Three Essays on the Effects of Political Crises on Economic Attitudes

Universität Hamburg

Fakultät für Wirtschafts- und Sozialwissenschaften

Kumulative Dissertation zur Erlangung der Würde eines Doktors der Wirtschafts- und Sozialwissenschaften „Dr. rer. pol.“

(gemäß der PromO vom 18. Januar 2017)

Vorgelegt im August 2020 von Dan Torge Dammann aus Buchholz i.d.N.

Februar 2021

Prüfungskommission:

Vorsitzender: Prof. Dr. Gerd Mühlheußer

Erstgutachter: Prof. Dr. Thomas Siedler

Zweitgutachterin: Prof. Dr. Miriam Beblo

Zweitbetreuer: Prof. Dr. Jan Marcus

Datum der Disputation:

22. Januar 2021

Abstract

The literature shows that economic attitudes are malleable and reactive to environmental influences. However, relatively little has been done to understand how political crises affect economic attitudes. This thesis approaches this question and how to measure such effects.

An unorthodox difference-in-differences approach is discussed, which is promising for analyzing medium-term effects of political crises on economic attitudes with panel data. This approach calculates two differences over the same time dimension, one within the survey year of the crisis and another across survey years. It is shown how to apply this approach to perform correct inference. Although it has been used by some authors, the approach is underused in the economic literature.

Two empirical applications are presented. The first is the so-called 2015 refugee crisis. German panel data show that individuals became on average more anti-immigrant as a consequence of the crisis. Moreover, it is tested which demographic groups were the most reactive. This exercise shows that no demographic group can be identified which became more pro-immigrant during the crisis. Hence, it is suggested that the crisis induced a general swing to the right.

The second application is the case of the 2014 Crimea crisis. From a German perspective, this crisis has been interpreted as a Russian aggression against the West. Although it could hardly affect any economic outcomes, it did increase the willingness to take risk as measured by Germany panel data. It is argued that the reason for this increase is adaptive behavior in the sense that humans adapt to situations which they cannot control or change.

Kurzfassung

Die Fachliteratur zeigt, dass ökonomische Einstellungen veränderbar sind und von Umweltfaktoren beeinflusst werden. Weniger klar ist, ob und wie sich politische Krisen auf ökonomische Einstellungen auswirken. Diese Arbeit behandelt diese Frage und diskutiert, wie man derartige Effekt messen kann.

Es wird eine unkonventionelle Differenz-in-Differenzen-Strategie diskutiert, welche mittelfristige Effekte von Krisen auf ökonomische Einstellungen in Paneldaten misst. Dieser Ansatz bildet zwei Differenzen über dieselbe Zeitdimension, eine innerhalb einer Umfragewelle und eine über unterschiedliche Umfragewellen hinweg. Es wird gezeigt, wie man diese Methode für korrekte Inferenz verwendet. Obwohl sie bereits verwendet wurde, scheinen ihre Potentiale in der Literatur noch nicht erschlossen.

Zwei empirische Anwendungen der Methode werden vorgestellt. Die erste zeigt mit Paneldaten aus Deutschland, dass die sogenannte Flüchtlingskrise von 2015 die Einstellungen zu Immigration im Durchschnitt negativ beeinflusst hat. Es wird getestet, welche demographischen Gruppen am stärksten reagierten. Es kann keine Gruppe identifiziert werden, die durch die Krise positiver über Immigration dachte. Dies legt nahe, dass es einen Rechtsruck in der Gesellschaft gegeben hat.

Die zweite Anwendung untersucht die Krimkrise von 2014. Aus deutscher Sicht wurde die Krise als russische Aggression gegenüber dem Westen interpretiert. Obwohl es kaum Auswirkungen auf ökonomische Größen gab, hat sich durch diese Krise die Risikobereitschaft in deutschen Paneldaten erhöht. Diese Erhöhung kann ein adaptives Verhalten darstellen, wobei Menschen sich Situationen anpassen, welche sie weder verhindern noch kontrollieren können.

Contents

List of Tables

List of Figures

Chapter 0

Introduction

When I was 10 years old, I went to piano classes every Tuesday. One Tuesday after class, my mother picked me up with her car and on our way home we listened to live commentary on the radio about a plane hitting a skyscraper in New York City. This was the day when I became interested in politics, it was September 11, 2001. This event clearly unsettled my beliefs about how the world works.

Crises kept coming. A few years later I struggled to understand the 2008 Lehman bankruptcy and the financial and economic crisis that followed. This phase of perceived economic instability culminated in the Greek government-debt crisis in 2010. This was the year I finished secondary schooling and I decided to study economics to gain a better understanding of the latest crises. During the next years, I followed coverage of the Arab Spring, the start of the Syrian war in 2011, the nuclear catastrophe of Fukushima, the annexation of the Crimea in 2014, and the so-called European refugee crisis in 2015.

Crises were the reason I started studying economics and they kept fascinating me during my studies. Did a piece of information that did not change any bit of my daily life actually change my perception of the world? Did a very distant event such as the nuclear catastrophe in Fukushima actually change not only my personal evaluation of the risks of nuclear technology, but also that of the German government? And how come that a terror attack that occurred 10,000 kilometers away occupied a 10-year-old boy to think about risk?

The answer is both trivial and very complicated. Humans constantly try to identify and understand patterns in the present and the past and build expectations about the future. Let an individual i's understanding of how the world works at time t be represented by a set of logical statements Ω_{it} the individual believes are true or not yet falsified. Note that neither does the individual need to be conscious of this set, nor does the set need to be built rationally. Moreover, this set may encompass contradictory statements which the individual cannot decide upon, or which the individual finds useful in different situations.[1] Based on this set, i tries to build rationales that guide her behavior. Now suppose that this set is in some way responsive to information. Let the set of information available to the individual at time t be denoted by F_{it} and the conscious or unconscious process which updates i's set of beliefs be defined as $p(\Omega_{it}, F_{it}) \rightarrow \Omega_{i,t+1}$. The process is a complicated interaction between Ω_{it} and F_{it}, for instance because i's current set of beliefs will determine which news she considers relevant. Changes in F_{it} might change the composition of

[1] Thus, Ω_{it} may encompass all the complexity and contradiction that leads to multiple selves which are relevant in different situations. See, for instance, the discussion of models with multiple selves in Frederick et al. (2002).

Ω_{it}, the weighting of its elements or the situational usage of them.

A shock to the information set will potentially challenge the rationales for behavior, i.e., the individual's attitudes and preferences. Throughout this thesis, I will define a political crisis as an event that drastically changes the set of public information in a negatively perceived way. I therefore analyze shocks to F_{it}. Analyzing such shocks allows us to get a better understanding of the process $p(\cdot)$, which is a black box in classical economic theory. The outcomes of interest are economic preferences and attitudes as a sub-set of Ω_{it}. This is particularly relevant to economists because the updated set of beliefs, $\Omega_{i,t+1}$, will guide subsequent behavior in economic domains. In this way, political crises can have long-lasting systematic effects on the economy.

There have been different ways of analyzing the determinants of attitudes and preferences. Most obviously, attitudes are determined by factors that do not or not suddenly vary. It has been stressed that time invariant factors such as parental background, height or gender and exogenous factors such as age determine attitudes (see, e.g., Dohmen et al., 2011; Croson and Gneezy, 2009). Studies on the historical origins of preference have used geographical variation in historical conditions to show that the latter shaped preferences over centuries (see, e.g., Alesina et al., 2011; Galor and Özak, 2016; Ang and Fredriksson, 2017; Falk et al., 2018; Ang, 2019; Bakker et al., 2020). The idea of historical conditions shaping preferences can be condensed into a theory on the evolution of preferences (Robson, 2001; Netzer, 2009). This theory, however, also assumes that evolution brought up individuals who could adapt to different situations (Netzer, 2009, p. 943).

Indeed, individuals' preferences can adapt over time within the span of a lifetime. Most studies showing this ability are long-term studies relating significant experiences to later life attitudes (see, for instance, Alesina and Fuchs-Schündeln, 2007; Giuliano and Spilimbergo, 2014; Siedler, 2011; Dohmen et al., 2011). Other studies have shown that information can even have situational effects on preferences (see, for instance, Kuziemko et al., 2015; Alesina and La Ferrara, 2005; Margalit, 2013; Luttmer, 2001).

This thesis contributes to the literature by discussing and highlighting a promising approach to test for medium-term effects of political crises on economic attitudes, and by applying this methodology to two case examples. It thereby delivers evidence both for the malleability of preferences and for the direct interaction between information and economic attitudes for two particular historical events.

The first chapter discusses the difficulty of identifying counterfactual attitudes in political crises. It argues that an unorthodox application of difference-in-differences can be applied to panel data by using both time variation within one survey year and time variation across survey years. Although it has already been used by some authors, this design is argued to be underused in the literature to understand historical events. To study its inferential properties and to discuss its specialties, Monte Carlo evidence of a data generating model is provided and illustrates how to use this design.

The second chapter applies this approach to understand how the so-called refugee crisis, which took place in Germany in 2015, affected the local population's attitudes towards immigration. It is shown by using different data sets that on average, attitudes became more anti-immigrant, that the willingness to vote for the determined right-wing party Alternative for Germany increased and that the population's approval of the government decreased. To allow for the possibility of sub-groups to behave differently, a new machine learning approach is applied to find heterogeneity in the treatment effect. Although the effects are heterogeneous, no sub-group is found that became more pro-immigrant, less likely to vote for the Alternative for Germany or more approving of the government. It is therefore concluded that Germany experienced a political swing to the right after the so-called refugee crisis.

The third chapter applies the approach to the case of the Crimea annexation from 2014 to understand how interstate conflict in neighboring regions affects the willingness to take risk. The hypothesis underlying this study is that individuals are able to adapt to different environments. The Crimea crisis was interpreted as an increase in the risk of war and general risk of living in Europe, and this risk could not be hedged or diversified by individuals. Hence, individuals may adapt to the higher environment risk by becoming more willing to take risk. This hypothesis cannot be rejected by the data since it is found that the Crimea crisis increased individuals' willingness to take risk significantly.

The motivation for this thesis is to understand whether my own experience that political crises changed my attitudes and preferences can be generalized to entire societies, with a special interest in Germany. I believe that I can convincingly support the evidence that crises affect attitudes in the short run with two new case examples. Moreover, I am convinced that this thesis contributes to the understanding of the specific two case examples picked in chapters 2 and 3. Hopefully, it can help to predict how future crises may affect individual attitudes and behavior, or at least create awareness of the fact that future crises are going to affect individual attitudes, and in this way behavior.

Bibliography

Alesina, A. and Fuchs-Schündeln, N. (2007). Goodbye Lenin (or Not?): The Effect of Communism on People's Preferences. *American Economic Review*, 97(4):1507–1528.

Alesina, A., Giuliano, P., and Nunn, N. (2011). Fertility and the Plough. *American Economic Review*, 101(3):499–503.

Alesina, A. and La Ferrara, E. (2005). Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5–6):897–931.

Ang, J. B. (2019). Agricultural legacy and individualistic culture. *Journal of Economic Growth*, 24(4):397–425.

Ang, J. B. and Fredriksson, P. G. (2017). Wheat agriculture and family ties. *European Economic Review*, 100:236–256.

Bakker, J. D., Maurer, S., Pischke, J.-S., and Rauch, F. (2020). Of Mice and Merchants: Connectedness and the Location of Economic Activity in the Iron Age. *The Review of Economics and Statistics*, pages 1–44. Publisher: MIT Press.

Croson, R. and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2):448–474.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692. Publisher: Oxford Academic.

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature*, 40(2):351–401.

Galor, O. and Özak, O. (2016). The Agricultural Origins of Time Preference. *American Economic Review*, 106(10):3064–3103.

Giuliano, P. and Spilimbergo, A. (2014). Growing up in a Recession. *The Review of Economic Studies*, 81(2):787–817. Publisher: Oxford Academic.

Kuziemko, I., Norton, M. I., Saez, E., and Stantcheva, S. (2015). How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments. *American Economic Review*, 105(4):1478–1508.

Luttmer, E. F. P. (2001). Group Loyalty and the Taste for Redistribution. *Journal of Political Economy*, 109(3):500–528. Publisher: The University of Chicago Press.

Margalit, Y. (2013). Explaining Social Policy Preferences: Evidence from the Great Recession. *American Political Science Review*, 107(1):80–103. Publisher: Cambridge University Press.

Netzer, N. (2009). Evolution of Time Preferences and Attitudes toward Risk. *American Economic Review*, 99(3):937–955.

Robson, A. J. (2001). The Biological Basis of Economic Behavior. *Journal of Economic Literature*, 39(1):11–33.

Siedler, T. (2011). Parental unemployment and young people's extreme right-wing party affinity: evidence from panel data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(3):737–758.

Chapter 1

Through Time and Time: An Unorthodox Difference in Differences Event Study Design

1.0 Abstract

National or worldwide news shocks and crises affect all individuals in the relevant population simultaneously. Hence, it is challenging to find an empirical counterfactual to outcome measures during such events without imposing very strong assumptions on the data. This paper discusses an unorthodox application of difference-in-differences in panel data as a solution which only requires weak assumptions. The design exploits the same time dimension within a survey year (treatment versus control group) and across survey years (before versus after treatment). Although being a promising approach that has been used by some authors, this is still a rare application in the economics literature. The design is superior to single difference analyses such as regression discontinuity designs when studying attitudinal changes with adjustment times. To study its inferential properties, Monte Carlo evidence of a data generating model is provided alongside discussions of potential problems and challenges that arise from the specialty of this design. It is shown that individual-fixed effects estimations with cluster-robust standard errors perform correct inference.[1] (JEL C23, D01, D83)

[1] All calculations and figures in this chapter are created with the software package R, version 3.5.3.

1.1 Introduction

When studying the effects of an unanticipated event such as a crisis or a news shock, social scientists often rely on secondary data. We cannot plan survey studies or even experiments to understand the effects of such sudden events. Instead, we have to rely on surveys that by chance take place at the same time as the events of interest.

A simple way to evaluate an event is to compare participants who were interviewed before an event with participants who were interviewed after an event in a single difference or regression discontinuity design. A problem with this approach is that survey timing may be correlated with participant characteristics for various reasons. It could happen by chance, because the interviewers plan to interview different regions or groups of individuals at particular times, because the form of data collection changes (Marcus, 2009), or because of seasonality. Studying the effect of an event may then, conditional on the timing of the event, result in a treated subsample with different average characteristics than the control group. This endangers single difference comparisons to be confounded.

Instead of single difference comparisons, it is possible to apply difference-in-differences (DiD) estimations to study news shocks in panel data. Suppose that the event occurs on day D in survey wave S. The control group are those individuals who are interviewed before D in wave S, and the treatment group are those interviewed on or after day D in wave S. If a panel survey is available, it will be possible to observe these two groups both in the treatment period S and in the pre-treatment periods S-t, where t>0.

This DiD approach is special because both of the two differences are calculated along the same time dimension (see, e.g., Angrist and Pischke, 2008, for a discussion of conventional DiD). It works by making a one-time cut through the sample during the treatment period. The first difference (treatment group versus control group) differentiates between different points in time within the treatment wave S, and the second difference (after treatment versus before treatment) differentiates between different points in time across the waves. Although this design has been used by some authors (see, for instance, Caliendo and Wrohlich, 2010; Dahlberg et al., 2020), the literature review in the next section shows that this is still an underused application of the DiD idea. This paper reflects on the general idea, provides a stylized model that allows to test the design in simulations and discusses potential pitfalls and how to avoid or test for them. To the best of my knowledge, this is the first structured discussion of this empirical strategy.

The empirical problem is formulated in a data generating model with two groups of individuals, where group membership is correlated with survey timing. It is shown that the design allows to control for potential confounders which are time constant, and for seasonality. The main identifying assumption is the standard DiD assumption, i.e., that the outcomes of the two groups would have developed parallel in absence of treatment. A simulation of the problem shows how one can deal with the serial correlation that this strategy incorporates by design (Bertrand et al., 2004). Most importantly, controlling for individual-fixed effects and using cluster-robust standard errors with clusters at the individual level performs correct inference. It is concluded that the presented DiD strategy is capable of dealing with the most critical empirical problems when analyzing societal dynamics that are caused by or interfere with sudden evens such as political crises or news shocks.

The paper is structured as follows. The next section provides a literature review. Section 3 provides a model of the core empirical problem, discusses how it can be used to simulate data, and why DiD is the preferred method in this case. Section 4 elaborates on potential problems and pitfalls with the help of the simulated data. Section 5 finally concludes on the presented insights. All tables and figures are appended in Section 6.

1.2 Literature Review

To assess how common the idea of a DiD on one time dimension is, I review the applications of DiD in the American Economic Review (AER). I interpret this journal as a representation of the current best practices in economics. The contributions published in this journal are often used as an orientation for applied work in economics, and if one searches for a good example of a DiD application, the AER will be a natural place to start.

To identify relevant articles I use the JSTOR archive's search engine[2] which allows me to search through all AER publications between 2010 and 2016. Excluding papers and proceedings, presidential addresses, comments and replies, there is a total of 801 AER publications during this time. Searching for all papers that explicitly mention the term „difference-in-differences“ yields 85 papers, of which I exclude 8 because they do not apply DiD although explicitly mentioning the method. I include papers that use DiD only as a robustness check or additional evidence and papers which

[2]The archive can be accessed on https://www.jstor.org, last accessed 11 December 2019.

apply fuzzy DiD (de Chaisemartin and D'HaultfŒuille, 2018). Papers which use estimation strategies that could in principle be represented by fuzzy DiD coefficients but do not explicitly apply a DiD strategy are not included (see, as an example, Gentzkow et al., 2011).[3]

Table 1.1 summarizes the review. The second column reports the total number of AER articles every year, the third column shows the frequency count for DiD papers, and the fourth column shows the share of DiD articles. Over all years, DiD papers account for a share of about 10 percent of all articles. Columns 5-9 report details about the DiD strategy, namely whether it was a region-time (R/T), group-time (G/T), group-group (G/G), region-group (R/G), or time-time (T/T) design. Region R refers to geographical entities like cities, counties, or countries. Time T refers to observation time. Group G refers to groups unrelated to time such as groups of students, product categories, technologies or business sectors. G also contains groups of individuals with the same birth year, i.e., cohorts, or more generally groups based on characteristics that depend on time in the past (Kostøl and Mogstad, 2014). It may be debatable whether this is a time dimension, but it is not exactly the same time dimension as the observation time. Different birth cohorts can be interviewed at the exact same point in time together during the treatment period. This is not possible for groups divided on observation times as proposed in this paper.[4]

None of the reviewed papers entails the problem of dividing the groups along the

[3]Note that this exercise probably misses many fuzzy DiD papers similar to the case of Gentzkow et al. (2011) because many of them do not seem to be aware that their estimates can be represented by Wald-DiDs as argued by de Chaisemartin and D'HaultfŒuille (2018). However, these accidental DiDs are not relevant to the review here, which searches for explicit and purposeful grouping designs.

[4]A detailed listing of the reviewed papers is provided in Table 1.2.

observation time while also using observation time as the second source of variation. The latter is a special case because it raises special pitfalls. Since, to the best of my knowledge, there are neither prominent applications of this design nor in-depth theoretical discussions, this paper provides a detailed discussion of the method.

1.3 A Simulation of the Problem

This section provides a framework for a simulation that will help to understand the empirical challenge of analyzing sudden events with secondary data. The first subsection presents a data generating model that incorporates various aspects of the empirical challenge and the second subsection discusses how to simulate data with this model. The third subsection reflects on different estimation strategies and argues why the DiD design on two time dimensions is reasonable for many applications. The fourth subsection specifies different empirical specifications that will be compared on simulated data.

1.3.1 A Simple Model of the Problem

As a minimal example of the problem suppose the following data generating population model. There are two groups g∈{A,B} in the population of interest. Most importantly, these groups can not be identified by observers. The outcome variable Y consists of a partly unobservable systematic part X with a group-dependent distribution[5], an average treatment effect θ and a random error ϵ with a group-dependent

[5] One may think of the systematic part as a linear combination of many characteristics Z_j, such that $X = \sum_{j=1}^{J} \beta_j Z_J$.

distribution. Let the outcome of individual i in group g in survey year s on the day d be defined as

$$Y_{igsd} = X_{igsd} + \theta 1_{\{s \geq S \text{ and } d \geq D\}} + \epsilon_{igsd}, \tag{1.1}$$

where $1_{\{\}}$ is the indicator function which equals one if the expression in brackets is true, and zero otherwise. Treatment occurs based on an event that starts at time D in survey year S such that all interviewees who participate on day $d \geq D$ in survey year $s \geq S$ are treated.

This data generating model has several important properties which one should be aware of. First, there is unobserved heterogeneity based on group membership. Group membership may be correlated with event timing which may confound estimations of the treatment effect. Furthermore, the group-dependent distribution of unobserved heterogeneity bears the possibility of clustering. Second, a correlation between the variance of unobservables and observable components of X_{igsd} or the treatment status introduces heteroskedasticity. Third, because the same individuals are observed over several survey waves s, there is the potential for strong serial correlation in the outcome (Bertrand et al., 2004). Moreover, the distribution of X_{igsd} may depend on s and d, which opens up a lot of possibilities for time trends and seasonality.

