	What tools are used for recruitment and selection of white-collar employees?	General knowledge test	0.27
wcoll_recruit_persint	G.1	What tools are used for recruitment and selection of white-collar employees?	Personal Interview	0.89
wcoll_recruit_groupdyn	G.1	What tools are used for recruitment and selection of white-collar employees?	Group Dynamics	0.08
wcoll_recr_outsourced	G.1	What tools are used for recruitment and selection of white-collar employees?	Outsourced	0.07
wcoll_eval_per_is0	G.2	Percentage of white-collar workers that undergo an evaluation process?	0%	0.39
wcoll_eval_per_b0l50	G.2	Percentage of white-collar workers that undergo an evaluation process?	between 0 and 50%	0.26
wcoll_eval_per_mt50	G.2	Percentage of white-collar workers that undergo an evaluation process?	More than 50%	0.35
wcoll_train_is0	G.3	Percentage of white-collar workers that got training in 2005 paid by the firm.	0%	0.16

continued on next page . . .

... continued from previous page

Practice indicator	#	Question text	Answer	Mean
wcoll_train_b0l50	G.3	Percentage of white-collar workers that got training in 2005 paid by the firm.	Between 0 and 50%	0.47
wcoll_train_mt50	G.3	Percentage of white-collar workers that got training in 2005 paid by the firm.	More than 50%	0.37
wcoll_vac_no_pref	G.5	How are white-collar workers promoted? Criteria.	No preference	0.14
wcoll_vac_only_extern	G.5	How are white-collar workers promoted? Criteria.	Only external	0.12
wcoll_vac_only_internal	G.5	How are white-collar workers promoted? Criteria.	Only internal	0.32
wcoll_vac_pref_extern	G.5	How are white-collar workers promoted? Criteria.	Prefer external	0.07
wcoll_vac_pref_internal	G.5	How are white-collar workers promoted? Criteria.	Prefer internal	0.34
autoeval_efqm	A.21	Auto-evaluation of EFQM?	Yes?	0.15
wcoll_info_all	G.7	How often white-collar workers are informed of the financial status of the plant?	All information	0.48
wcoll_info_no	G.7	How often white-collar workers are informed of the financial status of the plant?	No information	0.23

continued on next page ...

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
wcoll_info_reps	G.7	How often white-collar workers are informed of the financial status of the plant?	Periodically?	0.29
wcoll_job_rot_is0	G.8	Percentage of white-collar workers that change jobs, work in teams, contribute to improvement of processes?	Change jobs: 0	0.62
wcoll_job_rot_b0l50	G.8	Percentage of white-collar workers that change jobs, work in teams, contribute to improvement of processes?	Change jobs: between 0 and 50%	0.33
wcoll_job_rot_mt50	G.8	Percentage of white-collar workers that change jobs, work in teams, contribute to improvement of processes?	Change jobs: more than 50%	0.05
wcoll_team_is0	G.8	Percentage of white-collar workers that change jobs, work in teams, contribute to improvement of processes?	Work in teams: 0	0.32
wcoll_team_b0l50	G.8	Percentage of white-collar workers that change jobs, work in teams, contribute to improvement of processes?	Work in teams: between 0 and 50%	0.31

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
wcoll_team_mt50	G.8	Percentage of white-collar workers that change jobs, work in teams, contribute to improvement of processes?	Work in teams: more than 50%	0.38
wcoll_improvm_is0	G.8	Percentage of white-collar workers that change jobs, work in teams, contribute to improvement of processes?	Contribute to improvement of processes: 0	0.40
wcoll_improvm_b0l50	G.8	Percentage of white-collar workers that change jobs, work in teams, contribute to improvement of processes?	Contribute to improvement of processes: between 0 and 50%	0.33
wcoll_improvm_mt50	G.8	Percentage of white-collar workers that change jobs, work in teams, contribute to improvement of processes?	Contribute to improvement of processes: more than 50%	0.27

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
wcoll_mng_is0	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Management: 0	0.03
wcoll_mng_b0l50	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Management: 1 - 50%	0.94
wcoll_mng_mt50	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Management: 51%+	0.03
wcoll_tec_is0	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Technicians: 0	0.02
wcoll_tec_b0l50	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Technicians: 1 - 50%	0.85
wcoll_tec_mt50	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Technicians: 51% +	0.13

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
wcoll_admin_is0	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Clerical: 0	0.02
wcoll_admin_b0l50	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Clerical: 1 - 50%	0.87
wcoll_admin_mt50	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Clerical: 51%+	0.11
wcoll_interm_mng_is0	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Intermediate management: 0	0.13
wcoll_interm_mng_b0l50	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Intermediate management: 1- 50%	0.84
wcoll_interm_mng_mt50	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Intermediate management: 51%+	0.03