1.3.2 Simulation

To obtain a better understanding of the suggested estimation strategy and its potential problems, simulations of Eq. (1.1) are carried out. For the simulations of the problem I assume that time trends do not matter because they could easily be controlled for in a regression framework. I model the systematic part as $X_{igsd} \sim \mathcal{N}(\mu_g, \sigma_g^X)$ and the error as $\epsilon_{igsd} \sim \mathcal{N}(0, \sigma_g^\epsilon)$. Define $\Delta\mu \equiv \mu_B - \mu_A \geq 0$ as a measure for the expected group difference in outcomes.

The simulations of the data are performed for different levels of θ and $\Delta\mu$. For each combination of these parameters, the random simulation of the data is carried out 10,000 times, each time with 1,000 individuals.[6]

The days of interview are randomly assigned by a lottery with group-specific means. Suppose there are 90 possible days on which interviews can take place. For the depicted results the days were assigned by random draws from truncated normal distributions with mean 30 for group A and mean 60 for group B. A typical realization of this lottery is depicted in Figure 1.1a. Note that the choice of distribution for this lottery does not matter as long as it results in a correlation between the day of interview and the outcome variable, i.e., as long as the group-specific distributions assign major shares of the probability mass to different areas of the support.

As a representative example for outcome realizations, Figure 1.1b scatters the realized outcomes for a simulated sample of 1,000 individuals in the treatment year. The parameters of the distributions in this case are calibrated such that μ_B is 0.8 standard deviations higher than μ_A, where all standard deviations are assigned the same value $\sigma^X = \sigma^\epsilon$ irrespective of group membership.

The event of interest occurs randomly with a uniformly distributed timing. It does not make sense to choose a study design where the event occurs directly at the start or at the very end of the surveying procedure because this would result

[6] All simulations and graphical illustrations are carried out and created with the software R.

in extremely small control or treatment groups. I hence restrict the distribution of possible event times from day 11 to day 80 as depicted in Figure 1.2.

Repeated surveys that are publicly available like the German Socio-Economic Panel (SOEP, see Wagner et al., 2007), the British Household Panel Survey (BHPS) or the US-American Panel Study of Income Dynamics (PSID) can typically provide 10 or more years of consistent survey data around most more or less recent events of interest. Although the results are very consistent also for shorter and longer periods of survey data, I assume a survey panel of 10 years for depiction, where in year 10 the event of interest occurs. In the simulation I assume that a balanced panel is available, such that 10,000 observations are at hand for the 1,000 individuals in each of the 10,000 steps of the simulation.

1.3.3 Choice of Method

The design at hand is especially interesting to study exogenous news shocks and political crises. I thus discuss its application for the example of such an event.

The easiest approach to study exogenous news shocks or political crises is to compare survey participants interviewed shortly before an event to participants interviewed shortly after an event. Thus, a first idea may be to apply a single difference or regression discontinuity design. Especially a regression discontinuity approach would hinge on the assumption that attitudes and behaviors change immediately after the shock. However, the production and perception of news is a very complex process and in the real world, it may take some time until news actually change reported attitudes and behaviors.

News diffusion is relatively fast, but still takes some time. The quantity of information forwarded to us every day makes it impossible to be aware of every event immediately (Bawden and Robinson, 2009). Moreover, while news spread very fast within a network, it takes some time until news spread across all networks. This phenomenon is sometimes referred to as filter bubble problem (Pariser, 2011). Further, new information need to be processed by humans. For instance, Dijk (2013) divides the process of news comprehension into six steps. First, attention and perception for a news item is required. Usually, this first step of news comprehension is measurable by click rates or Google search counts. Second, the news item needs to be consumed, e.g., by reading it in form of an article. Quantitative measures on news consumption that are usually available lose track at this step. Third, the news needs to be decoded and interpreted. Fourth, a representation of the item needs to be built in episodic memory. Fifth, the formation, uses, and updating of situation models needs to take place. Then, as a sixth step, the uses and changes of general social knowledge and beliefs like attitudes may take place. It is only this last step which causes the alteration of behavior that economists are concerned with.

The idea of press (or, more generally, media) freedom is that free and unrestricted discourse is a society's most effective tool to approach and approximate truth (see Rauch, 1993, for a vital discussion). Being aware of this discursive nature of media and news requires readers to be skeptical about single voices and news items. Even if the news about a crisis spread fast and individuals are very aware of it, the skepticism that is inherent in a liberal news system will cause a time lag between the different steps outlined by Dijk (2013). A representation of the latest news may be built immediately in episodic memory, but skepticism may lead individuals to wait for a medium-term media consent on a topic before one adjusts attitudes and behavior accordingly.[7]

Adjustment times between the first media occurrence of a crisis and attitudinal or behavioral reactions to it reveal a trade-off for single difference or regression discontinuity designs. Since sharp discontinuities in attitudes and behaviors are not to be expected around event dates, one needs to increase the window size around an event date to be able to measure the treatment effect. However, the larger the window size, the lower is the reliability of single difference designs because of the problem outlined in the model above. This trade-off is exactly what I experienced when I first tried to analyze effects of the Crimea crisis from 2014 on individual behavior, see Dammann (2020). Using a single difference design, I detected statistically strong effects of the crisis on individual behavior when choosing a large window size, but detected no jump in the immediate neighborhood of the event. However, when choosing the large window size, the treatment and the control group became entirely unbalanced, and I wondered whether there was a more elegant way of dealing with this problem than controlling for as many variables as possible, imposing strong empirical restrictions by explicitly modeling the dynamics of reaction, or using abstract balancing methods.

A difference-in-differences design is able to correct for bias when comparing two groups that differ in level characteristics, which makes it a better choice for analyzing

[7] If this were not true, one should observe a lot of short-term changes in attitudes and behaviors, depending on the news an individual consumed most recently. This would contradict the often stated assumption and observation that attitudes are rather stable over time.

news shocks in panel data. The necessary identifying assumption is that the group which was interviewed in the weeks and months after an event would have followed the same trend as the group interviewed in the weeks and months before an event starting from the years before the event.

1.3.4 Estimation Strategies

Three estimation strategies are applied to each simulated data set: simple ordinary least squares (OLS), difference-in-differences (DiD) and a fixed effects version of the DiD model (FEM).

Define $Treat$ to equal one for individual i if she is treated in the treatment year, and zero otherwise. The variable $After$ equals one in the treatment year, and zero in all other periods. First, a naive OLS is carried out that compares outcomes after treatment with outcomes before treatment in the treatment year:

$$\text{OLS: Restrict data to s=S, then estimate } Y_i = \alpha_0 + \beta^{OLS} Treat_i + u_i \tag{1.2}$$

Second, a usual difference-in-differences estimation is performed that uses before and after comparisons during the treatment year combined with before and after comparisons over the survey years:

$$\text{DiD: } Y_{is} = \alpha_0 + \alpha_1 Treat_i + \alpha_2 After_s + \beta^{DID} Treat_i * After_s + u_{is} \tag{1.3}$$

Third, a fixed effects model (FEM) is estimated which is based on the DiD idea but controls for individual level effects:

$$\text{FEM: } Y_{is} = a_i + \alpha_2 After_s + \beta^{FEM} Treat_i * After_s + u_{is}. \tag{1.4}$$

In all these regressions, which are repeated 10,000 times for each combination of θ and $\Delta\mu$, the coefficient of interest is β. Heteroskedasticity-robust standard errors are estimated for OLS and DiD, and robust standard errors that account for clustering at the individual level are calculated for FEM.

In academic debates it is sometimes unclear what is meant by the distinction between DiD and FEM. The core idea of DiD is the explicit modeling of the counterfactual trend for the treatment group using the trend of the control group. This is most obviously done in the DiD specification, but the same idea translates into the FEM specification. Although FEM controls for higher-resolution unobserved heterogeneity, it still measures the treatment effect by projecting two general trends over time, one for the treatment and one for the control group. While it is always possible to fit the DiD idea in an FEM with panel data, not all fixed effects models project differentiated trends in the DiD sense. Hence, both DiD and FEM as presented above identify β using the core DiD idea, the only difference being the set of controls.

1.4 Discussion of Potential Problems and Pitfalls

When applying DiD in a non-standard sense as proposed here, one should be aware of obstacles that both may or may not usually be part of DiD. In this section I discuss potential problems and pitfalls that the proposed DiD design comes with.

The results from the simulation of the model are used as a common thread and illustration.

Event Timing A very basic problem when studying news shocks or political crises is to define the start of an event. The definition should satisfy two conditions. First, it should be driven by data. Both crises and news shocks are social events, i.e., events which are constituted by the fact that a critical mass of social individuals subjectively judge them to be these events. The best way to find out when this critical mass of individuals occurs is to analyze representative data which is informative about individuals' subjective judgments. Second, this data should be unrelated to the data used for the main analysis to avoid data snooping (White, 2000). External data sources such as Google Trends or news paper archives provide many opportunities for a careful examination of social dynamics regarding information and can be used to determine the start of a crisis or news shock, for instance by plotting the frequency of search terms related to an event. The latter is a proxy for the public awareness of and interest in the topic. Less precise definitions of the event timing will cause attenuation bias.

Identification Problem The problem for identification of θ is that group membership g may be correlated with both X_{igsd} and d. At the same time, X_{igsd} is often (partly) unobservable, of unknown composition or of unknown functional form. An estimation strategy that performs before and after comparisons is likely to pick up group differences even if one is optimistic to be able to control for some fraction of the systematic part.

Figure 1.3 shows box plots of the estimated treatment effects $\hat{\beta}$ from Eq. (1.2), (1.3) and (1.4) on 10,000 random data sets, each containing 10,000 observations of 1,000 individuals. Panel (a) shows the results for a data generating model that does not contain a difference between groups A and B. On average, OLS, DiD and FEM all yield the estimated treatment effect. All three estimations can be said to be unbiased in this case, although the OLS estimates are a little more dispersed. Note that this does not necessarily mean that the estimated standard errors are also larger for OLS than for the other methods. It only shows the distribution of estimated coefficients. As will be shown below, the standard errors for DiD and FEM differ, although their effect distributions are exactly the same.

Panels (b), (c), and (d) show the box plots of estimated effects for data that were generated with successively increasing unobserved group heterogeneity. While the DiD and FEM coefficients remain unbiased and stable as compared to panel (a) where no confounding factors where present, the single difference OLS strategy picks up the group difference in addition to the treatment effect. The estimated treatment effect rises, on average, by a little less than the group difference since both groups are present before and after the event.

In practice, there is no chance to check the validity of an estimation strategy like in Figure 1.3. But the DiD design allows for other simple and credible validity checks. The first step to investigate the standard DiD assumption of parallel trends is to plot the averages of the two groups' outcomes over time in all pre-treatment and treatment periods. As in the usual case, parallel trends in the pre-treatment periods suggest that the assumption is likely to hold. In addition, placebo tests that alter the timing of the treatment can be performed to check the validity. Placebo tests usually manipulate treatment timing. Because of the specialty of dividing the data twice along the survey time, both timing and grouping depend on time. Thus, not only placebo timing but also placebo grouping can be tested.

A cause for concern is that general time trends may affect both the group dimension and the time dimension since both are eventually the time dimension. As a consequence, one might worry that the estimated effect might not reflect a treatment effect but a general time trend. Since the long-run time trends from one year to the next are already captured by the trend comparison between treatment and control group, it is seasonality that needs to be additionally controlled for. The easiest way to achieve this is to include monthly or quarterly fixed effects. If a group deviates from the general long-run trend only because it is interviewed in a particular month or quarter, these seasonal fixed effects will capture the difference.

Note that seasonality will only be a problem if the interview timing differs across years for the same individuals. If every single individual is interviewed at the same time of the year in every survey wave, the DiD will control for seasonality by design since the treatment group's pre-treatment difference to the control group will contain the seasonality difference. Hence, another approach to check this problem would be to analyze whether interview timing eventually differs over survey waves.

Co-Treatment A further cause for concern is whether it is actually the event in question that causes the observed change. The longer the time period that is observed within the treatment period S, the greater is the chance that other events affect the outcome as well. A robustness check should scrutinize whether the results change drastically when only observations relatively close to the event are included in the sample. However, the idea of going arbitrarily close to the cutoff as, for instance, in a regression discontinuity design is unlikely to deliver promising results. As described above, the nature of the problem introduces a trade-off between statistical accuracy and unbiased estimation. To affect, for instance, economic attitudes or preferences, the new information need to work through psychological processes. Considering information where the human processing of new information is not finished is likely to deliver inconclusive results since individuals may be uncertain about the information content and its consequences for them. Thus, while going closer to the cutoff may be a good idea, going arbitrarily close to an event may be a bad idea.

Stable Group Compositions Another potential problem that might confound the estimation are changing group compositions in the treatment or the control groups over the survey waves. If, for instance, a major share of the treatment group is missing in pre-treatment years, the observation of a common trend during the pre-treatment periods will be uninformative about the actual trend parallelism, and the estimated treatment effect might simply reflect the fact that different people are interviewed before and after the treatment. Since a panel structure is necessary for the empirical design of interest, it is rather unlikely that the form of data collection changes drastically from one panel wave to another and that the composition of the samples changes for this reason (Marcus, 2009). However, panel attrition or the start of new sub-samples to increase the data pool might cause problematic changes in the sample structure (Wagner et al., 2007). Balancing restrictions can help to circumvent such problems and to ensure stable group compositions.

Attenuation Bias If attitudinal adjustment to a news shock takes its time, individuals whose response is measured briefly after the event will behave as if untreated. This would lead to an attenuation bias since the effect on the treatment group would be under-estimated. A strategy to deal with this attenuation bias is to exclude observations that were gathered briefly after the event of interest. This kind of doughnut strategy would at least remedy the bias.

Serial Correlation and Clustering Even when confounding factors are convincingly controlled for, clustering and serial correlation remain potential problems. Abadie et al. (2017) provide guidance for when using clustered standard errors is necessary. They suggest that standard errors should be clustered only if the sample is clustered by design or if the treatment assignment mechanism is clustered. There is no obvious clustering in the treatment assignment mechanism of a political crisis or news shock because it hits the entire population simultaneously. Thus, if clustering is necessary in this design, it is due to clustering in the sample. According to Abadie et al. (2017), this is especially true for two-stage sampling schemes where first regions are drawn out of the set of all regions, and then observation units are randomly drawn from the selected regions. It is straightforward to account for sample-induced clustering with standard software packages by adjusting standard errors at the regional level.

In the simulation, neither the treatment nor the sampling are clustered. However, serial correlation remains an additional problem for inference (Bertrand et al., 2004). Serial correlation is likely to occur because the treatment and control groups consist of the same individuals over time. To understand the inferential problems that may arise from serial correlation, Figure 1.4 plots the cumulative distributions of the p-values from testing $H_0 : \theta = 0$ versus the two-sided alternative after estimating the three empirical equations. Most importantly, the depicted distributions stem from a data generating model where $\theta = 0$, i.e., where the null hypothesis is true.

The p-value represents the probability that a test statistic is as extreme as it is or even more extreme under the null hypothesis. Hence, when estimating a correctly specified model over and over again in data where the null hypothesis is true, the fraction of p-values which is less or equal to a certain value (say, 5 percent) should be about that value (i.e., 5 percent). In completely random data, for instance in Eq. (1.1) when setting $X_{igsd} = \theta = 0$, the cumulative distribution function of p-values in this case is the 45 degree line between zero and one. This is the benchmark behavior of inference measures that one expects from a correct empirical specification.

Figure 1.4 gives interesting insights into the inferential behavior of the three estimation strategies under the null hypothesis. Panel (a) shows the behavior when the group difference $\Delta\mu$ is zero, and panels (b), (c) and (d) successively increase this group difference. While the simple difference OLS obviously performs good in panel (a) and bad as soon as there are confounding factors, inference with DiD and FEM is stable over all panels.[8] However, only FEM delivers the standard errors that lead to the preferred inferential behavior, while the simple DiD estimator noticeably deviates from the benchmark. The difference is that the FEM explicitly models serial correlation at the individual level. It is important to note that standard Software

[8]This result is robust to introducing group-specific error variances as long as robust standard errors are used. This suggests that clustering in the treatment assignment mechanism is not a problem in this application as suggested by the argument of Abadie et al. (2017). Standard errors may still need to be adjusted for clustering by sample design, but this is not part of this simulation.

like Stata clusters standard errors at the level of the panel variable, hence at the individual level in this exercise. Other software may require to program this adjustment more explicitly. Figure 1.5 provides the same graphs for cases where the null hypothesis is not true.

This simulation result provides three insights. First, explicitly modeling serial correlation by including individual-fixed effects is sufficient to obtain the expected inferential behavior in absence of a clustered sampling strategy. Cluster-robust standard errors that control for clustering at the individual level should be used to avoid problems in this domain. If the sample is clustered by design, an adjustment on the sample clusters should take place in addition. Second, a simple DiD is not necessarily more efficient than an FEM in presence of strong unobserved heterogeneity, as opposed to the often stated conventional wisdom (Lechner et al., 2016). Third, even if DiD was more efficient, that would not imply that it provides the inferential behavior that a researcher would usually expect.

Local Average Treatment Effects and Further Issues One advantage of DiD is that the average treatment effect can be identified. However, this is not necessarily the case in the special application proposed here. The sub-samples that are treated and for which the effect is estimated do not necessarily represent the average individual in the survey or the population of interest. The effect measured by the DiD estimator proposed here is thus local in the sense that it is measured only for the potentially special sub-group that experienced treatment. Standard heterogeneity checks or more advanced analyses of heterogeneity should be carried out to understand the importance of this potential problem in a particular application (see, for instance, Chernozhukov et al., 2018; Mullainathan and Spiess, 2017).

A convenient property of this estimation strategy is that the Stable Unit Treatment Value Assumption (SUTVA, see Rubin, 1980; Cox, 1958), which is an implicit assumption in causal modeling, holds almost tautologically. Because the assignment is based on the timing of an event, anticipation effects would be necessary for the SUTVA to be violated. However, it is a core characteristic of the shocks discussed in this paper that they are unanticipated, which makes it very unlikely for pre-treatment interviewees to respond to it. The fact that the events of interest are unanticipated crises also circumvents the so-called Ashenfelter's dip (Ashenfelter, 1978) where behavior of treated individuals changes in anticipation of treatment, or where participation in treatment depends on pre-treatment developments. Because anticipation effects are unlikely to occur and all individuals are treated at the same point in time, correlation between treatment and prior outcomes can not invalidate DiD in the design proposed here and will only result in level differences between treated and untreated individuals that the DiD can control for.

Some events occur very early or very late during a survey. Applying the DiD event study in these cases may lead to small treatment or control groups. In very extreme cases, an estimation may not make sense for the obvious reason that an effect would not be representative for any population of interest. In cases where the numbers of observation suggest that an estimation may make sense but treatment and control groups are strongly imbalanced, an over-sampling technique may be helpful (Chawla et al., 2002).

1.5 Concluding Remarks

This paper discusses a study design that is still very rare in the literature. The design exploits variation across individuals over time within a treatment year, and variation across years to identify the causal effect of an event. In other words, it is a DiD design that relies on the same time dimension for both differences. It is argued that this design is very promising for analyzing the impact of any unanticipated event in secondary data.

The design's advantage over, for instance, a regression discontinuity design is that it allows for adjustment times in human behavior after a news shock while controlling for biases that occur because of group specialties. Its disadvantages are that it suffers from the threat of co-treatments and the problem of serial correlation. Monte Carlo evidence suggests that an individual-fixed effects model with clustered standard errors at the individual level is the preferred estimation method for inference. In addition, standard errors may need to be adjusted for clustering in case the sampling strategy is clustered.

News shape our minds on politics, economics and the society in general. Understanding how particular news phenomena and information shocks impact on economic or social attitudes and behaviors is thus an important task for any discipline that seeks to understand the society. The empirical design presented in this paper provides an easy and convenient opportunity to identify the effects of such phenomena in secondary data. Future research should be aware of and exploit this opportunity.

Bibliography

Abadie, A., Athey, S., Imbens, G., and Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering? *arXiv:1710.02926 [econ, math, stat]*. arXiv: 1710.02926.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earnings. *The Review of Economics and Statistics*, 60(1):47–57.

Bawden, D. and Robinson, L. (2009). The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2):180–191.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, 119(1):249–275.

Caliendo, M. and Wrohlich, K. (2010). Evaluating the German 'Mini-Job' reform using a natural experiment. *Applied Economics*, 42(19):2475–2489. Publisher: Routledge _eprint: https://doi.org/10.1080/00036840701858125.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2018). Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments. Working Paper 24678, National Bureau of Economic Research.

Cox, D. R. (1958). *Planning of experiments*. Planning of experiments. Wiley, Oxford, England.

Dahlberg, M., Edin, P.-A., Grönqvist, E., Lyhagen, J., Östh, J., Siretskiy, A., and Toger, M. (2020). Effects of the COVID-19 Pandemic on Population Mobility under Mild Policies: Causal Evidence from Sweden. *arXiv:2004.09087 [econ, q-bio, q-fin]*. arXiv: 2004.09087.

Dammann, D. (2020). Adaptive Risk-Taking Behavior and the Crimea Crisis. *Working Paper - Chapter 3 of this Dissertation*.

de Chaisemartin, C. and D'HaultfŒuille, X. (2018). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, 85(2):999–1028.

Dijk, T. A. v. (2013). *News As Discourse*. Routledge. Google-Books-ID: Zeb-WAQAAQBAJ.

Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2011). The Effect of Newspaper Entry and Exit on Electoral Politics. *American Economic Review*, 101(7):2980–3018.

Kostøl, A. R. and Mogstad, M. (2014). How Financial Incentives Induce Disability Insurance Recipients to Return to Work. *The American Economic Review*, 104(2):624–655.

Lechner, M., Rodriguez-Planas, N., and Kranz, D. F. (2016). Difference-in-difference estimation by FE and OLS when there is panel non-response. *Journal of Applied Statistics*, 43(11):2044–2052.

Marcus, J. (2009). Der Einfluss von Erhebungsformen auf den Postmaterialismus-Index. *Methoden, Daten, Analysen (mda)*, 3(2):137–166.

Mullainathan, S. and Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2):87–106.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you.* Penguin UK.

Rauch, J. (1993). *Kindly inquisitors: The new attacks on free thought.* University of Chicago Press.

Rubin, D. B. (1980). Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association*, 75(371):591–593.

Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch - Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127(1):139–169.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica*, 68(5):1097–1126.

1.6 Appendix

Table 1.1: Counts of AER Articles

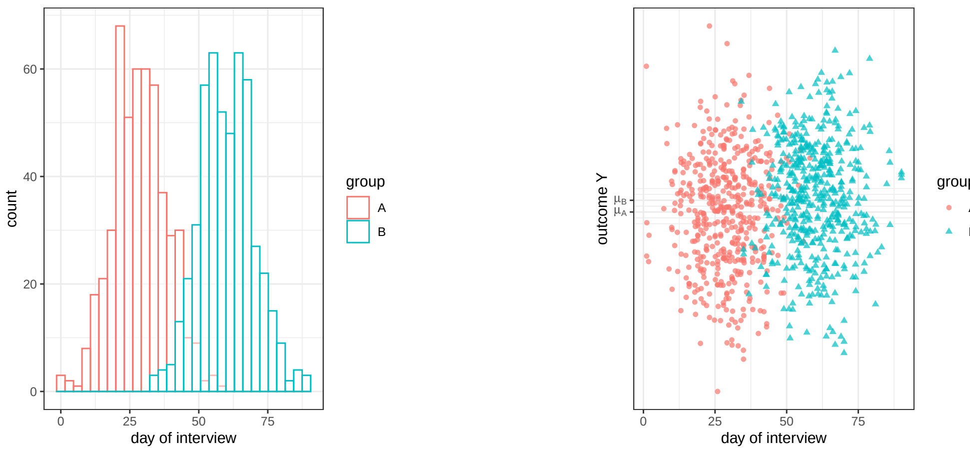
Year	Total	DiD	DiD Share	R/T	G/T	G/G	R/G	T/T
2010	95	5	5.3 %	1	3	0	1	0
2011	115	10	8.7 %	7	3	0	0	0
2012	127	9	7.1%	1	7	1	0	0
2013	101	5	5.0 %	3	2	0	0	0
2014	137	20	14.6 %	7	12	0	1	0
2015	112	13	11.6 %	6	7	0	0	0
2016	114	14	12.3 %	1	9	1	3	0
Sum	*801*	*76*	*9.5 %*	*26*	*43*	*2*	*5*	*0*

Notes: The numbers (except in column 1 and 4) represent frequency counts of papers in the American Economic Review (AER). DiD refers to articles which explicitly mention and apply a difference-in-differences design. In columns 5-8, the DiD counts are subdivided into papers using region (R), time (T), or other time unrelated groups (G) in the design.

Figure 1.1: Examples of Realized Data

(a) Survey Dates

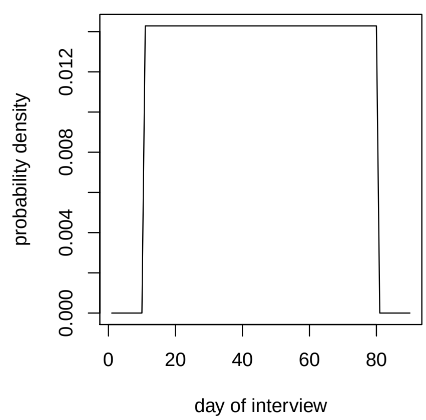
(b) Outcomes



Source: Own depiction. *Notes:* Example distribution of realized survey days during the treatment year for the two groups A and B. The y-axis depicts frequency counts. N=1,000.

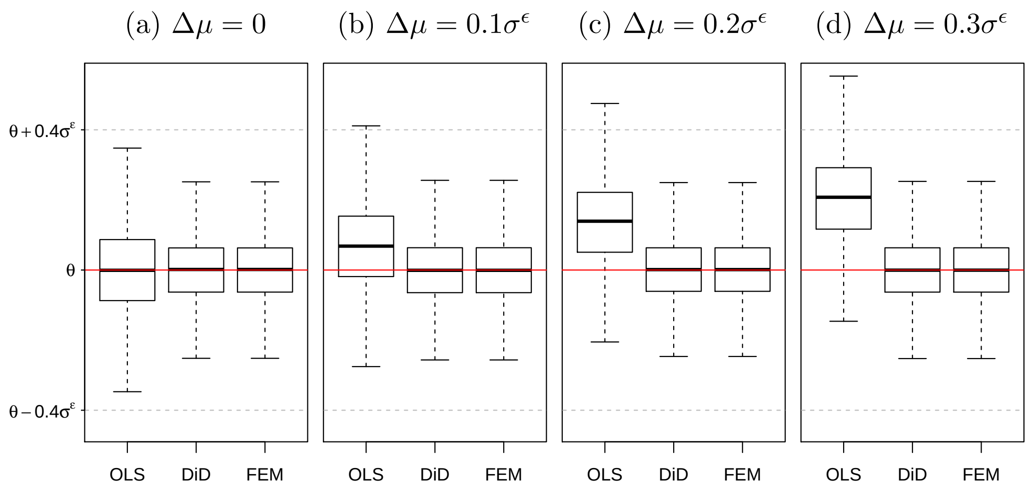
Source: Own depiction. *Notes:* Example of realized outcomes during the treatment year for the two groups A and B. The y-axis depicts the values of the outcome variable Y. N=1,000.

Figure 1.2: Event Timing Distribution



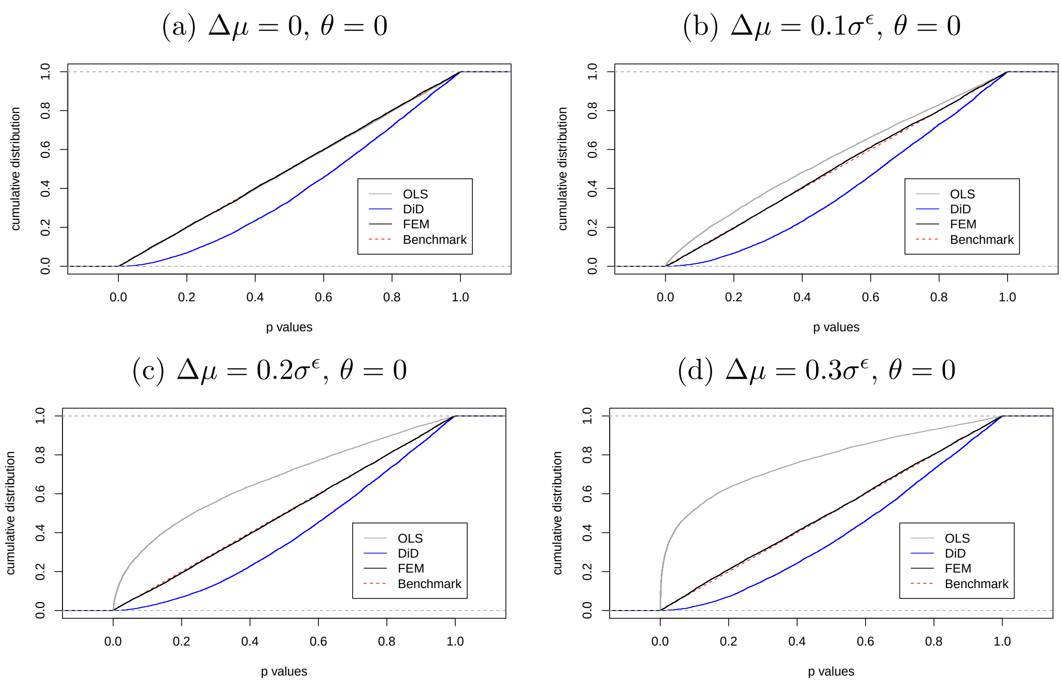
Source: Own depiction. *Notes:* The graph depicts the designed lottery for the event timing. The support of the uniform distribution ranges from 11 to 80.

Figure 1.3: Estimated Treatment Effects



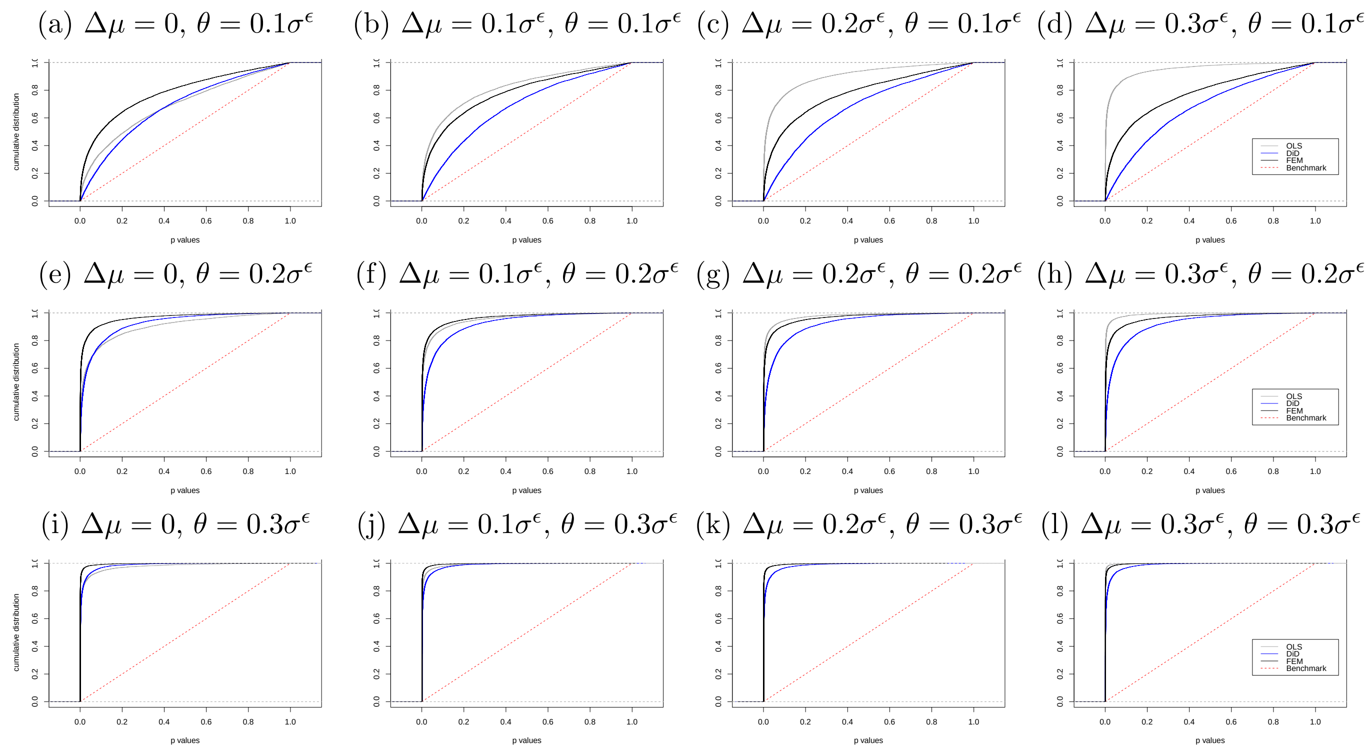
Source: Own depiction. *Notes:* The graphs show the box plots of estimated treatment effects. Each box plot is based on the results of 10,000 regressions, each with 10,000 observations. The label OLS refers to results from estimating Eq. (1.2), DiD refers to estimating Eq. (1.3), and FEM refers to estimating (1.4). The true treatment effect θ is shown by the red line. The mean difference in the distributions of X between groups A and B is $\Delta\mu$.

Figure 1.4: Cumulative Distributions of p-Values



Source: Own depiction. *Notes:* The graphs show the cumulative distributions of p values for estimated treatment effects. Each line is based on the results of 10,000 regressions, each with 10,000 observations. The label OLS refers to results from estimating Eq. (1.2), DiD refers to estimating Eq. (1.3), and FEM refers to estimating (1.4). The Benchmark line is 45 degrees line through the origin. The true treatment effect is θ and $\Delta\mu$ is the mean difference in the distributions of X between groups A and B.

Figure 1.5: Cumulative Distributions of p-Values 2



Source: Own depiction. *Notes:* The graphs show the cumulative distributions of p values for estimated treatment effects. Each line is based on the results of 10,000 regressions, each with 10,000 observations. The label OLS refers to results from estimating Eq. (1.2), DiD refers to estimating Eq. (1.3), and FEM refers to estimating (1.4). The Benchmark line is 45 degrees line through the origin. The true treatment effect is θ and $\Delta\mu$ is the mean difference in the distributions of X between groups A and B.

Table 1.2: Detailed List of AER Articles with Specific Characteristics

#	Author	Title	Year	Category	Variable 1	Variable 2	Variable 3
1	Søren Leth-Petersen	Intertemporal Consumption and Credit Constraints: Does Total Expenditure Respond to an Exogenous Shock to Credit?	2010	Group/Time	Household type	Time	
2	Craig E. Landry, Andreas Lange, John A. List, Michael K. Price and Nicholas G. Rupp	Is a Donor in Hand Better than Two in the Bush? Evidence from a Natural Field Experiment	2010	Group/Time	Type of household / solicitor characteristic	Time	
3	Martha J. Bailey	"Momma's Got the Pill": How Anthony Comstock and Griswold v. Connecticut Shaped US Childbearing"	2010	Region/Time	State	Time	
4	Kenneth A. Couch and Dana W. Placzek	Earnings Losses of Displaced Workers Revisited	2010	Group/Time	Matched individuals	Time	
5	Ann Harrison and Jason Scorse	Multinationals and Anti-Sweatshop Activism	2010	Region/Group/Time	Region / Sector	Export / ownership structure	Time
6	Jeffrey L. Furman and Scott Stern	Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research	2011	Group/Time	Time	Type of article	
7	Mirko Draca, Stephen Machin and Robert Witt	Panic on the Streets of London: Police, Crime, and the July 2005 Terror Attacks	2011	Region/Time	Boroughs	Time	
8	Xiaoquan (Michael) Zhang and Feng Zhu	Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia	2011	Group/Time	Type of contributor	Time	
9	Maximilian Auffhammer and Ryan Kellogg	Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality	2011	Region/Time	County	Time	
10	Atif Mian and Amir Sufi	House Prices, Home Equity—Based Borrowing, and the US Household Leverage Crisis	2011	Region/Time	Region	Time	(Region - smaller resolution)

Continued on next page

Table 1.2 Continued: Detailed List of AER Articles with Specific Characteristics

#	Author	Title	Year	Category	Variable 1	Variable 2	Variable 3
11	Janet Currie	Inequality at Birth: Some Causes and Consequences	2011	Region/Time	Proximity	Time	
12	Shing-Yi Wang	State Misallocation and Housing Prices: Theory and Evidence from China	2011	Group/Time	Household characteristic	Time	
13	John Y. Campbell, Stefano Giglio and Parag Pathak	Forced Sales and House Prices	2011	Region/Time	Proximity	Time	
14	Taryn Dinkelman	The Effects of Rural Electrification on Employment: New Evidence from South Africa	2011	Region/Time	Community	Time	
15	Ruben Enikolopov, Maria Petrova and Ekaterina Zhuravskaya	Media and Political Persuasion: Evidence from Russia	2011	Region/Time	Geographic location	Time	
16	Petra Moser and Alessandra Voena	Compulsory Licensing: Evidence from the Trading with the Enemy Act	2012	Group/Time	Product category / Patent type	Time	
17	Jean-Francois Houde	Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline	2012	Region/Time	Geographic location	Time	
18	Timothy Simcoe	Standard Setting Committees: Consensus Governance for Shared Technology Platforms	2012	Group/Group	Type of request for comments (RFC)	Distributional Conflict	
19	Michael Faye and Paul Niehaus	Political Aid Cycles	2012	Group/Time	Type of administration	Time	
20	Thomas Chaney, David Sraer and David Thesmar	The Collateral Channel: How Real Estate Shocks Affect Corporate Investment	2012	Group/Time	Firm	Time	
21	Santosh Anagol and Hugh Hoikwang Kim	The Impact of Shrouded Fees: Evidence from a Natural Experiment in the Indian Mutual Funds Market	2012	Group/Time	Type of fund	Time	

Continued on next page

Table 1.2 Continued: Detailed List of AER Articles with Specific Characteristics

#	Author	Title	Year	Category	Variable 1	Variable 2	Variable 3
22	Meredith Fowlie, Stephen P. Holland and Erin T. Mansur	What Do Emissions Markets Deliver and to Whom? Evidence from Southern California's NO x Trading Program	2012	Group/Time	Program group / matched facilities	Time	
23	Scott A. Imberman, Adriana D. Kugler and Bruce I. Sacerdote	Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees	2012	Group/Time	School grade characteristic	Time	
24	Esther Duflo, Rema Hanna and Stephen P. Ryan	Incentives Work: Getting Teachers to Come to School	2012	Group/Time	Teacher category	Time	
25	Amit K. Khandelwal, Peter K. Schott and Shang-Jin Wei	Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters	2013	Group/Time	Export good category	Time	
26	Liran Einav, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf and Mark R. Cullen	Selection on Moral Hazard in Health Insurance	2013	Group/Time	Groups of employees	Time	
27	Henrik Jacobsen Kleven, Camille Landais and Emmanuel Saez	Taxation and International Migration of Superstars: Evidence from the European Football Market	2013	Region/Time	Country (synthetic control)	Time	
28	Raj Chetty, John N. Friedman and Emmanuel Saez	Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings	2013	Region/Time	Cities	Time	
29	Matias Busso, Jesse Gregory and Patrick Kline	Assessing the Incidence and Efficiency of a Prominent Place Based Policy	2013	Region/Time	Geographic	Time	
30	Michael Greenstone and Rema Hanna	Environmental Regulations, Air and Water Pollution, and Infant Mortality in India	2014	Region/Time	City	Time	

Continued on next page

Table 1.2 Continued: Detailed List of AER Articles with Specific Characteristics

#	Author	Title	Year	Category	Variable 1	Variable 2	Variable 3
31	Christopher Mayer, Edward Morrison, Tomasz Piskorski and Arpit Gupta	Mortgage Modification and Strategic Behavior: Evidence from a Legal Settlement with Countrywide	2014	Group/Time	Loan type	Time	
32	Sumit Agarwal and Wenlan Qian	Consumption and Debt Response to Unanticipated Income Shocks: Evidence from a Natural Experiment in Singapore	2014	Group/Time	population group	Time	
33	Jeffrey Clemens and Joshua D. Gottlieb	Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?	2014	Region/Time	Region	Time	
34	William Jack and Tavneet Suri	Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution	2014	Group/Time	Type of household	Time	
35	Steven J. Davis, John Haltiwanger, Kyle Handley, Ron Jarmin, Josh Lerner and Javier Miranda	Private Equity, Jobs, and Productivity	2014	Group/Time	Type of firm	Time	
36	Katrina Jessoe and David Rapson	Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use	2014	Group/Time	Household group	Time	
37	Ing-Haw Cheng, Sahil Raina and Wei Xiong	Wall Street and the Housing Bubble	2014	Region/Group	Location of property	Type of market participant	
38	Koichiro Ito	Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing	2014	Region/Time	Provider territory	Time	
39	Jason Brown, Mark Duggan, Ilyana Kuziemko and William Woolston	How Does Risk Selection Respond to Risk Adjustment? New Evidence from the Medicare Advantage Program	2014	Group/Time	Time spent in Medicare Advantage	Time	

Continued on next page

Table 1.2 Continued: Detailed List of AER Articles with Specific Characteristics

#	Author	Title	Year	Category	Variable 1	Variable 2	Variable 3
40	Elliot Anenberg and Edward Kung	Estimates of the Size and Source of Price Declines Due to Nearby Foreclosures	2014	Region/Time	Proximity	Time	
41	Nathan Nunn and Nancy Qian	US Food Aid and Civil Conflict	2014	Region/Time	Country	Time	
42	Dominique Goux, Eric Maurin and Barbara Petrongolo	Worktime Regulations and Spousal Labor Supply	2014	Group/Time	Employee category	Time	
43	Robert W. Fairlie, Florian Hoffmann and Philip Oreopoulos	A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom	2014	Group/Time	Minority status student	Minority status instructor	
44	Liran Einav, Dan Knoepfle, Jonathan Levin and Neel Sundaresan	Sales Taxes and Internet Commerce	2014	Region/Time	State	Time	
45	Alessandro Tarozzi, Aprajit Mahajan, Brian Blackburn, Dan Kopf, Lakshmi Krishnan and Joanne Yoong	Micro-Loans, Insecticide-Treated Bednets, and Malaria: Evidence from a Randomized Controlled Trial in Orissa, India	2014	Region/Time	Geographic location	Time	
46	Petra Moser, Alessandra Voena and Fabian Waldinger	German Jewish Émigrés and US Invention	2014	Group/Time	Technology Category	Time	
47	Dina Mayzlin, Yaniv Dover and Judith Chevalier	Promotional Reviews: An Empirical Investigation of Online Review Manipulation	2014	Group/Time	Review site	Neighborhood and ownership / affiliation	
48	Andreas Ravndal Kostøl and Magne Mogstad	How Financial Incentives Induce Disability Insurance Recipients to Return to Work	2014	Group/Time	Entry into insurance (2 months around date)	Entry into insurance in 1-year steps	
49	Alberto Alesina and Eliana La Ferrara	A Test of Racial Bias in Capital Sentencing	2014	Group/Time	Race of victim	Race of defendant	
50	Rafael Lalive, Camille Landais and Josef Zweimüller	Market Externalities of Large Unemployment Insurance Extension Programs	2015	Group/Time	Employee category	Age	Time
51	Neale Mahoney	Bankruptcy as Implicit Health Insurance	2015	Region/Time	State	Time	