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
wcoll_sale_is0	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Salesforce: 0	0.35
wcoll_sale_b0l50	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Salesforce: 1 - 50%	0.62
wcoll_sale_mt50	G.9	Percentage of white-collar workers that belong to management, technicians, clerical, intermediate management, salesforce.	Salesforce: 51%+	0.03
mng_age_young	H.1	Age	Young or not	0.23
mng_educ_belowSecond	H.2	Highest degree obtained.	Below secondary school	0.10
mng_educ_second	H.2	Highest degree obtained.	Secondary school	0.19
mng_educ_univ	H.2	Highest degree obtained.	University education	0.69
mng_tenure_b5	H.4	Years on the job.	Below 5 years	0.23
mng_tenure_5to15	H.4	Years on the job.	From 5 to 15 years	0.32

continued on next page . . .

. . . continued from previous page

Practice indicator	#	Question text	Answer	Mean
mng_tenure_mt15	H.4	Years on the job.	More than 15 years	0.38
mng_prev_sameplant	H.5	Where did he/she work before?	Same plant or not	0.44
mng_equ	H.7	Does he/she own equity?	Owns equity	0.58
mng_sex_female	H.9	Gender.	Male recorded as 1	0.08

Bibliography

ABADIE, A., S. ATHEY, G. W. IMBENS, AND J. M. WOOLDRIDGE (2020): "Sampling-based versus design-based uncertainty in regression analysis," *Econometrica*, 88(1), 265–296.

ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2001): "The colonial origins of comparative development: An empirical investigation," *American Economic Review*, 91(5), 1369–1401.

ADENA, M., AND A. HAGER (2020): "Does online fundraising increase charitable giving? A nation-wide field experiment on Facebook," Discussion paper, WZB Discussion Paper No. SP II 2020-302.

AGHION, P., N. BLOOM, B. LUCKING, R. SADUN, AND J. VANREENEN (2020): "Turbulence, Firm Decentralization and Growth in Bad Times," *American Economic Journal: Applied Economics*, forthcoming.

AKER, J., R. BOUMNIJEL, A. MCCLELLAND, AND N. TIERNEY (2016): "Payment mechanisms and anti-poverty programs: Evidence from a mobile money cash transfer experiment in Niger," *Economic Development and Cultural Change*, 65(1).

ALAN, S., T. BONEVA, AND S. ERTAC (2019): "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit," *Quarterly Journal of Economics*, 134(3), 1121–1162.

ALAREF, J., S. BRODMANN, AND P. PREMAND (2020): "The medium-term impact of entrepreneurship education on labor market outcomes: Experimental evidence from university graduates in Tunisia," *Labour Economics*, 62, 101787.

ALDRICH, J. (1993): "Reiersol, Geary and the idea of instrumental variables," *Economic and Social Review*, 24(3), 247–273.

ALMUNIA, M., P. ANTRAS, D. LOPEZ-RODRIGUEZ, AND E. MORALES (2018): "Venting Out: Exports during a Domestic Slump," Unpublished.

ALTMANN, S., A. FALK, P. HEIDHUES, R. JAYARAMAN, AND M. TEIRLINCK (2019): "Defaults and donations: Evidence from a field experiment," *Review of Economics and Statistics*, 101(5), 808–826.

Andersen, S., A. Di Girolamo, G. W. Harrison, and M. I. Lau (2014): "Risk and time preferences of entrepreneurs: evidence from a Danish field experiment," *Theory and Decision*, 77(3), 341–357.

Anderson, C. (2006): *The long tail: Why the future of business is selling less of more*. Hachette Books.

Andreoni, J., A. Payne, and S. Smith (2014): "Do grants to charities crowd out other income? Evidence from the UK," *Journal of Public Economics*, 114, 75–86.

Andreoni, J., and A. A. Payne (2003): "Do government grants to private charities crowd out giving or fund-raising?," *American Economic Review*, 93(3), 792–812.

——— (2011): "Is crowding out due entirely to fundraising? Evidence from a panel of charities," *Journal of Public Economics*, 95(5-6), 334–343.

Andreoni, J., and A. A. Payne (2013): "Charitable giving," in *Handbook of Public Economics*, vol. 5, pp. 1–50. Elsevier.

Angrist, J. D., and A. B. Krueger (1999): "Empirical strategies in labor economics," in *Handbook of Labor Economics*, vol. 3, pp. 1277–1366. Elsevier.

Angrist, J. D., and J.-S. Pischke (2008): *Mostly harmless econometrics: an empiricist's companion*. Princeton university press.

Ann, B., C. Ichniowski, and K. Shaw (2004): "Using" insider econometrics" to study productivity," *American Economic Review*, 94(2), 217–223.

Ashraf, N., O. Bandiera, E. Davenport, and S. S. Lee (2020): "Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services," *American Economic Review*, 110(5), 1355–94.