Continued on next page

Table 1.2 Continued: Detailed List of AER Articles with Specific Characteristics

#	Author	Title	Year	Category	Variable 1	Variable 2	Variable 3
52	Esther Ann Bøler, Andreas Moxnes and Karen Helene Ulltveit-Moe	R&D, International Sourcing, and the Joint Impact on Firm Performance	2015	Group/Time	Type of firm	Time	
53	Janet Currie, Lucas Davis, Michael Greenstone and Reed Walker	Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings	2015	Group/Time	Geographical Location	Time	
54	Serguey Braguinsky, Atsushi Ohyama, Tetsuji Okazaki and Chad Syverson	Acquisitions, Productivity, and Profitability: Evidence from the Japanese Cotton Spinning Industry	2015	Group/Time	Type of firm	Time	
55	Danny Yagan	Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut	2015	Group/Time	Type of company	Time	
56	Lucija Muehlenbachs, Elisheba Spiller and Christopher Timmins	The Housing Market Impacts of Shale Gas Development	2015	Region/Time	Proximity	Area	Time
57	Martin B. Hackmann, Jonathan T. Kolstad and Amanda E. Kowalski	Adverse Selection and an Individual Mandate: When Theory Meets Practice	2015	Region/Time	State	Time	
58	Katherine Casey	Crossing Party Lines: The Effects of Information on Redistributive Politics	2015	Region/Time	Geographic location	Type of election	Radio ownership
59	Dina Pomeranz	No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax	2015	Group/Time	Experiment group	Time	
60	Claudio Michelacci and Hernán Ruffo	Optimal Life Cycle Unemployment Insurance	2015	Region/Time	State characteristics	Time	
61	Alessandra Voena	Yours, Mine, and Ours: Do Divorce Laws Affect the Intertemporal Behavior of Married Couples?	2015	Region/Time	State	Time	

Continued on next page

Table 1.2 Continued: Detailed List of AER Articles with Specific Characteristics

#	Author	Title	Year	Category	Variable 1	Variable 2	Variable 3
62	Michael Kosfeld and Devesh Rustagi	Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia	2015	Group/Time	Type of leader	Time	
63	Justin R. Pierce and Peter K. Schott	The Surprisingly Swift Decline of US Manufacturing Employment	2016	Group/Time	Industry	Time	
64	Sandra Sequeira	Corruption, Trade Costs, and Gains from Tariff Liberalization: Evidence from Southern Africa	2016	Group/Time	Product category	Time	
65	Bryan Kelly, Hanno Lustig and Stijn Van Nieuwerburgh	Too-Systemic-to-Fail: What Option Markets Imply about Sector-Wide Government Guarantees	2016	Group/Time	Financial market sector	Time	
66	Peter Koudijs and Hans-Joachim Voth	Leverage and Beliefs: Personal Experience and Risk-Taking in Margin Lending	2016	Group/Time	Type of lender	Time	
67	Massimo Bordignon, Tommaso Nannicini and Guido Tabellini	Moderating Political Extremism: Single Round versus Runoff Elections under Plurality Rule	2016	Region/Type	Municipality	Time	
68	Atila Abdulkadiroğlu, Joshua D. Angrist, Peter D. Hull and Parag A. Pathak	Charters without Lotteries: Testing Takeovers in New Orleans and Boston	2016	Group/Time	School	Time	
69	Hilary Hoynes, Diane Whitmore Schanzenbach and Douglas Almond	Long-Run Impacts of Childhood Access to the Safety Net	2016	Region/Cohort	County	Year of birth	
70	Arik Levinson	How Much Energy Do Building Energy Codes Save? Evidence from California Houses	2016	Group/Group	House vintage	Experienced temperature / state	
71	Naomi E. Feldman, Peter Katuščák and Laura Kawano	Taxpayer Confusion: Evidence from the Child Tax Credit	2016	Cohort/Time	Day of birth	Time	

Continued on next page

Table 1.2 Continued: Detailed List of AER Articles with Specific Characteristics

#	Author	Title	Year	Category	Variable 1	Variable 2	Variable 3
72	Sibylle Lehmann-Hasemeyer and Jochen Streb	The Berlin Stock Exchange in Imperial Germany: A Market for New Technology?	2016	Group/Time	IPO status (firm categories)	Time	
73	Paula Bustos, Bruno Caprettini and Jacopo Ponticelli	Agricultural Productivity and Structural Transformation: Evidence from Brazil	2016	Region/Time	Municipality characteristics	Time	
74	Jessica Calfee Stahl	Effects of Deregulation and Consolidation of the Broadcast Television Industry	2016	Group/Time	Station	Time	
75	Lawrence Schmidt, Allan Timmermann and Russ Wermers	Runs on Money Market Mutual Funds	2016	Group/Time	Type of fund	Time	
76	David Atkin	Endogenous Skill Acquisition and Export Manufacturing in Mexico	2016	Region/Cohort	Geographic location characteristic	Cohort	

Chapter 2

A Swing to the Right? The Refugee Crisis in Germany

2.0 Abstract

This paper analyzes whether Germans' attitudes toward immigration were affected and polarized by the so-called 2015 refugee crisis in the short run. Applying a difference-in-differences design to German panel data shows that the crisis increased individuals' average likelihood to be concerned about immigration and the likelihood to express anti-immigration sentiment significantly in both practical and statistical terms. Moreover, the average propensity to vote the Alternative for Germany increased, whereas the average probability to report approval of Chancellor Merkel decreased. Potential polarization in the latter two measures is tracked using a new machine learning approach and the most reactive groups are characterized demographically. Although the effects of the crisis entail significant heterogeneity, no polarization of attitudes is found. The results therefore suggest a general swing to the right in political attitudes.[1] (JEL D72, J15, H12)

[1] All calculations and figures in this chapter using the German Socio-Economic Panel are created with the software Stata 14. All calculations and figures in this chapter using the Politbarometer are created with the software package Python, version 3.7.6, in particular the package scikit-learn, and the open-source distribution of Anaconda.

2.1 Introduction

Anti-immigration claims are increasingly successful in elections. Examples are the Brexit campaign, Donald Trump's presidential election, Hungary's Fidesz, Poland's Law and Justice (PiS), Italy's Lega Nord, the National Rally in France, and the Party for Freedom in the Netherlands. The objective of this paper is to understand how the so-called refugee crisis impacted on Germans' attitudes toward immigration and the right-wing populist party Alternative for Germany (AfD).

About two million foreigners moved to Germany in 2015, and more than 800,000 of these were asylum seekers and refugees (BAMF, 2016). This constitutes an all-time high in immigration and refuge to Germany since the beginning of the records and makes Germany the largest provider for foreign refugees in Europe and America (UNHCR, 2020). In the course of the year 2015, a highly sensitive debate about asylum and migration policies flared up in Germany, a debate that was (and still is) penetrated regularly by the much-noticed outbursts of the AfD.

The public awareness of the high influx to Germany erupted in August and September 2015. After the German government raised its forecast of incoming refugees to 800,000 on August 19, it decided to receive several thousands of refugees who at the time resided in Hungary at the beginning of September. This has been interpreted by many as a signal of an open-door policy for refugees and illegal migrants (Blume et al., 2016), a state of affairs that has been called refugee crisis. I interpret the crisis as a shock to the public awareness of the high inflow of migrants and refugees to Germany.

A growing body of literature is analyzing various aspects of the refugee crisis.

Hatton (2017) reviews the political obstacles regarding the European refugee crisis and discusses potential solutions, and first empirical analyses investigate the impact of Syrian refugees on Turkey's labor market (Tumen, 2016; Akgündüz and Torun, 2018). The paper at hand analyzes the short-term effects of the 2015 refugee crisis on the strength of concerns about immigration in Germany, the rise of the AfD, and the increasing disapproval of Chancellor Angela Merkel.

During the work on this study, a very related working paper by Sola (2018) arose which analyzes the same effect and reaches similar conclusions. I can contribute to the discussion beyond this by providing a very clean identification of the beginning of the crisis, therefore avoiding attenuation of the estimated effect, and by isolating concerns which can clearly be attributed to anti-immigration sentiment from more general concerns. To my knowledge, this is the first study to show that the refugee crisis both raised anti-immigration concerns and increased the poll rating of the AfD. Moreover, it can be shown that the satisfaction with Chancellor Merkel's work decreased by a very comparable fraction. Potential polarization in these measures of attitude is analyzed along socio-demographic characteristics using a new machine learning approach proposed by Chernozhukov et al. (2018), and the most reactive groups are characterized in detail.

Using the German Socio-Economic Panel (SOEP), a difference-in-differences strategy is applied as discussed by Dammann (2020). It compares individuals in the SOEP who were interviewed during the refugee crisis in 2015 (treatment group) with those who were interviewed earlier in the year (control group). Because the treatment group might be selected and special in many dimensions, the strategy exploits that both these groups can be observed in the SOEP survey years prior to 2015. The estimation results show that the refugee crisis increased the likelihood to be very concerned about immigration to Germany by about 21 percentage points above the general upward trend of about 9 percentage points in 2015, and raised the likelihood of having clear anti-immigrant concerns by about 7 percentage points.

Moreover, a second data set, the German Politbarometer, is scrutinized to learn the effect of the crisis on AfD poll success and on the approval rate of Chancellor Angela Merkel applying simple pre-post comparisons. It is found that the crisis increased the likelihood of voting the AfD by an average of approximately two percentage points during the time of the crisis, and lowered the likelihood to approve Chancellor Merkel by an average of about seven percentage points. Applying a new machine learning approach suggested by Chernozhukov et al. (2018) reveals significant heterogeneity in these effects. In particular, there are groups in the population who are much more likely to vote for the AfD and much less likely to approve of Merkel as a result of the crisis than the average. These findings are of particular interest because they are data-driven without imposing any ex-ante assumptions about functional forms or the relevancy of factors to the data. A classification analysis reveals an over-representation of males, middle school graduates, East Germans, non-Christians, and older people (ages 50-70) among those who are most likely to vote for the AfD and to disapprove Merkel during the crisis.

The empirical design that is applied to the Politbarometer bears enough flexibility to detect potential polarization in political attitudes as a response to the refugee crisis. Polarization might occur if some individuals become more anti-immigration, while others become more pro-immigration. For instance, since Merkel was popular for her pro-refugee attitude, some individuals might become more disapproving of her while others might become more approving of her. However, the empirical findings do not suggest polarization of attitudes but rather a general swing to right, i.e., a swing towards being more critical of immigration. Based on demographic characteristics, no subgroup can be identified which is less likely to vote for the AfD or more likely to approve Merkel.

Additional evidence from the Politbarometer is consistent with the predictions that labor market concerns, group conflict, social identity, and hatred pave the way for the negative reactions to the refugee crisis. In particular, groups that are socio-economically more vulnerable to labor market competition react stronger to the refugee crisis. Furthermore, right-wing voters are more likely to be concerned that government spending on refugees comes at the expense of others, that more refugees increase criminality and that refugees jeopardize societal values in Germany.

The refugee crisis provides the opportunity to analyze the effect of a rare change in more general societal conditions on attitudes over time. This study helps to better understand that attitudes toward immigration can react strongly to political crises in the short run. Moreover, it shows that the political polarization, which is often discussed in public media, did not increase quantitatively during the refugee crisis. To the best of my knowledge, this is the first study that analyzes the causal effect of the 2015 refugee crisis on political attitudes in combination with the voting success of the AfD and the approval of Chancellor Merkel.

The next section provides additional information on and an empirical definition of the crisis. Section 3 reviews related literature and develops hypotheses about the potential effects of the crisis. Section 4 discusses the data and empirical designs in detail as well as the estimation results and robustness checks. Section 5 concludes on the analysis. All figures and tables are appended in Section 6.

2.2 The Refugee Crisis

The inflow of non-Germans into Germany increased throughout the entire year 2015 (BAMF, 2016). Because increasing migratory or refugee inflows are not a problem *per se*, one needs to define at which point this inflow was considered to become a problem or even a crisis. To find this point in time, Figure 2.1 shows the number of articles mentioning the term 'refugee crisis' in five major German newspapers on a weekly basis. It is obvious that in all these newspapers the term is hardly present before August, that its appearance skyrockets during the end of August and that it remains high afterwards.

As an indication of the public interest in the topic, Figure 2.2 shows the Google Trends indices for the two terms 'refugee' and 'migrant' in Germany throughout the year 2015. The search indices depict the number of searches per week relative to the highest point of the chart, i.e., a value of 100 indicates the highest number of searches during the observation period, a value of 50 indicates half the number of searches as compared to the highest point, and so forth. It can be seen that both search indices increase during the end of August, show a pronounced peak in September, and stay high afterwards. Although the search for the term 'migrant' is more frequent and more volatile relative to its peak than the search for the term 'refugee', it is clear that the public interest in both terms receive a significant stimulus from the refugee crisis. Hence, I interpret the refugee crisis as a shock to the public awareness about the high inflow of refugees and migrants to Germany.

Media reports have carefully reconstructed the events that were henceforth called the refugee crisis in Germany (see, for instance, the comprehensive report in Blume et al., 2016). On August 19, the federal government raised the expected number of refugee arrivals during 2015 from 450,000 to 800,000, giving rise to first speculations about a refugee crisis in the media. This can also be seen in Figure 2.1, where the vertical red line separates the week of August 19 from previous observations. On August 31, Chancellor Merkel held a much-noticed speech on the situation of the refugee inflow, which was the first time when she stated her famous dictum of 'we can do it.' This dictum was meant to express that Germany could handle the refugee inflows that were about to come. Even more publicity for the topic was reached during the weekend after September 4, when thousands of refugees who at the time resided in Hungary approached Germany, and the German government decided to let them cross the German border. This has been interpreted by many as a signal of an open-door policy for refugees and illegal migrants. Regarding these historical dates in combination with the indicators for public awareness and interest in the topic, I define the start of the refugee crisis in Germany as August 2015.

Between September and December 2015, an influx of 678,359 asylum seekers was registered via the German EASY system (BAMF, 2016),[2] which constitutes an un-

[2]The EASY system is a government tool to distribute applicants for asylum across federal states in Germany. Please note that this number is probably overstating the true inflow due to multiple

precedented inflow of non-Germans to the Federal Republic of Germany in such a short period of time. A heated public debate emerged which focused on topics like the possibility of border controls, the duty to help people in need, the right to migrate and the right to direct migration, and cultural identity. While parts of the German population strongly advocated a pro-refugees and pro-immigration welcoming culture, the so-called *Willkommenskultur*, other parts of the population where very critical of these developments. According to the Federal Crime Police Office in Germany, the refugee crisis is linked to agitation and numerous politically motivated crimes, for instance infringements against refugee camps (Bundeskriminalamt, 2016). Thus, the general impression one could gain during the crisis was a strong dichotomy in attitudes toward immigration and in the reactions to the refugee crisis.

An omnipresent participant in the public debate was the AfD. The party was founded in 2013 as a non-parliamentary opposition to the Eurozone policies. During the refugee crisis, the party enhanced its far-right wing profile by starkly opposing the pro-refugee policies of the German government. The party's then-chairwoman even suggested to use armed forces against refugees to protect the German border. Eventually, the party was elected into parliament in 2017 and fills 89 of 709 seats in the current (2019) Bundestag.[3]

registrations of the same individuals. However, it represents the impression of the size of the inflow at the time of the crisis.

[3]Note that the AfD was initially awarded 94 seats in 2017. The party lost seats because some parliamentarians left the party.

2.3 Theoretical Considerations

Economists, sociologists and psychologists have formulated various theories about the determinants of attitudes toward immigration. The approach most familiar to economists is a simple consideration of the demand and supply of labor. If labor demand curves slope downwards and immigration increases the supply of labor, then immigration will lower natives' wages (Borjas, 2003).[4] In an economy like Germany where high-skill labor is relatively abundant and immigration is to a great extent low-skilled, the expectation is that low-skill or low-income workers are more likely to be against immigration (Mayda, 2006). Moreover, xenophobic, anti-Semitic and racist attitudes are magnified when people think that their personal economic prospect is bad (Mocan and Raschke, 2016).

The labor demand and supply considerations are one version of group conflict theory, which assumes that an in-group's attitudes toward an out-group will be hostile if the two groups compete for scarce resources (see, e.g., Sherif and Sherif, 1953). Sudden changes in minority group sizes change the level of competition and, therefore, the attitudes toward immigration (Meuleman et al., 2009). In particular, one would expect socio-economically vulnerable groups to react more hostile toward foreigners due to the refugee crisis than less vulnerable groups (Lancee and Pardos-Prado, 2013). This prediction will be tested in the empirical analysis.

Even without competing for scarce resources, hostility toward out-groups can arise according to social identity theory. Following Akerlof and Kranton (2000), the

[4]Although the actual effect of immigration on the wage structure is disputable (Dustmann et al., 2016), a negative perception of immigration may arise from the expectation of downward-sloping demand curves for labor.

fact that the refugee crisis increased the group of non-Germans in Germany may be perceived as a threat to identity by natives, who would then suffer a loss in their sense of self. These individuals' attitudes toward immigration might in turn become (more) negative. This is consistent with empirical evidence that the size of inflows of asylum seekers is negatively associated with average attitudes toward immigration in cross-country comparisons (Mayda, 2006).

Related to both group conflict and identity theory, Glaeser (2005) constructs a political economy of hatred. In his model, an in-group forms beliefs about the likelihood that an out-group will impose some cost on each member of the in-group, which in turn causes hatred. Politicians may have an incentive to supply hate-creating stories against the out-group to increase their share of the vote. An increase in minority group size raises the incentive to broadcast hate-creating stories as well as the support for anti-immigration candidates. An example for hate-creating stories is the allegation that foreigners are criminal and that immigration therefore raises crime rates. Thus, if the economy of hatred plays a role for the development of attitudes during the refugee crisis, concerns about criminality will most likely increase. Whether they eventually do will be tested below.

2.4 Empirical Analysis

The empirical analysis uses two data sets that complement each other. The first data set is the German Socio-Economic Panel (SOEP), which provides information on attitudes toward immigration and exhibits a data structure that enables a clean causal identification. However, the SOEP documents self-reported attitudes and is thus far from observing actual behavior. Moreover, the majority of SOEP participants are interviewed in spring and the sample becomes rather small during the crisis, making it difficult to perform subgroup analyses. The second data set is the Politbarometer, which provides rich information on declared voting behavior. The Politbarometer encompasses reasonably consistent cross-sectional samples over the whole year such that although the identification is not as clean as in the SOEP, the sample is sufficiently large for heterogeneity analyses. It is shown that the clean identification with the SOEP yields very similar results as the less clean estimation with the Politbarometer, and the Politbarometer is then used to scrutinize the heterogeneity of the effect.

2.4.1 Attitudes Toward Immigration - Panel Data

The information about the attitudes toward immigration stem from the German Socio-Economic Panel (SOEP). The SOEP is described in great detail by Wagner et al. (2007). It contains the question: 'How concerned are you about the following issues? - Immigration to Germany.' Respondents could answer that they are very concerned, somewhat concerned, or not concerned at all. I define a dummy variable VC^I that takes on the value one if a respondent reports to be very concerned about immigration to Germany, and zero in the other two cases.

I estimate the causal effect of the refugee crisis on individuals' attitudes toward migration using a difference-in-differences design as discussed in Dammann (2020). The design exploits the fact that some individuals in the 2015 wave of the SOEP were interviewed before the events that were henceforth referred to as the refugee crises, and some were interviewed afterwards. Moreover, most of these individuals were also interviewed in the years prior to 2015. I define the observations before 2015 as pre-treatment and the observations in 2015 as post-treatment observations. The scientific use file of the SOEP contains information of the month of interview, and I exclude all individuals from the analysis who where interviewed in August 2015 since it is not clear whether these should be classified as treatment or control units. I define the respondents who were interviewed before August 2015 as the control group and those respondents interviewed during or after September 2015 as the treatment group. To ensure a reasonable representation of the treatment group, I include pre-treatment observations ranging from 2012 to 2014.[5] Included individuals are required to be observed in 2015 and in at least two out of the three pre-treatment years.[6]

Abadie et al. (2017) discuss when and why it is necessary to adjust standard errors for clustering. It is necessary if either the sample is clustered such that the ex ante probability of cluster C being part of the sample is less than one, or if the treatment assignment mechanism is such that some clusters are treated with a higher probability. It is unlikely that the treatment mechanism is clustered in the

[5] The question I use to measure attitudes toward immigration was not posed to all sub-samples of the SOEP in all these survey years. Thus, the data are restricted to the sub-samples to which it was stated in all years between 2012 and 2015 to prevent structural changes in either the control or treatment group between any two years. Survey years before 2012 are excluded because this would either decrease data size significantly under the condition that all included samples were interviewed about their concerns about immigration in all sampled years, or it would change the composition of the data between 2011 and 2012.

[6] Note that although far less information are available under a fully balanced sample, the main results remain stable when only using individuals with full record.

population since all individuals are eventually treated by the crisis. However, the SOEP sampling is clustered in a two-stage procedure. At the first stage, regions C are drawn from a set of regions. The SOEP calls these regions primary sampling units (Spiess and Kroh, 2007) for which, ex ante, $Prob(C \text{ in sample}) < 1$. At the second stage, households are randomly drawn from the selected primary sampling regions. Following Abadie et al. (2017), I adjust standard errors in the SOEP for clustering at the region of primary sampling from the first stage of the sampling procedure.

The identifying assumption is that, absent the refugee crisis, the attitudes of those interviewed after September 2015 would have developed along the same trend as the attitudes of those interviewed before September 2015. Panel (a) of Figure 2.3 shows the development of VC^I for both the treatment and the control group from 2012 to 2015. It can be seen that the two groups develop decently parallel in the three years prior to the crisis, which is reassuring regarding the identifying assumption.[7] In 2015, the treatment group's indicator for being very concerned increases rapidly as compared to the prior common trend, suggesting a strong treatment effect. Table 2.1 provides descriptive statistics for the two groups and shows that the difference in differences between the treatment and control group amounts to about 20.4 percentage points and that it is statistically significant at the 1 percent level. At the same time, the difference in differences for the baseline characteristics gender, age, and living in East Germany are practically small and statistically insignificant.

[7]None of the small deviations from a perfect common trend is statistically significant at the 10 percent level, see Table 2.7.

Table 2.2 reports the results for the fixed-effects regression

$$y_{it} = \alpha_i + \theta_t + \beta treat_i * year2015_t + u_{it}, \tag{2.1}$$

where y_{it} is the outcome for individual i in survey year t, α_i is an individual-fixed effect, θ_t is a survey year-fixed effect, and $treat_i * year2015$ is the treatment group-treatment period interaction. Thus, β is supposed to measure the causal effect of the refugee crisis on the outcome. The first column of Table 2.2 confirms the descriptive findings for VC^I with a point estimate of 20.5 percentage points for β. This effect is statistically different from zero at very small significance levels.

An increase in the concerns about immigration may both represent pro- and anti-immigration concerns. Especially in light of the heated public debate and severe crimes against refugee homes at the time, increased concerns about immigration may as well represent increasing worries about xenophobia. The SOEP provides a measure of concern regarding xenophobia analogous to the measure of concern about immigration, and I define VC^X as an indicator for being very concerned about xenophobia, as opposed to being somewhat concerned or not concerned at all. Individuals who report concerns about immigration and who are not concerned about xenophobia despite the increase in xenophobic violence during the crisis are very likely to be anti-immigrant. To isolate concerns that can unambiguously be attributed to anti-immigration sentiment, I use $VC^I(1 - VC^X)$ as a new measure of concern that equals one if a respondent is very concerned about immigration but not very concerned about xenophobia, and zero in all other cases.

Panel (b) of Figure 2.3 shows the group trends for this measure over the observation period. Treatment and control group develop parallel during the pre-treatment periods, and the treatment group deviates upwards in the treatment period. The comparison of mean values in Table 2.1 shows that the difference in differences for $VC^I(1 - VC^X)$ averages to about 6.8 percentage points, an estimate that is statistically distinguishable from 0 at the 5 percent level. Adding individual and year fixed effects yields a point estimate of 6.7 percentage points that is significant at the 5 percent level, as shown in the second column of Table 2.2. Thus, the refugee crisis is estimated to increase the likelihood of having unambiguous anti-immigration attitudes by about 6.7 percentage points.

Note, however, that being worried about xenophobia does not necessarily indicate pro-immigrant sentiment. Increasing concerns about xenophobia may relate to concerns about xenophobic violence and crimes at the time of the crisis and may therefore just indicate an individual's refusal of violence and criminal activity in general. Hence, while the difference between the increase in VC^I and the increase in $VC^I(1 - VC^X)$ suggests a potential for polarization in the attitudes toward immigrants, it does not prove the existence of polarization. The following subsections seek to shed more light on potential polarization.

2.4.2 AfD Polls and Satisfaction with Chancellor Merkel - Repeated Cross Sections

Changes in the attitudes toward immigration may be of little impact if they do not change behavior. To check whether political behavior was affected as well, I analyze data from the German Politbarometer. The Politbarometer is a representative repeated cross-sectional survey that takes place at least once, usually twice per month. It is affiliated with the German public television channel ZDF and is widely recognized by the German public. It contains individual-level data on many questions, including the famous question which party a participant would vote for if elections were to take place next Sunday. This enables me to measure the effect of the crisis on AfD opinion poll success and other political outcomes directly. For the year 2015, 18 independent cross sections including a total of 30,051 individual interviews are available of which 21,068 contain information on all relevant variables.

Each cross-section in the Politbarometer is a representative random draw of the German population, where households are randomly drawn from the directory of landline phone numbers. The sample design has the benefit that the data support is relatively continuous throughout the entire year. The disadvantage is the lacking panel structure, which makes it impossible to use the same difference-in-differences approach as in the SOEP and only allows single difference before and after comparisons. However, as opposed to the SOEP, the treatment group is not selected based on the interview timing because new random samples are drawn for every survey.

Of particular interest in light of the refugee crisis is the vote share of the AfD, which is the first outcome to be studied. I define the variable AfD to equal one if a person would vote for this party next Sunday, and zero otherwise. Another interesting measure that is informative about political attitudes is contained in the Politbarometer, namely the statement whether or not a participant beliefs that Chancellor Merkel does a good job. Merkel was strongly associated with an open-border regime at the time of the crisis and was famous for her dictum of ‘we can do it,’ leading some media to even call her the Refugee Chancellor. Thus, individuals who become less satisfied with Merkel because of the refugee crisis are likely to be disappointed with her pronounced pro-refugees and pro-migration attitude. A decrease in the satisfaction with her over the course of the refugee crisis may thus be interpreted as an expression of anti-immigration sentiment. At the same time, an increase in the satisfaction with Merkel may be interpreted as an expression of pro-immigrant sentiment, which makes this question particularly interesting for the analysis of potential polarization. The variable $Merkel$ is defined to take on the value one if a person is satisfied with Merkel's work, and zero otherwise.

Because only before and after comparisons can be carried out, the identifying assumption is that survey timing is not correlated with other factors than the refugee crisis. Table 2.3 shows that even without any controls, the timing of the refugee crisis is not statistically significantly correlated with the baseline characteristics available in the Politbarometer including gender, school degree, living in East Germany, marital status, being Christian, being unemployed, and having children. The only difference between the treatment and control group that is significant at the 5 percent level is a slight difference in the proportion of the age groups of 40-44 years. Eventually, the timing of the crisis does not seem to be correlated with individual characteristics.

Using Machine Learning to Detect Heterogeneity and Polarization

The previous results show that the average individual became more critical of immigration because of the refugee crisis. This suggests that, on average, the refugee crisis should increase the support of the AfD and reduce the satisfaction with Merkel's work. However, not all individuals need to react in the same way to the rising awareness about a high migratory and refugee inflow. Suppose that each individual has a prior attitude toward immigrants to Germany, and that each individual updates her beliefs after learning about the new developments during the end of August 2015. Polarization due to the refugee crisis will occur if the new information about the high inflows lead some individuals to become more pro-refugee, and others to become more anti-refugee.[8]

Although machine learning has relatively little to offer for the identification of average treatment effects, it is a strong tool for detecting treatment effect heterogeneity (Athey, 2018). Moreover, both picking the relevant factors along which attitudes might diverge and choosing the functional form to fit the effect heterogeneity are relatively arbitrary choices that are usually taken by the researcher. To avoid such decisions, I apply machine learning tools and a data-driven approach to learn about potential effect heterogeneity.[9]

I apply the approach of Chernozhukov et al. (2018, CDDF henceforth) and use random forest classifiers for both estimating the average treatment effect and detecting it's heterogeneity in the Politbarometer (see, e.g., Efron and Hastie, 2016, for a discussion of random forests). The CDDF approach allows to apply any generic machine learning tool and does not impose assumptions on these tools other than that they have predictive power for the outcome. Most importantly, the CDDF approach provides guidance for valid inference when applying machine learning prediction tools

[8] See, for instance, Acemoglu et al. (2016) and Dixit and Weibull (2007) for theoretical definitions and discussions of the concept of political polarization.

[9] All calculations in this subsection are done in Python using the scikit-learn package.

without making assumptions about how these tools work and facilitates a data-driven exploration of heterogeneity in the treatment effect. Because it is relatively unknown so far, I discuss the approach of CDDF in detail in this subsection and thereby outline the results and their interpretation.

Following CDDF, suppose that the potential outcomes are y^0 in absence of the refugee crisis and y^1 in presence of the refugee crisis. Moreover, characteristics Z of the individuals can be observed and the average treatment effect might be heterogeneous along these characteristics. The baseline conditional average is given by $b_0(Z) = E[y^0|Z]$, and the conditional average treatment effect (CATE) is given by $s_0(Z) = E[y^1 - y^0|Z]$.

Using random forests to estimate $b_0(Z)$ and $s_0(Z)$ is generally known to deliver inconsistent results. The approach of CDDF focuses on learning key features about $b_0(Z)$ and $s_0(Z)$ rather than the functions themselves by estimating the inconsistent proxies $B = B(Z)$ for $b_0(Z)$ and $S = S(Z)$ for $s_0(Z)$ with machine learning tools and projecting them back on the observed outcome y.

The first step to implement the CDDF approach is to randomly split the data into an auxiliary sample A and a main sample M, each of which I assign to encompass half of the data. The tuning and training of the random forest is performed on sample A and the trained random forest is used to predict B and S in sample M. To avoid the problem of over-fitting, all inference is done using the main sample, leaving 10,534 observations for inference.[10] For a given data split, three steps are performed once the random forest is trained. First, the best linear prediction of

[10] The random data splits are stratified on the treatment status to ensure a representative presence of post-treatment observations.

$s_0(Z)$ based on S is calculated. Second, the sorted group average treatment effects (GATES) are calculated, for instance as the average treatment effect for the most affected group and the average treatment effect for the least affected group. Third, a classification analysis (CLAN) is performed to learn the average characteristics of groups of different treatment impact, for instance the average characteristics of the most and the least affected groups.

To account for the uncertainty of the results arising from randomly splitting the sample, CDDF suggest to repeat the fitting and estimations over many different random splits of the data into A and M. From the set of all results for the GATES, CLAN, and best linear projections of the CATE for the different sample splits, the median is used as the point estimator. Analogously, the median of the p-values is calculated and adjusted upwards to account for the additional uncertainty. CDDF call this procedure variational estimation and inference method (VEIN). In this analysis, I fix the number of sample splits to 100.

Best Linear Predictor Suppose that as in case of the SOEP, the treatment status is given by the binary indicator $treat$, and $p(Z) = P(treat = 1|Z)$ denotes the treatment propensity. Following CDDF, the treatment propensities $p(Z)$ are obtained by linearly predicting $treat$ given Z in the entire data set. Thus, any confounding factors that may be contained in Z are indirectly controlled for in the following regressions. Given a random split into A and M, the best linear prediction of the CATE given S can, according to CDDF, be obtained by estimating the equation

$$y = \alpha_0 + \alpha_1 B + \beta_1(treat - p(Z)) + \beta_2(treat - p(Z))(S - E[S]) + u, \tag{2.2}$$

$$s.t.\ E[(p(Z)(1 - p(Z))^{-1} uX] = 0, \tag{2.3}$$

where X is a vector containing B, $treat - p(Z)$, and $(treat - p(Z))(S - E[S])$. It follows that

$$\beta_1 = E[s_0(Z)],\ \text{and} \tag{2.4}$$

$$\beta_2 = Cov(s_0(Z), S)/Var(S) \tag{2.5}$$

(CDDF, Theorem 2.1). Thus, while β_1 is supposed to measure the average treatment effect, rejecting that $\beta_2 = 0$ implies that the treatment effect is heterogeneous and that S is a relevant predictor for that heterogeneity.

Table 2.4 presents the estimated versions of β_1 and β_2 for the two political outcomes AfD and $Merkel$ in the first and third column. The coefficients are calculated with the generalized method of moments (GMM) to satisfy the weighted moment condition in Eq. (2.3). It is estimated that the refugee crisis increases the likelihood of voting for the AfD by on average about 2.3 percentage points, an effect that is highly statistically significant. The heterogeneity coefficient β_2 is significant at the five percent level. The estimated satisfaction with Merkel decreases by on average 6.9 percentage points due to the refugee crisis, an effect which is statistically distinguishable from 0 at conventional significance levels. Moreover, the heterogeneity coefficient is significant at the five percent level. Hence, the null hypothesis of no heterogeneity or no predictive power regarding it, i.e., $H_0 : \beta_2 = 0$, can be rejected at conventional significance levels for both outcome variables. This is strong evidence that heterogeneity exists and that the Random Forest classifier is informative about it for the variables AfD and $Merkel$.

Sorted Group Average Treatment Effects The sample is divided into three groups based on the estimated CATE function S in the main sample. Group $k = 1$ has the lowest average treatment effect, group $k = 2$ has an intermediate treatment effect and group $k = 3$ has the largest treatment effect. In the case of the binary outcome y which is estimated with the Random Forest classifier, I define the group $k = 1$ as those individuals who, as a consequence of the refugee crisis, are classified to change their mind from $y = 1$ to $y = 0$, I define group $k = 2$ as those who are classified to stick with their opinion although they might become more likely to change it, and I define group $k = 3$ as those who are classified to change their opinion from $y = 0$ to $y = 1$.

In order to estimate the average treatment effect for Group k, which is denoted as $E[s_0(Z)|k]$, the following equation can be estimated according to CDDF:

$$y = \alpha_0 + \alpha_1 B + \sum_{k=1}^{3} \gamma_k (treat - p(Z))1(k) + v, \tag{2.6}$$

$$s.t.\ E[(p(Z)(1 - p(Z))^{-1} vX] = 0, \tag{2.7}$$

where $1(k)$ is a binary indicator for group membership in k and X is a vector containing B and $\sum_{k=1}^{3}(treat - p(Z)) * 1(k)$. CDDF show that the γ_k in Eq. (2.6) identify the GATEs for the respective groups under simple monotonicity restrictions (see Theorem 2.3 in CDDF).

Column 2 of Table 2.4 reports the results from estimating Eq. (2.6) under the moment condition in Eq. (2.7) using GMM with the outcome AfD. For the group with the smallest treatment effect as predicted by S, the average treatment effect is estimated to be about -0.004 and is statistically insignificant. The average effect for the group with the intermediate treatment effect size is estimated to be about 2.2 percentage points, an effect which is highly statistically significant and very similar to the overall average treatment effect in terms of magnitude. The group with the largest predicted effect size is estimated to be about 14.2 percentage points more likely to vote for the AfD, and this surge is statistically significant at the 1 percent level.

The last four rows of Table 2.4 report p-values for different hypothesis tests regarding the GATEs. The first test tests the joint hypothesis $H_0 : \gamma_1 = \gamma_2 = \gamma_3$ against the two-sided alternative using a Wald test. In the case of AfD, the p-value is about 5.5 percent, yielding moderate evidence for treatment effect heterogeneity. To test the hypothesis $H_0 : \gamma_1 = \gamma_2$ against its two-sided alternative, the second test (as all following ones) uses a simple t-test. The hypothesis cannot be rejected in case of AfD, even at very high significance levels. This implies that the estimate for γ_1 is not a precisely estimated zero and cannot be distinguished from the overall average treatment effect. The third test for $H_0 : \gamma_2 = \gamma_3$ can reject the null against the two-sided alternative at the five percent significance level, which implies that heterogeneity of the treatment effect exists at the upper end of the effect distribution.

The group $k = 3$ is affected significantly stronger by the refugee crisis in terms of voting for the AfD than the average person.

In the last column of Table 2.4, the GATE results for the outcome $Merkel$ are presented. Here, the group with the smallest, i.e., most negative predicted treatment effect is estimated to be about 15.4 percentage points less likely to be satisfied with Merkel's work, an effect which is significant at the 1 percent level. The intermediate-size treatment effect group is estimated to be significantly less satisfied with Merkel at the 1 percent level as well, with a point estimate of about 6.4 percentage points. As in the case of AfD, the intermediate group's effect size is very similar to the estimated overall average treatment effect. Finally, the third group's estimated treatment effect is about -0.013, which is statistically insignificant. The joint hypothesis that all three group-specific coefficients are of the same size can be rejected at the five percent significance level, as well as the hypothesis $H_0 : \gamma_1 = \gamma_2$. It is thus clear that treatment effect heterogeneity exists at the lower end of the treatment effect distribution. However, the hypothesis $H_0 : \gamma_2 = \gamma_3$ cannot be rejected at conventional levels, which implies that the average treatment effect of group $k = 3$ is statistically indistinguishable from the overall average treatment effect.

Polarization Interesting conclusions about political polarization can be drawn from the results of the GATEs. Borrowing from the theoretical work on political polarization (see, e.g., Acemoglu et al., 2016; Dixit and Weibull, 2007), I define polarization as a situation where, as a reaction to the same information shock regarding the refugee inflows, part of the population becomes more pro-immigration, and another part becomes more anti-immigration. However, such a reaction is not observable in the data at hand, neither for the propensity to vote AfD nor for the approval of Merkel. Although the coefficient for the least reactive group regarding AfD voting tendency has a negative sign, it is very close to zero. Furthermore, it is measured so imprecisely that it is statistically indistinguishable from the effect on the intermediate reaction group. There is hence no indication for polarization. Even less ambiguous is the effect on the approval of Merkel, where all groups, from the least reactive to the most reactive, are estimated to become less approving of Merkel. Although the impression in the public debate is that the refugee crisis polarizes the political factions, the data at hand hints toward a general swing to the right rather than a political polarization in the population.

Classification Analysis To learn along which characteristics the differently affected groups differ, a classification analysis (CLAN) is carried out as described in CDDF. The reported values give the sample analog to $E[Z|k]$, i.e., the mean of the observed characteristics given a group $k = 1, 2, 3$.

Because the group $k = 1$ in the case of the AfD and the group $k = 3$ in the case of $Merkel$ are statistically indistinguishable from the average, I pool these groups with the intermediate groups $k = 2$ in the CLAN. Table 2.5 shows the results. The first column shows the average characteristics for the group with the average reaction to the crisis regarding the AfD vote share, and the second column shows the average characteristics for the group with the harshest reaction. The third column reports the p-values of the differences in these averages. It is clear that the two groups are very different from each other, where the strongest reaction group is far less female, more likely to be married, less likely to have Abitur or a basic school degree, and more likely to have a middle school degree. Moreover, the strongest reaction group is less likely to be Christian, far more likely to live in East Germany, more likely to have children and more likely to be of the ages 50-69. However, the strongest reaction group is less likely to be of ages above 70.

Note that all these results are median values from 100 randomly drawn sub-samples of the data as advised by CDDF. As a consequence, the sample size of the groups $k = 1, 2, 3$ can differ between the draws. The median values are 10,388.5 observations for groups $k = 1, 2$, and 195.5 observations for group $k = 3$ in case of AfD.

The fourth column shows the averages for the group with the harshest reaction regarding the satisfaction with Merkel, and the fifth column shows the averages for the group with the average reaction for this outcome. Finally, column six denotes the p-values of the differences in average characteristics for the different reaction strengths regarding $Merkel$. It is easy to see the common patterns between these results and the group differences in the case of AfD. Several groups are over-represented in the groups with the strongest reactions to the refugee crisis. One may characterize those with the highest propensity to react strongly as males with a middle school degree who live in East Germany, are non-Christian, and who are between 50 to 69 years old. People of ages above 70 are strongly underrepresented.

2.4.3 Potential Channels

The CLAN results are in line with the predictions of group conflict theory and economic competition on the labor market. Individuals with middle school degrees are over-represented in the harsh reaction groups, which suggests that individuals with intermediate skill are most sensitive to the refugee inflow. This is consistent with the evidence on job polarization, which observes that intermediate skill levels are most likely to be replaced with automated machines and may therefore either do low-skill jobs or high-skill jobs instead of their traditional intermediate skill tasks (see, e.g., Autor and Dorn, 2013). Individuals who experience this kind of hardship on the labor market may therefore be more sensitive to even further increases in competition for their jobs.

Another observation from the CLAN that is consistent with group conflict theory is that unemployed individuals are over-represented in the strong reaction group to show disapproval of Merkel. Unemployed individuals may fear increasing competition for social welfare benefits if more refugees approach Germany who receive benefits as well. Thus, the prediction from group conflict theory that socio-economically vulnerable individuals are more likely to show anti-immigration reaction sentiment in response to the crisis is partly confirmed. Moreover, the age group which is not active on the labor market any more (70+) is strongly under-represented in the strong reaction groups.

The Politbarometer introduced questions directly related to the refugee crisis in October 2015. In particular, about 1,500 respondents were asked in October 2015 whether they believe that crime rates will increase because of the refugees, whether refugees jeopardize societal values and whether spending on refugees comes at the expense of others. The answers to these questions are informative about group conflict, social identity, and the political economy of hatred as channels for the reaction to the refugee crisis. Unfortunately, these questions were not asked prior to the crisis, which prohibits a causal analysis. In order to provide suggestive evidence, Table 2.6 reports the average approval to these statements for all respondents, for all respondents who approve Merkel, and for all respondents who would vote for the AfD. It can be seen that the approval to all three statements is strongly negatively correlated with approval of Merkel, and strongly positively associated with voting for the AfD.