Åstebro, T., H. Herz, R. Nanda, and R. A. Weber (2014): "Seeking the roots of entrepreneurship: Insights from behavioral economics," *Journal of Economic Perspectives*, 28(3), 49–70.

Åstebro, T., S. A. Jeffrey, and G. K. Adomdza (2007): "Inventor perseverance after being told to quit: The role of cognitive biases," *Journal of Behavioral Decision Making*, 20(3), 253–272.

Athey, S. (2019): "The Impact of Machine Learning on Economics," in *The Economics of Artificial Intelligence: An Agenda*, ed. by A. Goldfarb, J. Gans, and A. Agrawal, pp. 507–547. University of Chicago Press.

Athey, S., and G. W. Imbens (2019): "Machine learning methods that economists should know about," *Annual Review of Economics*, 11, 685–725.

Autor, D. H., and M. J. Handel (2013): "Putting tasks to the test: Human capital, job tasks, and wages," *Journal of labor Economics*, 31(S1), S59–S96.

Backbaud Institute (2018): "Charitable giving report," Discussion paper, Backbaud Institute.

Bae, T. J., S. Qian, C. Miao, and J. O. Fiet (2014): "The relationship between entrepreneurship education and entrepreneurial intentions: A meta–analytic review," *Entrepreneurship Theory and Practice*, 38(2), 217–254.

Bandiera, O., A. Prat, S. Hansen, and R. Sadun (2020): "Ceo behavior and firm performance," *Journal of Political Economy*, 128(4), 1325–1369.

Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan (2015): "The miracle of microfinance? Evidence from a randomized evaluation," *American Economic Journal: Applied Economics*, 7(1), 22–53.

Barsbai, T., V. Licuanan, A. Steinmayr, E. Tiongson, and D. Yang (2020): "Information and the formation of social networks," *National Bureau of Economic Research. Working Paper #27346.*

Bauer, T. K., C. Rulff, and M. M. Tamminga (2019): "Berlin calling-Internal migration in Germany," Discussion paper, Ruhr Economic Papers.

Bayo-Moriones, A., J. E. Galdon-Sanchez, and S. Martinez-de Morentin (2013): "The diffusion of pay for performance across occupations," *ILR Review*, 66(5), 1115–1148.

Bayo-Moriones, A., J. E. Galdon-Sanchez, and S. Martinez-de Morentin (2017): "Performance Measurement and Incentive Intensity," *Journal of Labor Research*, 38(4), 496–546.

BBSR Bonn (2019): "Indikatoren und Karten zur Raum- und Stadtentwicklung (INKAR)," Available online `https://www.inkar.de/`.

Beaman, L., J. Magruder, and J. Robinson (2014): "Minding small change among small firms in Kenya," *Journal of Development Economics*, 108, 69–86.

Becerra, O., E. Cavallo, and I. Noy (2010): "In the aftermath of large natural disasters, what happens to foreign aid?," Working Papers 201018, University of Hawaii at Manoa, Department of Economics.

Becker, G. S., W. H. Hubbard, and K. M. Murphy (2010): "The market for college graduates and the worldwide boom in higher education of women," *American Economic Review*, 100(2), 229–33.

Behaghel, L., B. Crépon, M. Gurgand, and T. Le Barbanchon (2015): "Please call again: Correcting nonresponse bias in treatment effect models," *Review of Economics and Statistics*, 97(5), 1070–1080.

Bellmann, L., M. Bossler, H.-D. Gerner, and O. Hübler (2015): "IAB-Betriebspanel: Reichweite des Mindestlohns in deutschen Betrieben," Discussion paper, IAB-Kurzbericht.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*, 81(2), 608–650.

Bernard, T., and A. S. Taffesse (2014): "Aspirations: An approach to measurement with validation using Ethiopian data," *Journal of African Economies*, 23(2), 189–224.

Bertrand, M., and J. Pan (2013): "The trouble with boys: Social influences and the gender gap in disruptive behavior," *American Economic Journal: Applied Economics*, 5(1), 32–64.

Bjorvatn, K., and B. Tungodden (2010): "Teaching business in Tanzania: Evaluating participation and performance," *Journal of the European Economic Association*, 8(2-3), 561–570.

Blattman, C., N. Fiala, and S. Martinez (2014): "Generating skilled self-employment in developing countries: Experimental evidence from Uganda," *Quarterly Journal of Economics*, 129(2), 697–752.

Blei, D. M. (2012): "Probabilistic topic models," *Communications of the ACM*, 55(4), 77–84.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003): "Latent dirichlet allocation," *Journal of Machine Learning Research*, 3(Jan), 993–1022.

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, and J. van Reenen (2019): "What Drives Differences in Management Practices?," *American Economic Review*, 109(5), 1648–1683.

Bloom, N., R. Lemos, R. Sadun, D. Scur, and J. van Reenen (2014): "The new empirical economics of management," *Journal of the European Economic Association*, 12(4), 835–876.

Bloom, N., R. Sadun, and J. van Reenen (2012a): "Americans Do IT Better: US Multinationals and the Productivity Miracle," *American Economic Review*, 102(1), 167–201.

——— (2012b): "The organization of firms across countries," *Quarterly Journal of Economics*, 127(4), 1663–1705.

Bloom, N., and J. van Reenen (2007): "Measuring and explaining management practices across firms and countries," *Quarterly Journal of Economics*, 122(4), 1351–1408.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the distribution of male and female wages accounting for employment composition using bounds," *Econometrica*, 75(2), 323–363.

Bøg, M., H. Harmgart, S. Huck, and A. M. Jeffers (2012): "Fundraising on the Internet," *Kyklos*, 65(1), 18–30.

Bossler, M., and H.-D. Gerner (2019): "Employment Effects of the New German Minimum Wage: Evidence from Establishment-Level Microdata," *ILR Review*, pp. 1–25.

Bouchouicha, R., and F. M. Vieider (2019): "Growth, entrepreneurship, and risk-tolerance: A risk-income paradox," *Journal of Economic Growth*, 24(3), 257–282.

Brown, P. H., and J. H. Minty (2008): "Media coverage and charitable giving after the 2004 tsunami," *Southern Economic Journal*, pp. 9–25.

Brown, S., M. N. Harris, and K. Taylor (2012): "Modelling charitable donations to an unexpected natural disaster: Evidence from the US Panel Study of Income Dynamics," *Journal of Economic Behavior & Organization*, 84(1), 97–110.

Bruhn, M., D. Karlan, and A. Schoar (2018): "The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico," *Journal of Political Economy*, 126(2), 635–687.

Bruhn, M., and B. Zia (2013): "Stimulating managerial capital in emerging markets: The impact of business training for young entrepreneurs," *Journal of Development Effectiveness*, 5(2), 232–266.

Brynjolfsson, E., Y. Hu, and D. Simester (2011): "Goodbye Pareto principle, hello long tail: the effect of search costs on the concentration of product sales," *Management Science*, 57(8), 1373–1386.

Brynjolfsson, E., Y. J. Hu, and M. D. Smith (2006): "From niches to riches: Anatomy of the long tail," *Sloan Management Review*, 47(4), 67–71.

Brynjolfsson, E., and P. Milgrom (2013): *Handbook of Organizational Economics*chap. Complementarity in Organizations, pp. 11–55. Princeton University Press.

Burbidge, J. B., L. Magee, and A. L. Robb (1988): "Alternative transformations to handle extreme values of the dependent variable," *Journal of the American Statistical Association*, 83(401), 123–127.

Burmeister, K., and C. Schade (2007): "Are entrepreneurs' decisions more biased? An experimental investigation of the susceptibility to status quo bias," *Journal of Business Venturing*, 22(3), 340–362.

CALIENDO, M., A. FEDORETS, M. PREUSS, C. SCHRÖDER, AND L. WITTBRODT (2018): "The short-run employment effects of the German minimum wage reform," *Labour Economics*, 53, 46–62.

CALIENDO, M., C. SCHRÖDER, AND L. WITTBRODT (2019): "The causal effects of the minimum wage introduction in Germany–an overview," *German Economic Review*, 20(3), 257–292.

CAMERON, A. C., J. B. GELBACH, AND D. L. MILLER (2008): "Bootstrap-based improvements for inference with clustered errors," *Review of Economics and Statistics*, 90(3), 414–427.

CAMERON, A. C., AND D. L. MILLER (2015): "A practitioner's guide to cluster-robust inference," *Journal of human resources*, 50(2), 317–372.

CAMPOLIETI, M., T. FANG, AND M. GUNDERSON (2005): "How minimum wages affect schooling-employment outcomes in Canada, 1993–1999," *Journal of Labor Research*, 26(3), 533.

CAMPOS, F., M. FRESE, M. GOLDSTEIN, L. IACOVONE, H. C. JOHNSON, D. MCKENZIE, AND M. MENSMANN (2017): "Teaching personal initiative beats traditional training in boosting small business in West Africa," *Science*, 357(6357), 1287–1290.

CARD, D., AND A. B. KRUEGER (1994): "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 84(4), 772–784.

CARD, D., AND A. B. KRUEGER (1995): *Myth and Measurement: The New Economics of the Minimum Wage-Twentieth-Anniversary Edition*. Princeton University Press, New Jersey.

CARLSON, N., AND A. RINK (2019): "Starting fresh? Evidence from a field experiment with young entrepreneurs in Zimbabwe," *Working Paper*.