First, the statement that spending on refugees comes at the expense of others directly refers to competition for scarce government resources. As outlined above, competition for scarce resources among groups of different ancestry can be interpreted directly as group conflict. Thus, the strong positive correlation between the approval to this statement and the tendency to vote AfD and to disapprove Merkel suggests that group conflict is a matter of concern for these groups. Second, respondents who affirm that refugees jeopardize societal values do not refer to competition for scarce resources but point towards an other phenomenon. Societal values are part of a societies' social identity. Refugees may be perceived as jeopardizing societal values if their mere existence is perceived as a disturbance in the collective sense of self. This statement is thus closely related to social identity, and the fact that the approval to this statement is relatively strong among those who disapprove Merkel and those willing to vote the AfD may suggest that social identity considerations play a role in the reactions to the refugee crisis. Third, because hate-creating stories about refugees often build on tales about criminal foreigners, the high approval to the statement that refugees increase criminality among those disapproving Merkel and among those voting for AfD is consistent with the predictions of the political economy of hatred partly driving the effects previously estimated.

2.4.4 Robustness

The most important requirement in the difference-in-differences design is that the common trends assumption holds. As discussed above, Figure 2.3 suggests that this is the case in this application. I nonetheless conduct placebo tests to see whether the slight deviations from the common trends in pre-treatment periods are statistically significant. In these tests the treatment period 2015 is excluded and the treatment group indicator is interacted with the year indicators for 2013 and 2014. The estimation results from these regressions are depicted in columns (1), (2), (4) and (5) of Table 2.7. The placebo interactions are close to zero and statistically insignificant. This reassures that the slight deviations from perfect common trends in Figure 2.3 are not significant.

In addition, the estimated equations are linear probability models. To allow for a more flexible functional form, the third and sixth column of Table 2.7 report estimation results from conditional Logit estimations for the two outcomes VC_I and $VC_I(1 - VC_X)$. The presented coefficients are marginal effects when assuming that the fixed effect is zero. The marginal effects are about twice as large as in the simple linear fixed effects model and lead to similar qualitative interpretations. However, the increase in the size of the coefficients may arise because the conditional Logit only uses information from individuals whose outcome variable actually changes over time, a fact that is also reflected in the relatively low numbers of observations for the Logit estimations. Of course this might be a subgroup which is more likely to be affected by the refugee crisis.

Another cause for concern regarding the difference-in-differences estimation may be the Paris terror attacks in November 2015. During the terror attacks, gunmen and suicide bombers killed 130 people in Paris and left hundreds wounded. Since the attacks affected the treatment group during the treatment period but not the control group, the estimation results might reflect reactions to these terror attacks rather than reactions to the refugee crisis. While the SOEP sample does not contain post-Paris attack observations, the terror attacks may well affect the results from the Politbarometer, where a substantial part of the treatment group is interviewed in November and December. The results after excluding all observations from November and December from the sample and repeating the CATE and GATE estimations are shown in Table 2.8. Surprisingly, the average effect on voting AfD is insignificant and close to zero in this sub-sample. It is thus plausible that a major share of the previously estimated 2.3 percentage point increase results from the November attacks in Paris. It is still difficult to attribute this increase solely to the November attacks because the refugee crisis sowed the seeds and provided the public interest for the subject. It is hard to imagine the strong responses to the November attacks in Paris without the refugee crisis in the first place. In addition, the results for the approval of Merkel are robust to excluding those influenced by the November attacks, as presented in Table 2.8 as well. The estimate for the average treatment effect and the GATEs resemble the full-sample estimates from Table 2.4 very closely. Hence, the observation that the effect for voting AfD diminishes when excluding the post-November observations while attitudes toward immigrants and satisfaction with Merkel are already affected before November suggests that a combined analysis of the interplay between the refugee crisis and the November terror attacks might be enlightening. Perhaps the attitudes are shaped by the crisis, but eventually voting a right-wing party necessitates the extra impulse provided by the November attacks in Paris. Future research might also be interested in analyzing whether the public communication strategies of right-wing populists differed during the refugee crisis and after the November attacks. Table 2.9 shows that the classification analysis is robust for the outcome $Merkel$ when excluding November and December 2015.[11]

2.5 Concluding Remarks

A difference-in-differences design is applied to estimate the causal effect of the so-called refugee crisis on attitudes toward immigration in Germany. Using data from the German Socio-Economic Panel, I find that the refugee crisis increased the likelihood to be very concerned about immigration by about 21 percentage points of which at least 33 percent can be unambiguously attributed to anti-immigrant sentiment. Moreover, analyzing a second data source, the Politbarometer, shows that the refugee crisis increased the likelihood to vote AfD by about two percentage points and lowered the likelihood to approve Merkel by about seven percentage points.

These effects of the crisis are heterogeneous along socio-demographic characteristics, but no socio-demographic group can be identified for which the likelihood to

[11] A robustness check on the classification analysis with outcome AfD is not provided since the CATE analysis did not show heterogeneity in the effect.

vote AfD decreases or the likelihood to show approval of Merkel increases. Thus, regarding the outcomes at hand, no evidence for political polarization as a consequence of the refugee crisis is found, but rather evidence for a swing to the right in the whole population.

The high inflow of refugees and migrants in 2015 and its strong representation in the media led to important attitudinal reactions toward immigration. I contribute to the public debate by showing that the German public became more anti-immigration and more willing to vote AfD as a consequence of the refugee crisis. In addition, the findings that labor market concerns and concerns regarding crime channel the increase in anti-immigration attitudes have important policy implications. For instance, a better communication of up-to-date scientific evidence about the effects of immigration on the labor market (Dustmann et al., 2016) and improving labor market programs would seem appropriate to counter the anti-immigration switch in attitudes due to the crisis. However, it should be kept in mind that other reasons for the switch in attitudes are equally plausible, such as social identity concerns and the political economy of hatred.

Because of the data constraints and the empirical design, I can only estimate the short-term effects of the crisis on attitudes toward immigration. Additional analyses are necessary to answer the question whether these effects were permanent or vanishing over time. Future research may help to gain further insights in this regard as well as regarding the interplay between the so-called refugee crisis and the 2015 terror attacks in Paris in determining voting behavior.

Bibliography

Abadie, A., Athey, S., Imbens, G., and Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering? *arXiv:1710.02926 [econ, math, stat]*. arXiv: 1710.02926.

Acemoglu, D., Chernozhukov, V., and Yildiz, M. (2016). Fragility of asymptotic agreement under Bayesian learning. *Theoretical Economics*, 11(1):187–225.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and Identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Akgündüz, Y. E. and Torun, H. (2018). Two and a half million Syrian refugees, skill mix and capital intensity. Working Paper 186, GLO Discussion Paper.

Athey, S. (2018). The Impact of Machine Learning on Economics. In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Autor, D. H. and Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5):1553–1597.

BAMF (2016). Bundesamt für Migration und Flüchtlinge - Publications - The 2015 Migration Report. Technical report.

Blume, G., Brost, M., Hildebrandt, T., Hock, A., Klormann, S., Köckritz, A., Krupa, M., Lau, M., Randow, G. v., Theile, M., Thumann, M., and Wefing, H. (2016). Grenzöffnung für Flüchtlinge: Was geschah wirklich? *Die Zeit*. 35/2016.

Borjas, G. J. (2003). The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. *The Quarterly Journal of Economics*, 118(4):1335–1374.

Bundeskriminalamt (2016). Kriminalität im Kontext von Zuwanderung - Bundeslagebild Kriminalität im Kontext von Zuwanderung 2015. Technical report.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2018). Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments. Working Paper 24678, National Bureau of Economic Research.

Dammann, D. (2020). Through Time and Time: An Unconventional Difference-in-Differences Event Study Design. *Working Paper - Chapter 1 of this Dissertation.*

Dixit, A. K. and Weibull, J. W. (2007). Political polarization. *Proceedings of the National Academy of Sciences*, 104(18):7351–7356.

Dustmann, C., Schönberg, U., and Stuhler, J. (2016). The Impact of Immigration: Why Do Studies Reach Such Different Results? *Journal of Economic Perspectives*, 30(4):31–56.

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference*, volume 5. Cambridge University Press.

Glaeser, E. L. (2005). The Political Economy of Hatred. *The Quarterly Journal of Economics*, 120(1):45–86.

Hatton, T. J. (2017). Refugees and asylum seekers, the crisis in Europe and the future of policy. *Economic Policy*, 32(91):447–496.

Lancee, B. and Pardos-Prado, S. (2013). Group Conflict Theory in a Longitudinal Perspective: Analyzing the Dynamic Side of Ethnic Competition. *International Migration Review*, 47(1):106–131.

Mayda, A. M. (2006). Who Is Against Immigration? A Cross-Country Investigation of Individual Attitudes toward Immigrants. *The Review of Economics and Statistics*, 88(3):510–530.

Meuleman, B., Davidov, E., and Billiet, J. (2009). Changing attitudes toward immigration in Europe, 2002–2007: A dynamic group conflict theory approach. *Social Science Research*, 38(2):352–365.

Mocan, N. and Raschke, C. (2016). Economic well-being and anti-Semitic, xenophobic, and racist attitudes in Germany. *European Journal of Law and Economics*, 41(1):1–63.

Sherif, M. and Sherif, C. W. (1953). *Groups in harmony and tension. An integration of studies on intergroup relations*. Harper & Brothers, New York.

Sola, A. (2018). The 2015 Refugee Crisis in Germany: Concerns About Immigration and Populism. SSRN Scholarly Paper ID 3169243, Social Science Research Network, Rochester, NY.

Spiess, M. and Kroh, M. (2007). Documentation of the Dataset DESIGN of the Socio-Economic Panel Study (SOEP). Technical report, DIW Berlin.

Tumen, S. (2016). The Economic Impact of Syrian Refugees on Host Countries: Quasi-experimental Evidence from Turkey. *American Economic Review*, 106(5):456–460.

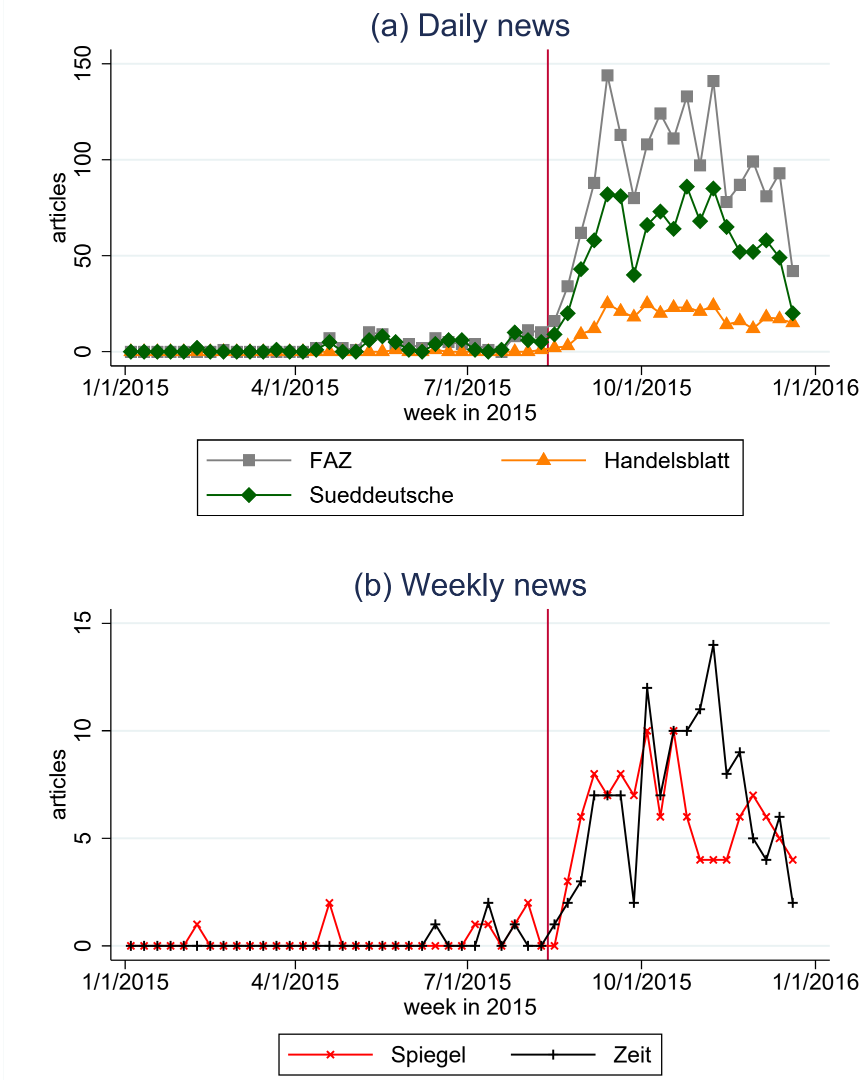
UNHCR (2020). UNHCR Statistics - The World in Numbers. From popstats.unhcr.org. Last accessed May 5, 2020.

Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch - Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127(1):139–169.

2.6 Appendix

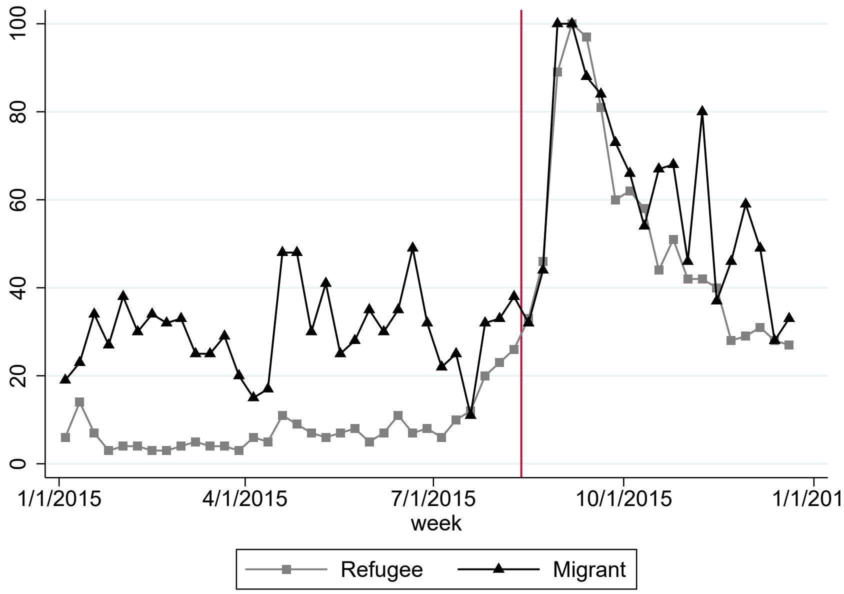
2.6.1 Figures

Figure 2.1: Newspaper Articles Mentioning the Refugee Crisis



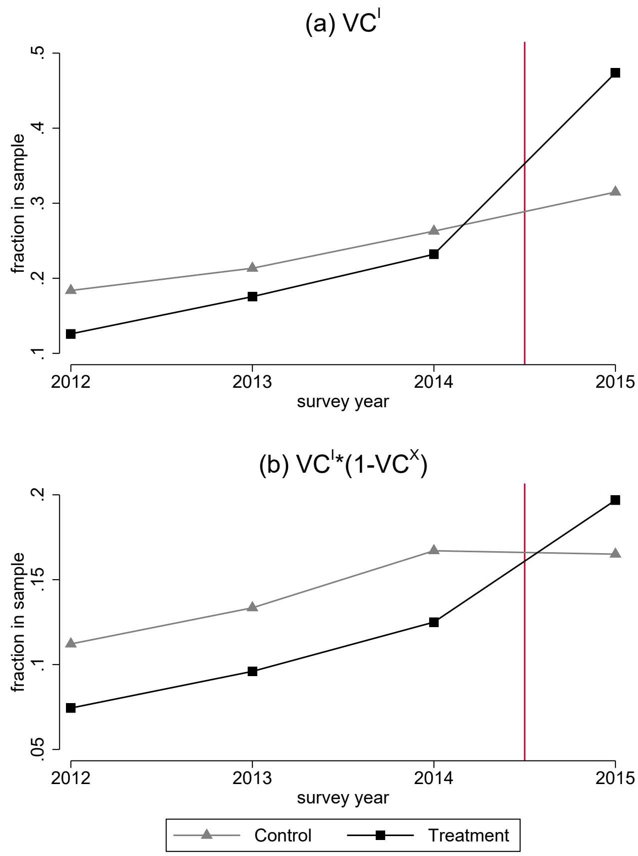
Source: F.A.Z.-Bibliotheksportal, Süddeutsche Zeitung Archiv, www.wiso-net.de. Own depiction.
Notes: Weekly data for German newspapers in 2015. The vertical axis depicts the number of articles mentioning the term 'refugee crisis'. The horizontal axis depicts the week in 2015. The red vertical lines are placed slightly before the week starting on August 16, 2015.

Figure 2.2: Trends in Google Searches Related to the Refugee Crisis in 2015



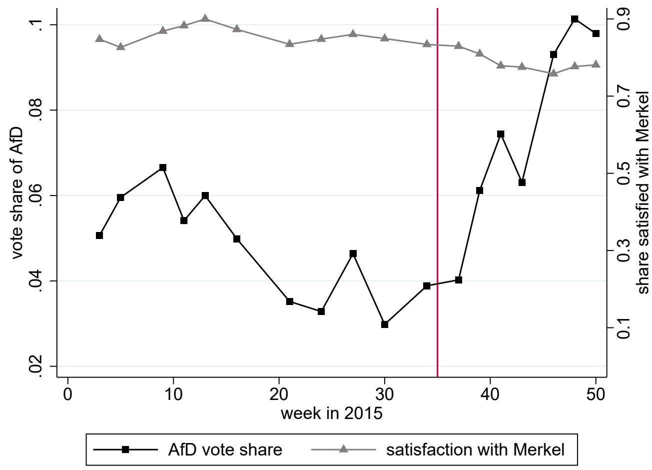
Source: Google Trends, weekly data for Germany in 2015. Own depiction. *Notes:* The trend index represents search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means that there was not enough data for this term (see https://trends.google.com). The red vertical line is placed slightly before the week starting on August 30, 2015.

Figure 2.3: Trends in Attitudes Toward Immigration by Treatment Status



Source: SOEP v32, own calculations. *Notes:* The treatment group consists of individuals who are interviewed during or after September in 2015, the control group consists of individuals who are interviewed before August in 2015. The red vertical line separates pre-treatment from treatment periods.

Figure 2.4: Trends in AfD Poll Share and Satisfaction with Merkel



Source: Politbarometer 2015, own calculations. *Notes:* The red vertical line marks the week before August 30, 2015.

2.6.2 Tables

Table 2.1: Descriptive Statistics of the Attitudes Sample

	2012-2014		2015		
	Control (C)	Treatment (T)	Control	Treatment	$\Delta_C - \Delta_T$
VC^I	0.219	0.178	0.312	0.474	0.204***
$VC^I(1 - VC^X)$	0.134	0.107	0.162	0.203	0.068**
Female	0.533	0.509	0.534	0.504	-0.006
Age	53.195	47.008	54.913	48.940	0.213
East German	0.246	0.274	0.244	0.278	0.006
N	44,894	383	15,239	133	-

Source: SOEP v32, own calculations. *Notes:* Standard errors are clustered at the region of primary sampling (1,884 clusters). Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level.

Table 2.2: Main Results for Attitudes Toward Immigration

	VC^I		$VC^I(1-VC^X)$	
$treat * year2015$	0.205***	0.221***	0.067**	0.108**
	(0.038)	(0.036)	(0.034)	(0.055)
Individual FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Population Weights		✓		✓
R^2	0.014	0.014	0.004	0.004
N	60,649	60,649	60,649	60,649

Source: SOEP v32, own calculations. *Notes:* The two columns show the results from estimating Eq. (2.1) for the two outcome variables VC^I and $VC^I(1-VC^X)$. Robust standard errors with clustering at the level of primary sampling region (1,884 clusters) are reported in parentheses. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level.

Table 2.3: Descriptive Statistics of the Political Outcomes

	Control (C)	Treatment (T)	Δ(C−T)
AfD	0.049	0.072	0.023***
Merkel	0.859	0.793	-0.066***
Female	0.460	0.462	0.002
Married	0.623	0.631	0.008
Parenting	0.809	0.809	0.001
Basic school	0.181	0.183	0.002
Middle School	0.348	0.352	0.004
Abitur	0.468	0.462	-0.006
Unemployed	0.021	0.018	-0.003
Christian	0.563	0.576	0.013*
East German	0.390	0.386	-0.004
Age groups:			
18-20	0.002	0.003	0.001
21-24	0.007	0.007	-0.000
25-29	0.019	0.016	-0.003
30-34	0.036	0.032	-0.004
35-39	0.055	0.053	-0.002
40-44	0.079	0.068	-0.011***
45-49	0.114	0.111	-0.002
50-59	0.239	0.243	0.004
60-69	0.219	0.229	0.010*
70+	0.230	0.238	0.008
N	11,843	9,225	-

Source: Politbarometer 2015, own calculations. *Notes:* The entire sample without sample splits is used for calculations. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level.

Table 2.4: Best Linear Predictions of CATE and GATE

	AfD		*Merkel*	
	CATE	GATE	CATE	GATE
ATE (β_1)	0.023***		-0.069***	
	[0.000]		[0.000]	
HET (β_2)	0.082**		0.087**	
	[0.048]		[0.019]	
γ_1		-0.004		-0.154***
		[0.999]		[0.000]
γ_2		0.022***		-0.064***
		[0.000]		[0.000]
γ_3		0.142***		-0.013
		[0.009]		[0.999]
$p(H_0)$:				
$p(\gamma_1 = \gamma_2 = \gamma_3)$		[0.055]		[0.038]
$p(\gamma_1 = \gamma_2)$		[0.832]		[0.020]
$p(\gamma_2 = \gamma_3)$		[0.038]		[0.944]
$p(\gamma_1 = \gamma_3)$		[0.093]		[0.233]
N	10,534		10,534	

Source: Politbarometer 2015, own calculations. *Notes:* The results stem from 100 randomly drawn main samples, each with 10,534 observations. Median values over these draws are reported. The outcomes are binary indicators for voting AfD, and for approving Merkel. Prediction is based on Random Forest classifiers. CATE refers to estimating Eq. (2.2), and GATE to Eq. (2.6). GMM is applied to satisfy the moment conditions in Eq. (2.3) and Eq. (2.7), respectively. Adjusted and heteroskedasticity-robust p-values are reported in brackets. The last four rows refer to p-values of the null hypotheses stated in the parentheses of $p(H_0)$ against the two-sided alternatives. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level.

Table 2.5: Classification Analysis Political Outcomes

	AfD			*Merkel*		
	$k = 1, 2$	$k = 3$	$p(\Delta)$	$k = 1$	$k = 2, 3$	$p(\Delta)$
Female	0.467	0.165	0.000***	0.128	0.482	0.000***
Married	0.624	0.743	0.002***	0.556	0.631	0.000***
Abitur	0.468	0.346	0.002***	0.416	0.469	0.014**
Middle school	0.345	0.552	0.000***	0.393	0.348	0.019**
Basic school	0.183	0.096	0.003***	0.185	0.182	0.217
Christian	0.573	0.378	0.000***	0.241	0.587	0.000***
East German	0.381	0.732	0.000***	0.539	0.379	0.000***
Unemployed	0.019	0.025	0.864	0.055	0.017	0.000***
Parenting	0.807	0.892	0.006***	0.745	0.813	0.000***
Age groups:						
18-20	0.002	0.000	0.955	0.002	0.002	0.622
21-24	0.007	0.000	0.503	0.014	0.006	0.034**
25-29	0.017	0.012	0.945	0.070	0.014	0.000***
30-34	0.035	0.012	0.135	0.037	0.034	0.156
35-39	0.055	0.052	0.572	0.082	0.053	0.007***
40-44	0.074	0.062	0.429	0.066	0.074	0.313
45-49	0.112	0.102	0.462	0.126	0.112	0.101
50-59	0.238	0.352	0.001***	0.276	0.239	0.024**
60-69	0.222	0.288	0.050**	0.263	0.221	0.037**
70+	0.236	0.120	0.000***	0.042	0.244	0.000***
N	10,338.5	195.5	-	598.5	9,935.5	-

Source: Politbarometer 2015, own calculations. *Notes:* The results stem from 100 randomly drawn main samples, each with 10,534 observations. Prediction is based on Random Forest classifiers. Reported values are the medians of group averages per draw, $p(\Delta)$ refers to the p-values of the differences between the groups. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level.

Table 2.6: Approval to Refugee-Related Statements

Refugees...	... increase criminality	... endanger societal values	... spending at expense of others
	(a) All respondents		
Approval rate	0.635	0.351	0.733
	(b) Respondents approving Merkel		
Approval rate	0.576	0.279	0.687
Difference to all	-0.059***	-0.072***	-0.046***
	(c) Respondents voting for AfD		
Approval rate	0.943	0.828	0.943
Difference to all	0.307***	0.477***	0.209***

Source: Politbarometer 2015, own calculations. *Notes:* N=1,499 observations are included. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level.

Table 2.7: Robustness of Main Results for Attitudes Toward Immigration

	VC_I			$VC_I(1-VC_X)$		
	(1)	(2)	(3)	(4)	(5)	(6)
$treat \times year2015$	-	-	0.400***			0.162**
			(0.094)			(0.081)
$treat \times year2014$	0.004	-	-	-0.008	-	-
	(0.031)			(0.034)		
$treat \times year2013$	-	0.009		-	-0.003	-
		(0.032)			(0.032)	
Excluding 2015	✓	✓		✓	✓	
Conditional Logit			✓			✓
N	45,277	45,277	22,176	45,277	45,277	16,839

Source: SOEP v32, own calculations. *Notes:* Each column refers to separate regression. All regressions include year-fixed effects and an indicator for treatment (or placebo-treatment) group. Robust standard errors with clustering at the level of primary sampling region (1,884 clusters) are reported in parentheses. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level.

Table 2.8: Robustness of Best Linear Predictions of CATE and GATE When Excluding November and December

	AfD		*Merkel*	
	CATE	GATE	CATE	GATE
ATE (β_1)	0.009		-0.056***	
	[0.205]		[0.000]	
HET (β_2)	0.062		0.084*	
	[0.368]		[0.089]	
γ_1		-0.028		-0.151***
		[0.948]		[0.002]
γ_2		0.008		-0.052***
		[0.261]		[0.000]
γ_3		0.092		-0.018
		[0.364]		[1.000]
Exclude November & December (Paris Attacks)	✓	✓	✓	✓
N	8,774		8,774	

Source: Politbarometer 2015, own calculations. *Notes:* Reported results stem from 100 randomly drawn main samples, each with 8,774 observations. Median values over these draws are reported. The outcome variables are binary indicators equal to one if a person would vote for the AfD next Sunday, or when a person agrees that Merkel does a good job, respectively. Prediction is based on Random Forest classifiers. CATE refers to estimating Eq. (2.2), and GATE to Eq. (2.6). GMM is applied to satisfy the moment conditions in Eq. (2.3) and Eq. (2.7), respectively. Adjusted and heteroskedasticity-robust p-values are reported in brackets, see Chernozhukov et al. (2018) for the adjustment.

Table 2.9: Robustness of Classification Analysis When Excluding November and December

	Merkel		
	$k = 1$	$k = 2, 3$	$p(\Delta)$
Female	0.108	0.476	0.000***
Married	0.485	0.629	0.000***
Abitur	0.467	0.465	0.127
Middle school	0.340	0.350	0.228
Basic school	0.186	0.182	0.366
Christian	0.224	0.579	0.000***
East German	0.470	0.382	0.002***
Unemployed	0.051	0.018	0.000***
Parenting	0.647	0.814	0.000***
Age groups:			
18-20	0.000	0.002	0.840
21-24	0.025	0.006	0.001***
25-29	0.063	0.016	0.000***
30-34	0.040	0.035	0.374
35-39	0.099	0.053	0.001***
40-44	0.060	0.077	0.336
45-49	0.130	0.113	0.225
50-59	0.258	0.237	0.242
60-69	0.256	0.220	0.048**
70+	0.036	0.241	0.000****
N	304	8,470	-

Source: Politbarometer 2015, own calculations. *Notes:* Median group averages are reported from 100 random samples, each with 8,774 observations. November and December 2015 are excluded. Prediction is based on Random Forest classifiers. $p(\Delta)$ refers to the p-values for group differences. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level.

Chapter 3

Adaptive Risk-Taking Behavior and the Crimea Crisis

3.0 Abstract

The paper argues that individuals get used to the riskiness of their environment. Such behavior may be expected in settings with unavoidable risk that cannot be perfectly hedged or diversified. This hypothesis is tested by estimating the effect of the 2014 Crimea annexation, which has been interpreted as an act of aggression against Europe by many and is therefore an event that increased the perceived riskiness of living in Europe. A difference-in-differences approach which controls for time-fixed group differences induced by interview timing is applied to rich German panel data. It is estimated that the annexation raised the willingness to take risk significantly.[1] (JEL D81, D83, H12)

[1] All calculations and figures in this chapter are created with the software Stata 14.

3.1 Introduction

Both psychologists and economists show that individuals adapt to changing conditions and realize stable levels of life satisfaction. This observation has been called hedonic treadmill in the literature (see, e.g., Brickman and Campbell, 1971; Easterlin, 1974, 1995; Diener and Diener, 1996; Riis et al., 2005; Clark et al., 2008). Even individuals who experience severe events like widowhood or childbirth return, on average and after some time, to the same level of subjective well-being as before these events. This paper seeks to translate this adaptive behavior to situations where individuals are exposed to a higher environment risk.

As a reaction to higher environment risk, individuals are assumed to adapt their willingness to take risk and realize the same or a similar level of satisfaction as before the increase.[2] This is particularly expected for risks that cannot be avoided, hedged or diversified. Otherwise, individuals may prefer to protect themselves from the increased risk. Protective reactions have been documented as a reaction to financial crises and financial losses (Dohmen et al., 2016; Guiso et al., 2018; Cohn et al., 2015; Imas, 2016). The idea of an unavoidable risk may seem odd at first, but it is a decent representation of many real-life settings. For instance, individuals may find it hard to avoid, hedge or diversify risks such as war, nuclear catastrophes or violent prosecution. The adaption hypothesis assumes that individuals realize constant levels of satisfaction no matter the changes of riskiness they face and adapt their willingness

[2]Whether or not adaption can take place most likely depends on individuals' psychological traits like resilience. This paper, however, focuses on the possible occurrence of a hardening effect in risk taking behavior, leaving the interaction with psychological factors such as resilience for future research.

to take risk respectively.[3]

A sensible possibility to analyze an isolated shock to the riskiness of the environment is the 2014 annexation of Crimea by Russia from the perspective of German citizens. The annexation increased the perceived riskiness of living in Germany without changing actual outcomes for the vast majority of citizens. An event study approach as suggested in Dammann (2020) is applied to the German Socio-Economic Panel (SOEP). Participants who were interviewed before the Crimea crisis in 2014 constitute the control group, and participants who were interviewed during or after the crisis are the treatment group. The risk attitudes of the two groups develop parallel prior to 2014 and diverge only in 2014, where the treatment group shows a significantly higher willingness to take risk. These results are robust to placebo tests, different functional forms and sample restrictions. In conclusion, the results are consistent with the hypothesis of adaptive risk taking behavior.

This paper contributes to the literature on subjective well-being by applying the idea of the hedonic treadmill or adaptive utility to risk-taking behavior. Moreover, it seeks to contribute to the understanding of the origins, the formation, and the malleability of the willingness to take risk. To the best of my knowledge, this is the first paper that analyses the effect of the Russian annexation of the Crimea and that provides a causal framework to analyze individual-level effects of this annexation.

Section 2 of this paper outlines theoretical considerations regarding adaptive risk-taking behavior. Based on these considerations, empirical predictions are deduced.

[3]Adaption depends on the kind of change an individual is exposed to. For instance, humans cannot adapt to chronic pain (Kahneman et al., 1997). It is assumed that humans are able to adapt to the risks induced by the Crimea crisis analyzed below.

Section 3 discusses related literature. The specific historical event of the Crimea annexation is discussed in Section 4, and Section 5 presents the empirical strategy and data used to test whether the hypothesized behavior can be observed in survey data. Section 6 discusses the results and Section 7 provides a discussion on the robustness of the estimations. Section 8 concludes on the analysis. All figures and tables are appended in Section 9.

3.2 Theoretical Considerations

Suppose an individual's preferences over lotteries are such that she prefers higher payoffs over lower ones. Moreover, suppose the individual is risk averse and, *ceteris paribus*, prefers less risky lotteries over more risky ones.[4]

Suppose the individual participates in a lottery and learns that its riskiness increased. She may thus reconsider her evaluation of the lottery. Obvious reactions are exiting the lottery if it becomes too risky to yield a positive net present value, or hedging or diversifying the risk. I refer to such reactions as protective reactions. However, there are lotteries that cannot be avoided, hedged or diversified. One example is whether a foreign power starts a war against one's country of residence. Assume that protective reactions are not part of the option set. If the individual does not react to the increase, she will become less satisfied with the lottery due to her risk aversion. But this seems to be at odds with the observation of the hedonic treadmill that most humans are satisfied with their life irrespective of a wide range of factors including severe illness and the death of close relatives (Brickman and

[4]See, e.g., Pratt (1964) elaborating on the coefficient of relative risk aversion.

Campbell, 1971; Easterlin, 1974, 1995; Diener and Diener, 1996; Riis et al., 2005; Clark et al., 2008).

There is a third way the individual could deal with the situation, and that is by adaption. Instead of despairing of unavoidable risk, most humans may learn to live well with it and report constant levels of life satisfaction. If the individual learns to accept the higher level of risk surrounding her, she may realize a reference level of utility that she derives from being alive. This paper states the hypothesis that adaption in this case can lead individuals to become more willing to take risk.

Kahneman et al. (1997) discuss that observed life satisfaction or subjective well-being is an experienced utility that may differ from the decision utility which economic theory is usually concerned with. The hedonic treadmill literature relates to experienced utility, which is reported to adapt to different circumstances. This paper tests the hypothesis that utility adaption to new situations changes parameters of the utility function that may as well enter the decision utility, i.e., that it changes the willingness to take risk. However, it has been common in the literature to equate experience and decision utility in practical applications (see, e.g., Clark and Oswald, 1996; Kahneman and Sugden, 2005; Boyce and Wood, 2011).

There are important limitations to the perfect adaption hypothesis. First, adaption may take some time and does not necessarily take place immediately, which results in dynamics that remain unexplained by the hypothesis outlined in this paper. Second, adaption may be imperfect, since different life circumstances may eventually lead to different levels of utility realization, although mean-reverting behavior may still result in adaptive risk taking to some degree. Third, the reference level of utility or satisfaction may depend on individual characteristics and change over time. Moreover, the psychological and evolutionary mechanisms that might explain such adaptive behavior are beyond the scope of this paper.

3.3 Related Literature

Only few empirical studies analyze time variation in risk attitudes at the individual level. Early experimental studies on this topic find that risk preferences are rather unstable in repeated sessions with the same individuals (see, for instance, Wehrung et al., 1984; Smidts, 1997). In contrast, Sahm (2012) estimates a correlated random effects model to analyze how much risk tolerance changes within individuals over time. She finds that after filtering out the noise in the data, most of the explainable variation in risk tolerance is persistent within individuals. However, she only considers the variation in risk tolerance that she can explain with individual-level variables in this exercise, which makes her omit factors such as environmental risks. Thus, the part of the variation that she neglects as being unsystematic may contain exactly the kind of variation this paper is interested in. Consistent with this interpretation are the findings of studies that scrutinize the effect of the 2008 world financial crisis on risk attitudes. Dohmen et al. (2016) find that the crisis reduced individuals' willingness to take risk in both Germany and the Ukraine in a comparative study. Likewise, Guiso et al. (2018) find similar effects of the financial crisis in Italy. In line with these arguments, Malmendier and Nagel (2011) show that individuals who experienced low returns on their investments become more averse to take future risks.

Cohn et al. (2015) add to the evidence on time-varying risk attitudes by showing that risk attitudes vary counter-cyclically over the business cycle.

These findings do not necessarily contradict the adaption hypothesis, but may complement it. They can be explained by the above theoretical considerations since financial risk can generally be hedged[5] and diversified and may thus evoke rather protective reactions. Moreover, the financial crisis may be interpreted as the realized outcome of previously taken risks, which results in ex-post risk evaluations of realized losses that may differ from ex-ante evaluations (see Imas, 2016, for a discussion of this effect in financial matters). This suggests that adaptive behavior, if present, may rather be observed during ex-ante evaluations than during ex-post evaluations and as a reaction to risk that cannot be hedged or diversified.

The study at hand seeks to shed light on a kind of environmental change that, to the best of my knowledge, has not yet been considered. Focusing on the ex-ante evaluation of uncontrollable risks, I formulate the possibility of positive adaptive risk taking behavior. To test the adaption hypothesis, I analyze a crisis that arguably constitutes an isolated change to an individuals ex-ante evaluation of the risk of life. For the vast majority of German individuals, the Crimea annexation did not have an actual effect on life outcomes, but did change the perceived risk of living in Europe as it has been interpreted by the public media as a relapse of war on European ground. However, there is little German citizens could do to hedge or diversify this risk. I thus expect adaption of risk taking behavior in Germany in response to the Crimea crisis.

[5] Even systematic market risk that investors usually cannot diversify can be hedged, for instance with a short position in a market index.

3.4 The Crimea Crisis

In November 2013, the Ukrainian government decided to suspend an association agreement with the European Union and to pursue closer ties to Russia (bpb, 2019). This led to protests and riots with hundreds of injured persons and more than 100 fatalities (Marxsen, 2014). These protests received worldwide media attention under the name Maidan. In February 2014, Ukraine was at the threshold of a civil war and its pro-Russian president Yanukovych had to flee the capital, which allowed the pro-European opposition to fill the most important government positions (bpb, 2019). Instantaneously after these events, there were violent clashes between pro-Ukrainian Crimean Tatars and pro-Russian supporters on the Crimea (Porsche-Ludwig, 2014). In the course of these clashes, the Council of the Russian Federation granted president Putin the power to deploy armed forces on the Crimea on March 1, and Russian troops actively engaged in the conflict and gathered at the Ukrainian border during the next weeks (Marxsen, 2014).

A new government was formed for the Autonomous Republic of the Crimea, which denied the pro-European interim government of Ukraine. The Crimean parliament declared its independence from Ukraine on March 11 (Porsche-Ludwig, 2014). In a referendum held on March 16, a clear majority of voters declared themselves in favor of joining the Russian Federation, and the Crimea formally requested accession the very next day. Finally, the application was ratified by the Council of the Russian Federation on March 21 (Porsche-Ludwig, 2014).

In the European Union and the United States, the annexation of the Crimea was interpreted as an act of aggression against the West in popular media and raised the fear of war among Western citizens (Dahlkamp et al., 2014; Strempel, 2014). Thus, I interpret these events as an exogenous shock to the perceived riskiness of living in Europe. However, while the fear of war increased, the annexation of the Crimea did not actually change everyday life, income or any other realized outcomes for the vast majority of European citizens. The Crimea crisis hence provides the opportunity to study a change in the risk of the environment without any changes to outcome realizations at the time of the crisis, and to observe the response with respect to risk attitudes.

Note that other crises that occurred during the observation period are very different in nature. Most of these crises did not change the riskiness of the environment in which people acted, but changed the risky acting of people within the given environment. For instance, the 2008 financial crisis did not represent the arrival of higher risks on the financial markets, but marked the end of a phase of excessive financial risk taking. If you took too much risk in the past, you can change that today, e.g., by hedging strategies. This crisis did hence not necessitate adaption to a new environment, but correction of previous behavior.

I use Google search trends to identify the point in time when the public became aware of the importance of the Crimea crisis. Figure 3.3 depicts the time trends for different Google search terms in Germany on a weekly basis from September 2013 to August 2014. The trends are indexed such that the value 100 marks the highest number of searches during the observation period, whereas the value 50 marks half this number, 33 marks approximately one third of it, and so on. The solid line shows the search index for the combination of the terms Crimea and crisis. This combination was not of interest before March 2014, where it peaked in the first week of March and returned to moderate values during April and May, and searches declined steadily afterwards. From this I infer that the public awareness of the Crimea crisis started at the beginning of March, and I define March 1, 2014 as the beginning of the Crimea crisis. It is enlightening to compare other trending searches with the search for the Crimea crisis. For instance, the search for the term „war" (dash-dotted line) peaked together with the Crimea crisis. Moreover, there is a high correlation between the Crimea crisis and the search for the term „cold war" (dashed line). This indicates that the Crimea annexation was interpreted as an aggression against the West and that it raised the fear of war in Germany and Europe.

3.5 Data and Empirical Strategy

I use the German Socio-Economic Panel (SOEP) for the analysis (see Wagner et al., 2007, for a detailed description). Since 2008, the SOEP includes a question about risk attitudes every year. I therefore use the survey years 2008-2014, which results in 6 pre-treatment periods in addition to the treatment year 2014. The question posed is: "Are you generally a person who is willing to take risks or do you try to avoid taking risks?" I define RA_{it} as a dummy variable that is one if individual i exhibits above median willingness to take risk in period t, and zero otherwise (see Figure 3.1 for details). Only those individuals for whom RA_{it} is available in all years are included in the analysis.

Analogously to the risk attitude measure, I define the variable $LSAT_{it}$ as taking on the value one if individual i reports an above median life satisfaction in period t (see Figure 3.2 for details). This measure is used to test whether the life satisfaction stays constant throughout the crisis, as would be predicted by the hedonic treadmill.

In the 2014 wave, those participants interviewed during or after March are defined as treated by the Crimea annexation, whereas those who are interviewed earlier in 2014 are defined as the control group. The variable *treat* is a binary indicator for group affiliation, where one indicates the treatment group and zero indicates the control group. Comparing the simple difference between treatment and control in the year 2014 may yield biased results because the two groups may differ from each other. Individuals who are interviewed before March in 2014 may be, either by chance or because of the SOEP sample design, systematically different from individuals who are interviewed later. I thus apply a difference-in-differences design to control for any level differences between these groups.