CASEY, K., R. GLENNERSTER, AND E. MIGUEL (2012): "Reshaping institutions: Evidence on aid impacts using a preanalysis plan," *Quarterly Journal of Economics*, 127(4), 1755–1812.

CHAPLIN, D. D., M. D. TURNER, AND A. D. PAPE (2003): "Minimum wages and school enrollment of teenagers: a look at the 1990's," *Economics of Education Review*, 22(1), 11–21.

CHEN, Y., X. LI, AND J. K. MACKIE-MASON (2005): "Online fund-raising mechanisms: a field experiment," *The BE Journal of Economic Analysis & Policy*, 5(2).

CLEMENS, J., L. KHAN, AND J. MEER (2018): "Dropouts need not apply: The minimum wage and skill upgrading," Discussion paper, National Bureau of Economic Research. Working Paper #27090.

CORONESE, M., F. LAMPERTI, K. KELLER, F. CHIAROMONTE, AND A. ROVENTINI (2019): "Evidence for sharp increase in the economic damages of extreme natural disasters," *Proceedings of the National Academy of Sciences*, 116(43), 21450–21455.

CURRIE, J., H. KLEVEN, AND E. ZWIERS (2020): "Technology and Big Data Are Changing Economics: Mining Text to Track Methods," *American Economic Review Papers and Proceedings*, 110, 42–48.

DAVID, A. C. (2011): "How do international financial flows to developing countries respond to natural disasters?," *Global Economy Journal*, 11(4).

DE MEL, S., D. MCKENZIE, AND C. WOODRUFF (2012): "Enterprise recovery following natural disasters," *Economic Journal*, 122(559), 64–91.

DEVELOPMENT INITIATIVES (2018): "Global humanitarian assistance report," Discussion paper, Development Initiatives, Global Humanitarian Assistance.

DEVLIN, R. A., AND D. ROWLANDS (2019): "Giving it away: charitable donations to overseas disasters," Discussion paper, University of Ottawa Department of Economics Working Paper #1905E.

DUCKWORTH, A. L., AND P. D. QUINN (2009): "Development and validation of the Short Grit Scale (GRIT–S)," *Journal of Personality Assessment*, 91(2), 166–174.

DUFLO, E., A. BANERJEE, A. FINKELSTEIN, L. F. KATZ, B. A. OLKEN, AND A. SAUTMANN (2020): "In praise of moderation: Suggestions for the scope and use of pre-analysis plans for RCTs in economics," *National Bureau of Economic Research. Working Paper #26993.*

DUSTMANN, C., A. LINDNER, U. SCHÖNBERG, M. UMKEHRER, AND P. VOM BERGE (2019): "Reallocation effects of the minimum wage: Evidence from Germany," Discussion paper, mimeo.

EISENSEE, T., AND D. STRÖMBERG (2007): "News droughts, news floods, and US disaster relief," *Quarterly Journal of Economics*, 122(2), 693–728.

EROSHEVA, E. A., S. E. FIENBERG, AND C. JOUTARD (2007): "Describing disability through individual-level mixture models for multivariate binary data," *Annals of Applied Statistics*, 1(2), 502–537.

FALK, A., A. BECKER, T. DOHMEN, B. ENKE, D. HUFFMAN, AND U. SUNDE (2018): "Global evidence on economic preferences," *Quarterly Journal of Economics*, 133(4), 1645–1692.

FEHR, E., AND L. GOETTE (2007): "Do workers work more if wages are high? Evidence from a randomized field experiment," *American Economic Review*, 97(1), 298–317.

FINK, G., AND S. REDAELLI (2011): "Determinants of international emergency aid: humanitarian need only?," *World Development*, 39(5), 741–757.

FISHER, R. A. (1935): *The Design of Experiments.* Oliver and Boyd. Edinburgh.

FITZENBERGER, B., AND A. DOERR (2016): "Konzeptionelle Lehren aus der ersten Evaluationsrunde der Branchenmindestlöhne in Deutschland," *Journal for Labour Market Research*, 49(4), 329–347.

FORTIN, N. M., P. OREOPOULOS, AND S. PHIPPS (2015): "Leaving boys behind gender disparities in high academic achievement," *Journal of Human Resources*, 50(3), 549–579.

FRESE, M., S. I. KRAUSS, N. KEITH, S. ESCHER, R. GRABARKIEWICZ, S. T. LUNENG, C. HEERS, J. UNGER, AND C. FRIEDRICH (2007): "Business owners' action planning and its relationship to business success in three African countries.," *Journal of Applied Psychology*, 92(6), 1481.

GANGULI, I., M. HUYSENTRUYT, AND C. LE COQ (2018): "How do nascent social entrepreneurs respond to rewards? A field experiment on motivations in a grant competition," *SITE Working Paper 46*.

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

GRIFFITHS, T. L., AND M. STEYVERS (2004): "Finding scientific topics," *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5228–5235.

GUHA-SAPIR, D. (2018): *EM-DAT: the Emergency Events Database*Brussels, Belgium, www.emdat.be.

GUZMAN, J., J. J. OH, AND A. SEN (2020): "What Motivates Innovative Entrepreneurs? Evidence from a Global Field Experiment," *Management Science*, 66(10), 4808–4819.

HAMILTON, B. H. (2000): "Does entrepreneurship pay? An empirical analysis of the returns to self-employment," *Journal of Political Economy*, 108(3), 604–631.

HELPER, S., AND R. HENDERSON (2014): "Management Practices, Relational Contracts and the Decline of General Motors," *Journal of Economic Perspectives*, 28(1), 49–72.

Herz, H., D. Schunk, and C. Zehnder (2014): "How do judgmental overconfidence and overoptimism shape innovative activity?," *Games and Economic Behavior*, 83, 1–23.

Hess, S. (2017): "Randomization inference with Stata: A guide and software," *The Stata Journal*, 17(3), 630–651.

Hicken, M. (2013): "Major charities raise millions for Typhoon Haiyan victims," `https://money.cnn.com/2013/11/15/pf/typhoon-haiyan-donations/?iid=EL`, Accessed: 2020-06-15.

Hsieh, C.-T., and B. A. Olken (2014): "The missing "missing middle"," *Journal of Economic Perspectives*, 28(3), 89–108.

Hurst, E., and B. W. Pugsley (2011): "What do small businesses do?," *National Bureau of Economic Research. Working Paper #17041.*

Hurst, E. G., and B. W. Pugsley (2015): "Wealth, tastes, and entrepreneurial choice," *National Bureau of Economic Research. Working Paper #21644.*

Hyslop, D., and S. Stillman (2007): "Youth minimum wage reform and the labour market in New Zealand," *Labour Economics*, 14(2), 201–230.

Ichniowski, C., K. Shaw, and G. Prennushi (1997): "The effects of human resource management practices on productivity: A study of steel finishing lines," *American Economic Review*, 87(3), 291–313.

Imai, K., L. Keele, D. Tingley, and T. Yamamoto (2011): "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies," *American Political Science Review*, pp. 765–789.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, pp. 467–475.

Jamison, D. T., B. Searle, K. Galda, and S. P. Heyneman (1981): "Improving elementary mathematics education in Nicaragua: An experimental study of the impact of textbooks and radio on achievement.," *Journal of Educational psychology*, 73(4), 556.

Just Giving (2012): "Survey of online fundraisers, sponsors, and donors: summary of responses," Discussion paper, Just Giving.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer (2015): "Prediction policy problems," *American Economic Review*, 105(5), 491–95.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007): "Experimental analysis of neighborhood effects," *Econometrica*, 75(1), 83–119.

Klinger, B., and M. Schündeln (2011): "Can entrepreneurial activity be taught? Quasi-experimental evidence from Central America," *World Development*, 39(9), 1592–1610.

Komro, K. A., M. D. Livingston, S. Markowitz, and A. C. Wagenaar (2016): "The effect of an increased minimum wage on infant mortality and birth weight," *American Journal of Public Health*, 106(8), 1514–1516.

Koudstaal, M., R. Sloof, and M. Van Praag (2016): "Risk, uncertainty, and entrepreneurship: Evidence from a lab-in-the-field experiment," *Management Science*, 62(10), 2897–2915.

Lee, D. S. (2009): "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *Review of Economic Studies*, 76(3), 1071–1102.

Levine, R., and Y. Rubinstein (2017): "Smart and illicit: Who becomes an entrepreneur and do they earn more?," *Quarterly Journal of Economics*, 132(2), 963–1018.

——— (2018): "Selection into entrepreneurship and self-employment," *National Bureau of Economic Research. Working Paper #5350.*

Lind, J. (1772): *A treatise on the scurvy.* S. Crowder.

List, J. A. (2011): "The market for charitable giving," *Journal of Economic Perspectives*, 25(2), 157–80.

List, J. A., A. M. Shaikh, and Y. Xu (2019): "Multiple hypothesis testing in experimental economics," *Experimental Economics*, 22(4), 773–793.

Manski, C. F., and J. V. Pepper (2018): "How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions," *Review of Economics and Statistics*, 100(2), 232–244.

McElheran, K., S. Ohlmacher, and M.-J. Yang (2020): "Strategy and Structured Management," Unpublished.

McKenzie, D. (2012): "Beyond baseline and follow-up: The case for more T in experiments," *Journal of Development Economics*, 99(2), 210–221.

——— (2017): "Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition," *American Economic Review*, 107(8), 2278–2307.

McKenzie, D., and C. Woodruff (2014): "What are we learning from business training and entrepreneurship evaluations around the developing world?," *World Bank Research Observer*, 29(1), 48–82.

——— (2016): "Business practices in small firms in developing countries," *Management Science*, 63(9), 2773–3145.

Milgrom, P., and J. Roberts (1990): "The Economics of Modern Manufacturing: Technology, Strategy and Organization," *American Economic Review*, 80(3), 511–528.

——— (1995): "Complementarities and fit strategy, structure, and organizational change in manufacturing," *Journal of Accounting and Economics*, 19(2-3), 179–208.

Mindestlohnkommission der Bundesregierung (2016): "Erster Bericht zu den Auswirkungen des gesetzlichen Mindeslohns," Discussion paper, Bundesregierung von Deutschland.

Mitchell, T. M. (1997): *Machine learning*. McGraw Hill.

Mohapatra, S., G. Joseph, and D. Ratha (2009): "Remittances and natural disasters: ex-post response and contribution to ex-ante preparedness," Discussion paper, The World Bank.

Moore, D. A., and P. J. Healy (2008): "The trouble with overconfidence.," *Psychological Review*, 115(2), 502.

Mullainathan, S., and J. Spiess (2017): "Machine learning: an applied econometric approach," *Journal of Economic Perspectives*, 31(2), 87–106.

Name-Correa, A. J., and H. Yildirim (2013): "A theory of charitable fundraising with costly solicitations," *American Economic Review*, 103(2), 1091–1107.

Neumark, D., and W. Wascher (1995): "Minimum-wage effects on school and work transitions of teenagers," *American Economic Review*, 85(2), 244–249.

——— (2003): "Minimum wages and skill acquisition: Another look at schooling effects," *Economics of Education Review*, 22(1), 1–10.

New York Times (2006): "Many Dismissing 'Donor Fatigue' as Myth: Despite Natural Disasters, Charities Say '05 Giving Held Steady," `https://www.nytimes.com/2006/04/30/us/many-dismissing-donor-fatigue-as-myth.html`, Accessed: 2020-09-11.

Obenauer, M. L., and B. von der Nienburg (1915): "Effect of minimum-wage determinations in Oregon," Discussion paper, Bulletin of the United States Bureau of Labor Statistics.

Pacheco, G., and A. Cruickshank (2007): "Minimum wage effects on educational enrollments in New Zealand," *Economics of Education Review*, 26(5), 574–587.

Payne, A., K. Scharf, S. Smith, et al. (2014): "Online fundraising-the perfect ask?," Discussion paper, No. 194, Competitive Advantage in the Global Economy (CAGE).

Penta, S. (2017): "Why Puerto Rico is getting the brunt of 'donor fatigue'," *The Conversation*, November 14.

Petrova, M., R. Perez-Truglia, A. Simonov, and P. Yildirim (2020): "Are political and charitable giving substitutes? Evidence from the United States," Discussion paper, National Bureau of Economic Research. Working Paper #26616.

Porta, R. L., and A. Shleifer (2008): "The unofficial economy and economic development," *National Bureau of Economic Research. Working Paper #14520.*

Premand, P., S. Brodmann, R. Almeida, R. Grun, and M. Barouni (2016): "Entrepreneurship education and entry into self-employment among university graduates," *World Development*, 77, 311–327.

Rammstedt, B., and O. P. John (2007): "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," *Journal of Research in Personality*, 41(1), 203–212.

Rigol, N., R. Hussam, and B. Roth (2018): "Targeting high ability entrepreneurs using community information: Mechanism design in the field," *Working Paper.*

Rose-Ackerman, S. (1982): "Charitable giving and "excessive" fundraising," *Quarterly Journal of Economics*, 97(2), 193–212.

Rosenbaum, P. R., and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41–55.

Scharf, K. A., S. Smith, and M. Wilhelm (2017): "Lift and shift: the effect of fundraising interventions in charity space and time," CESifo Working Paper Series #6694.

Schildberg-Hörisch, H. (2018): "Are risk preferences stable?," *Journal of Economic Perspectives*, 32(2), 135–54.

Schumpeter, J. (1934): "Das Grundphänomen der wirtschaflichen Entwicklung," in *Theorie der wirtschaftlichen Entwicklung*, pp. 88–140. Berlin: Duncker & Humblot.

Seaman, S. R., and I. R. White (2013): "Review of inverse probability weighting for dealing with missing data," *Statistical Methods in Medical Research*, 22(3), 278–295.

Shane, S. (2009): "Why encouraging more people to become entrepreneurs is bad public policy," *Small Business Economics*, 33(2), 141–149.

SHINNAR, R. S., D. K. HSU, AND B. C. POWELL (2014): “Self-efficacy, entrepreneurial intentions, and gender: Assessing the impact of entrepreneurship education longitudinally,” *International Journal of Management Education*, 12(3), 561–570.

SIR/DPA (2013): “In Deutschland wird weiter viel gespendet,” https://www.schwarzwaelder-bote.de/inhalt.taifun-haiyan-in-deutschland-wird-weiter-viel-gespendet.6c42645d-3a2d-456c-b569-58bc6db6927e.html, Accessed: 2020-06-15.

SMITH, S., M. OTTONI-WILHELM, AND K. SCHARF (2017): “The donation response to natural disasters,” Discussion paper, No. W17/19. IFS Working Papers.

SNOW, J. (1855): *On the mode of communication of cholera*. John Churchill.

SOLESVIK, M. Z. (2013): “Entrepreneurial motivations and intentions: Investigating the role of education major,” *Education+ Training*, 55(3), 253—-271.

STATISTISCHES BUNDESAMT (2014): “Verdienststrukturerhebung (VSE),” .

STEPHENS JR, M., AND D.-Y. YANG (2014): “Compulsory education and the benefits of schooling,” *American Economic Review*, 104(6), 1777–92.

STIRK, C. (2014): “Humanitarian assistance from Non-State donors,” Discussion paper, Development Initiatives, Global Humanitarian Assistance.

STOCK, J. H., AND F. TREBBI (2003): “Retrospectives: Who invented instrumental variable regression?,” *Journal of Economic Perspectives*, 17(3), 177–194.

STREICHER, M., L. ROSENDAHL HUBER, K. MOBERG, C. JØRGENSEN, AND D. REDFORD (2019): “Filling in the Blanks? The Impact of Entrepreneurship Education on European High School Students,” in *Academy of Management Proceedings*. Academy of Management Briarcliff Manor, NY 10510.

STRÖMBERG, D. (2007): “Natural disasters, economic development, and humanitarian aid,” *Journal of Economic perspectives*, 21(3), 199–222.

SUTCH, R. (2010): “The Unexpected Long-Run Impact of the Minimum Wage: An Educational Cascade,” Discussion paper, National Bureau of Economic Research. Working Paper #16355.

THE ECONOMIST (2013): “Generation joblesws,” https://www.economist.com/international/2013/04/27/generation-jobless, Accessed: 2020-09-03.

——— (2016): “Economists are prone to fads, and the latest is machine learning,” https://www.economist.com/finance-and-economics/2016/11/24/economists-are-prone-to-fads-and-the-latest-is-machine-learning, Accessed: 2020-09-07.

——— (2020a): "Africa is changing so rapidly, it is becoming hard to ignore," https://www.economist.com/special-report/2020/03/26/africa-is-changing-so-rapidly-it-is-becoming-hard-to-ignore, Accessed: 2020-09-03.

——— (2020b): "What harm do minimum wages do?," https://www.economist.com/schools-brief/2020/08/15/what-harm-do-minimum-wages-do, Accessed: 2020-08-21.

The Guardian (2015): "Life in the Philippines: preparing for the next typhoon Haiyan," https://www.theguardian.com/environment/2015/mar/25/life-in-the-philippines-preparing-for-the-next-typhoon-haiyan, Accessed: 2020-06-17.

The White House (2013): "FACT SHEET: U.S. Response to Typhoon Haiyan," https://obamawhitehouse.archives.gov/the-press-office/2013/11/19/fact-sheet-us-response-typhoon-haiyan, Accessed: 2020-06-15.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

Time Magazine (2014): "Tacloban Survived the World's Strongest Recorded Storm, but Only Just," https://time.com/3554112/tacloban-yolanda-supertyphoon-typhoon-haiyan-anniversary/, Accessed: 2020-09-11.

Ubfal, D., I. Arraiz, D. W. Beuermann, M. Frese, A. Maffioli, and D. Verch (2019): "The impact of soft-skills training for entrepreneurs in Jamaica," *IZA Discussion Paper #12325.*

United Nations (2018): "Global humanitarian overview," Discussion paper, United Nations Coordinated Support to People Affected by Disaster.

Wehby, G., W. Lyu, R. Kaestner, and D. Dhaval (2020): "Effects of the Minimum Wage on Child Health," Discussion paper, National Bureau of Economic Research. Working Paper #26691.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data.* MIT press.

Wright, P. G. (1928): *Tariff on animal and vegetable oils.* Macmillan Company, New York.

Yang, D. (2008): "Coping with disaster: the impact of hurricanes on international financial flows, 1970-2002," *The BE Journal of Economic Analysis & Policy*, 8(1).

Yang, M.-J., M. Christensen, N. Bloom, R. Sadun, and J. Rivkin (2020): "How Do CEOs Make Strategy?," .

Young, A. (2019): "Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results," *Quarterly Journal of Economics*, 134(2), 557–598.