The SOEP encompasses person identifiers that allow me to observe the very same individuals as in 2014 in the prior years.[6] The study design is special because it contains a difference-in-differences design over two time dimensions as discussed in Dammann (2020). The first difference is taken between the treatment and control group in 2014, the second difference is taken between the two groups prior to 2014. The identifying assumption is that the treatment group's risk attitude would have developed parallel to the control group's risk attitude without the Crimea annexation. Moreover, I need to assume that the only change that could cause any deviation from

[6]Note that although there is a positive correlation between survey months in different survey years, individuals who are interviewed in, for instance, March 2014 may well be interviewed in February or April, or any other month in prior years.

the parallel trends is the Crimea annexation.

Dohmen et al. (2011) argue that the key exogenous determinants of risk attitudes are gender, age, parental education and height. I conduct individual fixed effects regressions to rule out that gender, parental education, or height confound the results. To apply the difference-in-differences idea, I include survey year-fixed effects.[7] The estimation equation is

$$RA_{it} = \alpha_i + \gamma_t + \beta treat * year2014_{it} + \delta X_{it} + u_{it}, \tag{3.1}$$

where α_i is the individual fixed effect, γ_t is the year fixed effect, and X_{it} is a vector of time-varying individual characteristics such as location, marital status, and seasonality effects. The variable $treat * year2014_{it}$ interacts the treatment group identifier with the treatment year identifier, and the treatment effect of interest is measured by β under the parallel trends assumption.

A helpful guide for adjusting standard errors is provided by Abadie et al. (2017). They argue that cluster adjustment is necessary either if the sample is clustered (cluster C has a sampling probability below one) or if the treatment assignment mechanism assigns some clusters with a higher probability of treatment. Since all individuals are treated by the crisis, the assignment mechanism is not clustered. But the sampling strategy of the SOEP is clustered, where first regions C are drawn from a set of regions and then households are randomly drawn from the selected primary sampling regions C (Spiess and Kroh, 2007). Guided by Abadie et al. (2017), I adjust

[7]Note that including both individual-fixed effects and year-fixed effects implicitly controls for age.

standard errors in the SOEP for clustering at the region of primary sampling.

3.6 Results

Figure 3.4 plots the trends of the variable RA_{it} for the two treatment status groups over the entire observation period. The two groups develop reasonably parallel between 2008 and 2013, and eventually diverge in 2014, where the treatment group shows a more pronounced increase in the willingness to take risk than the control group. The first row of Table 3.1 shows that RA_{it} is about 28.1 percent for the control group during the pre-treatment period (column 1), and that it is about 30.4 percent for the treatment group during the same period of time (column 2).[8] Thus, the pre-treatment difference between the treatment and the control group is about 2.3 percentage points. During the treatment period 2014, those who were interviewed before the Crimea conflict have a 32.2 percent chance of showing an above-median willingness to take risk (column 3), whereas those who were interviewed during or after the crisis show a likelihood of 36.6 percent (column 4), yielding a difference of 4.4 percentage points in the treatment period. Assuming that the pre-treatment difference between the two groups indicates a stable difference between the two potential outcomes of RA_{it} and that the only difference between the treatment and the pre-treatment period is the appearance of the Crimea conflict, one can subtract the differences from each other, yielding a causal estimate of a 2.1 percentage point increase. As indicated by the asterisks, the causal estimate is statistically significant

[8]Note that the highest weight of the empirical distribution, i.e., about 20 percent, lies exactly on the median in this application.

at the five percent level even in this very descriptive setting without controlling for further factors.

The results from estimating Eq. (3.1) with the willingness to take risk as the dependent variable are depicted in Table 3.2. The first column reports the most basic specification, which only includes individual and year fixed effects and a treatment group-treatment period interaction. The estimated effect amounts to a 2.2 percentage points higher likelihood of having a high willingness to take risk due to the Crimea crisis, which is statistically significant at the 1 percent level. This estimate remains stable when including marital status and geographical location as control variables in the second column. To check whether seasonality plays a role in the estimation, the third column includes month fixed effects. The point estimate drops by 0.4 percentage points, which results in a statistically significantly estimated increase of 1.8 percentage points in the likelihood of having a high willingness to take risk.

Consider second the estimated effect of the crisis on the likelihood of reporting a high life satisfaction. The second row in Table 3.1 shows the descriptive evidence. The likelihood to show an above-median life satisfaction is about 1.5 percentage points lower for the treatment group than for the control group during the pre-treatment periods. During the treatment period, it is about 1.4 percentage points lower for the treatment group. Making the analogous assumption as in the case of the risk attitude and calculating the difference-in-differences estimate results in a 0.1 percentage point effect of the crisis on life satisfaction, an economically negligible effect that is statistically insignificant. This finding is confirmed in the fixed effects regression shown in Table 3.3. The estimated effect of the crisis is insignificant and close to zero for all specifications, the plain fixed effects specification as well as the one including additional time-varying characteristics and seasonality effects.

This evidence is clearly consistent with the adaption hypothesis. The crisis did not change overall life satisfaction but led to an adaption in the willingness to take risk, which is what the adaption hypothesis predicts.

3.7 Robustness

The main assumption made in the empirical analysis is that the potential outcomes of the analyzed variables follow a common trend. The lower panel of Figure 3.4 depicts the development of the level differences between the treatment and the control group. The group differences are indicated by the black squares in combination with their 95 percent confidence intervals. The average group difference during the pre-treatment periods is about 0.023, which is indicated by the horizontal red line. The pre-treatment differences from this average can be interpreted as placebo treatments. The fact that the placebo treatments are not statistically significant provides strong evidence in favor of the parallel trends assumption.

Figure 3.5 shows the same graph for the life satisfaction variable. The average difference between the treatment status groups is about -0.015, as is indicated by the horizontal red line. The data points in the single years are slightly more noisy in case of the life satisfaction variable as compared to the risk attitudes, but the two respective groups show no significant deviation from their average difference at all, neither during the pre-treatment nor the treatment periods. This may be interpreted as evidence for the hypothesis that the insignificant effect in 2014 is not a product of random noise, but resembles a stable relation that can be found over the entire observation period.

An additional check of the validity of the parallel trends assumption is to estimate the difference-in-differences estimator for exogenous variables. Under parallel trends, no effect should be found on variables that are not subjected to the treatment. Table 3.1 reports the estimates for the set of control variables discussed above. Reassuringly, the estimated placebo effects on indicators for being female, living in East Germany, and being married are very close to zero and statistically insignificant. However, the treatment correlates significantly with a slight increase in age of slightly less than one month. This slight age increase hardly affects the outcome and is rather an artifact of the design than a failure of balancing. It indicates that the survey timing within the year is not stable over survey waves. More specifically, individuals interviewed in or after march in 2014 are not always interviewed in or after march in the years prior to 2014, but are interviewed, on average, one month earlier. This highlights again the need to check for seasonality effects, which is done in Table 3.2.

Moreover, Table 3.1 shows that the treatment is slightly correlated with having a matriculation standard (Abitur). Although the effect is hardly significant at the 10 percent level and practically small at less than 0.2 percentage points, it may indicate a specialty of the German education system. The final examinations to graduate with an Abitur take place in spring every year, and this might actually lead the treatment to be correlated with a higher fraction of individuals who did graduate. To check whether Abitur graduations drive the treatment effect from the main specification, column (1) of Table 3.4 estimates the main specification for individuals who are 25 or older in 2015, therefore being well above the usual age of Abitur graduation which takes place between ages 17 and 19 for the majority of graduates. The estimated treatment effect remains robust and significant after excluding the younger population.

A further cause for concern may be the linearity of the dependencies imposed upon the estimation of Eq. (3.1). As an alternative to this, columns (2) and (3) of Table 3.4 provide the estimated average partial effects from Logit regressions of the same relationships. The second column reports the result of a simple, unconditional Logit estimation, which results in an estimated average effect of about 1.8 percentage points. This effect is significant at the five percent level and of comparable magnitude as the results from the linear model. The third column shows the result from a conditional Logit estimation to control for individual-fixed effects. The estimated effect is about 3.8 percentage points and statistically significant at the 5 percent level.

Column (4) of Table 3.4 shows the result from the fixed-effects regression without requiring the sample to be balanced on the outcome. The resulting coefficient amounts to an increase of about 1.6 percentage points, which is statistically significant at the 1 percent level.

It is important for the validity of the difference-in-differences design that the changes observed in the data are actually due to the Crimea annexation and not for some other, unrelated reasons. It may appear more plausible to interpret the observed changes as a consequence of the crisis when the data is restricted only to observations closely around the point in time that is defined as the start of the crisis. Column (5) of Table 3.4 reports the result of restricting the sample to individuals who were interviewed in the first four months in the 2014 survey wave. The estimated coefficient amounts to about 2.1 percentage points and is statistically significant. Furthermore, column (6) reports the result of restricting the observations to individuals who were interviewed in February or March during the 2014 wave, which yields a very similar and significant estimate of 1.8 percentage points. Thus, restricting the data to observations for which other influences than the treatment under scrutiny are unlikely delivers results which are very similar to the full-sample estimations.

Table 3.5 shows the same robustness checks for life satisfaction, without any statistically significant result. Hence, the empirical results are robust to various manipulations of the analysis.

3.8 Concluding Remarks

Adaptive utility behavior is well-documented in economics and psychology. Humans adapt to changes both in their individual lives and in their environments, where adaption means that they realize a similar level of life satisfaction as before these changes. Although such behavior is generally well-documented, it has not yet been translated to other settings such as risk-taking behavior. This paper develops the hypothesis of adaptive risk-taking behavior, i.e., the ability of humans to adapt to different levels of risk in their environment or their personal life.

The theoretical considerations encompass two components. First, learning about the riskiness of a lottery changes the ex-ante risk evaluation and, potentially, the risk attitudes. Risk attitudes may therefore stay stable, change in a protective, more conservative manner, or lead to an adaption. Second, the realization of a risky lottery and the accompanying ex-post risk evaluation may influence risk attitudes in a different manner for psychological reasons. This paper formulates the prediction that adaptive behavior is more likely to occur in the case of ex-ante risk evaluation for lotteries that cannot be avoided, hedged or diversified.

The hypothesis and its predictions are tested by analyzing the effect of the Crimea conflict and annexation from 2014 on the attitudes of German participants in the SOEP. To identify the causal effect of the crisis, a panel difference-in-differences strategy is employed that uses both within-time variation during the interviews in 2014 and within-time variation over the interview years. It is found that the crisis increased the willingness to take risk significantly, without altering the overall life satisfaction. This is interpreted as empirical evidence which is consistent with the risk adaption hypothesis.

This paper makes three contributions to the economic literature. First and despite of the ongoing research on utility adaption, this is the first paper that formulates the possibility of adaptive risk taking behavior. Second, the adaption hypothesis is shown to be consistent with the empirical evidence on the Crimea crisis. Third, this is the first paper that analyzes the effect of the Crimea crisis on economic attitudes.

Future literature should conduct experiments on adaptive risk-taking behavior to learn more about the economic consequences of the presented findings. Moreover, future research might formalize the theoretical construct outlined in this paper more formally to understand its consequences in, for instance, interactive games or financial decision making. Last but not least, researchers who conduct studies in this field should be aware of the possibility of adaptive risk-taking behavior.

Bibliography

Abadie, A., Athey, S., Imbens, G., and Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering? *arXiv:1710.02926 [econ, math, stat]*. arXiv: 1710.02926.

Boyce, C. J. and Wood, A. M. (2011). Personality and the marginal utility of income: Personality interacts with increases in household income to determine life satisfaction. *Journal of Economic Behavior & Organization*, 78(1):183–191.

bpb (2019). Vor fünf Jahren: Russlands Annexion der Krim. From http://www.bpb.de/politik/hintergrund-aktuell/287565/krim-annexion. Last accessed June 24, 2019.

Brickman, P. and Campbell, D. T. (1971). Hedonic relativism and planning the good society. In Appley, M. H., editor, *Adaptation level theory: A symposium*, pages 287–302. New York: Academic Press.

Clark, A. E., Diener, E., Georgellis, Y., and Lucas, R. E. (2008). Lags And Leads in Life Satisfaction: a Test of the Baseline Hypothesis*. *The Economic Journal*, 118(529):F222–F243.

Clark, A. E. and Oswald, A. J. (1996). Satisfaction and comparison income. *Journal of Public Economics*, 61(3):359–381.

Cohn, A., Engelmann, J., Fehr, E., and Maréchal, M. A. (2015). Evidence for Countercyclical Risk Aversion: An Experiment with Financial Professionals. *American Economic Review*, 105(2):860–885.

Dahlkamp, J., Gathmann, M., Hujer, M., Kurbjuweit, D., Middelhoff, P., Müller, P., Puhl, J., Schmid, B., and Wensierski, P. (2014). "Wir denken an den August 1914 zurück". *DER SPIEGEL*, (18):18–21.

Dammann, D. (2020). Through Time and Time: An Unconventional Difference-in-Differences Event Study Design. *Working Paper - Chapter 1 of this Dissertation.*

Diener, E. and Diener, C. (1996). Most People Are Happy. *Psychological Science*, 7(3):181–185.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550.

Dohmen, T., Lehmann, H., and Pignatti, N. (2016). Time-varying individual risk attitudes over the Great Recession: A comparison of Germany and Ukraine. *Journal of Comparative Economics*, 44(1):182–200.

Easterlin, R. A. (1974). Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In David, P. A. and Reder, M. W., editors, *Nations and Households in Economic Growth*, pages 89–125. Academic Press.

Easterlin, R. A. (1995). Will raising the incomes of all increase the happiness of all? *Journal of Economic Behavior & Organization*, 27(1):35–47.

Guiso, L., Sapienza, P., and Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics*, 128(3):403–421.

Imas, A. (2016). The Realization Effect: Risk-Taking after Realized versus Paper Losses. *American Economic Review*, 106(8):2086–2109.

Kahneman, D. and Sugden, R. (2005). Experienced Utility as a Standard of Policy Evaluation. *Environmental and Resource Economics*, 32(1):161–181.

Kahneman, D., Wakker, P. P., and Sarin, R. (1997). Back to Bentham? Explorations of Experienced Utility. *The Quarterly Journal of Economics*, 112(2):375–406.

Malmendier, U. and Nagel, S. (2011). Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?*. *The Quarterly Journal of Economics*, 126(1):373–416.

Marxsen, C. (2014). The Crimea Crisis – An International Law Perspective. *Zeitschrift für ausländisches öffentliches Recht und Völkerrecht (Heidelberg Journal of International Law)*, 74(2):367–391.

Porsche-Ludwig, M. (2014). *Krimkrise und Völkerrecht.* Traugott Bautz, Nordhausen.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica*, 32(1/2):122–136.

Riis, J., Loewenstein, G., Baron, J., Jepson, C., Fagerlin, A., and Ubel, P. A. (2005). Ignorance of Hedonic Adaptation to Hemodialysis: A Study Using Ecological Momentary Assessment. *Journal of Experimental Psychology: General*, 134(1):3–9.

Sahm, C. R. (2012). How Much Does Risk Tolerance Change? *The Quarterly Journal of Finance*, 2(4).

Smidts, A. (1997). The Relationship Between Risk Attitude and Strength of Preference: A Test of Intrinsic Risk Attitude. *Management Science*, 43(3):357–370.

Spiess, M. and Kroh, M. (2007). Documentation of the Dataset DESIGN of the Socio-Economic Panel Study (SOEP). Technical report, DIW Berlin.

Strempel, M. (2014). ARD-DeutschlandTrend: Große Sorge vor neuem Kalten Krieg. *tagesschau.de*. Retrieved from https://www.tagesschau.de.

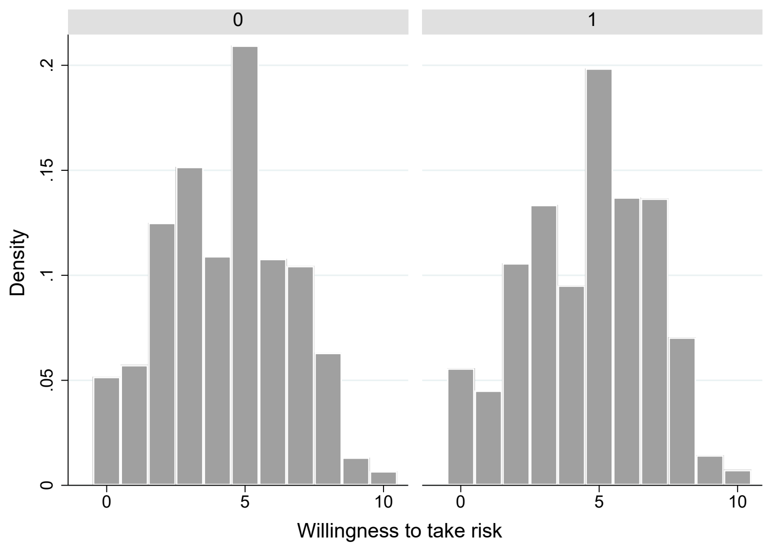
Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch - Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127(1):139–169.

Wehrung, D. A., Maccrimmon, K. R., and Brothers, K. M. (1984). Utility Assessment: Domains, Stability, And Equivalence Procedures. *INFOR: Information Systems and Operational Research*, 22(2):98–115.

3.9 Appendix

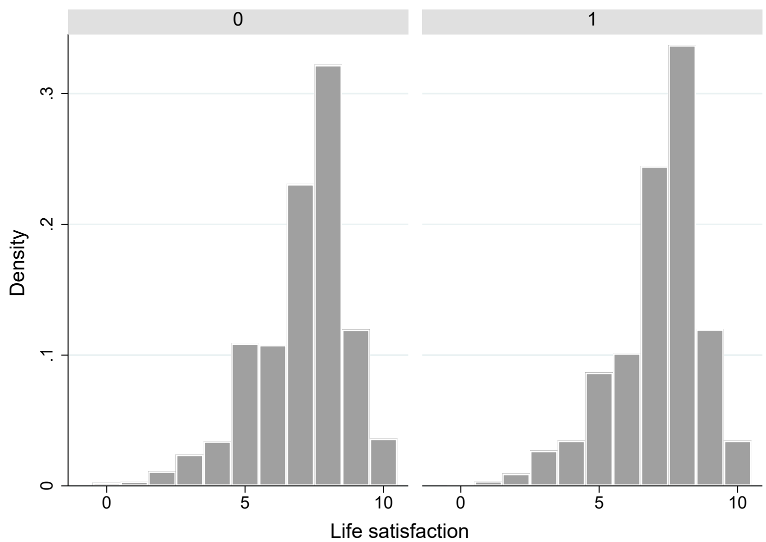
3.9.1 Figures

Figure 3.1: Risk Attitudes by Treatment Status



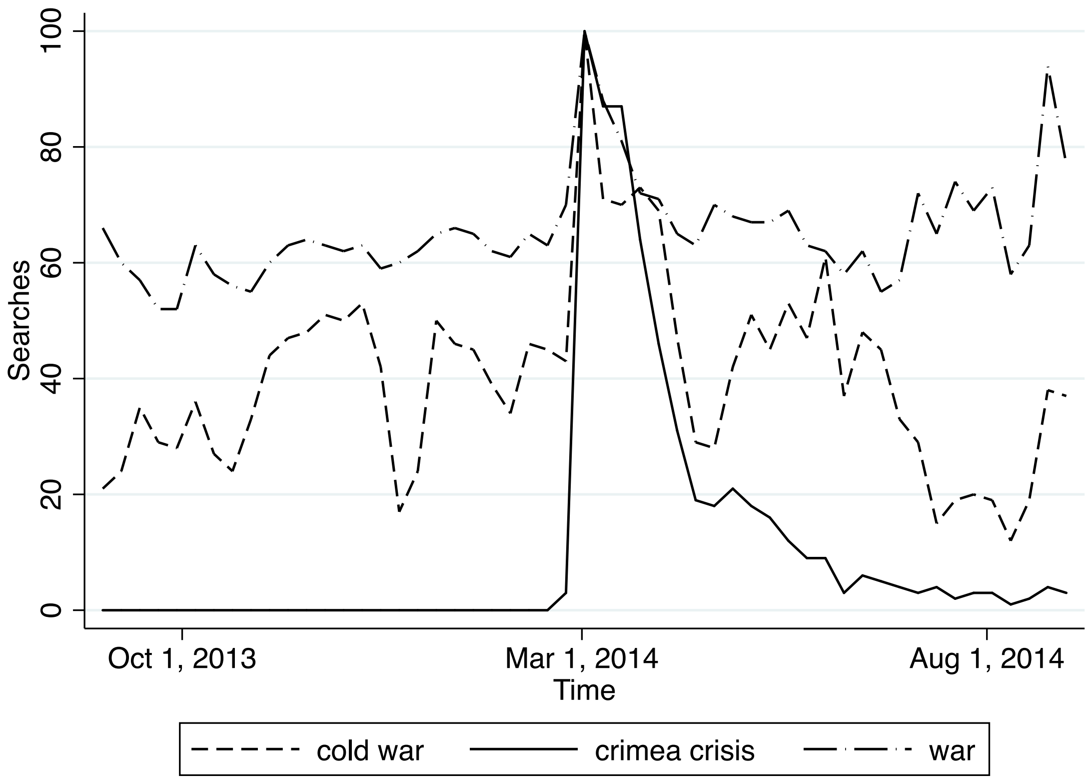
Source: SOEP v33, own calculations. *Notes:* Sample zero includes both treatment and control group in non-treatment periods. Sample one includes treatment group in the treatment period. The median willingness to take risk is five.

Figure 3.2: Life Satisfaction by Treatment Status



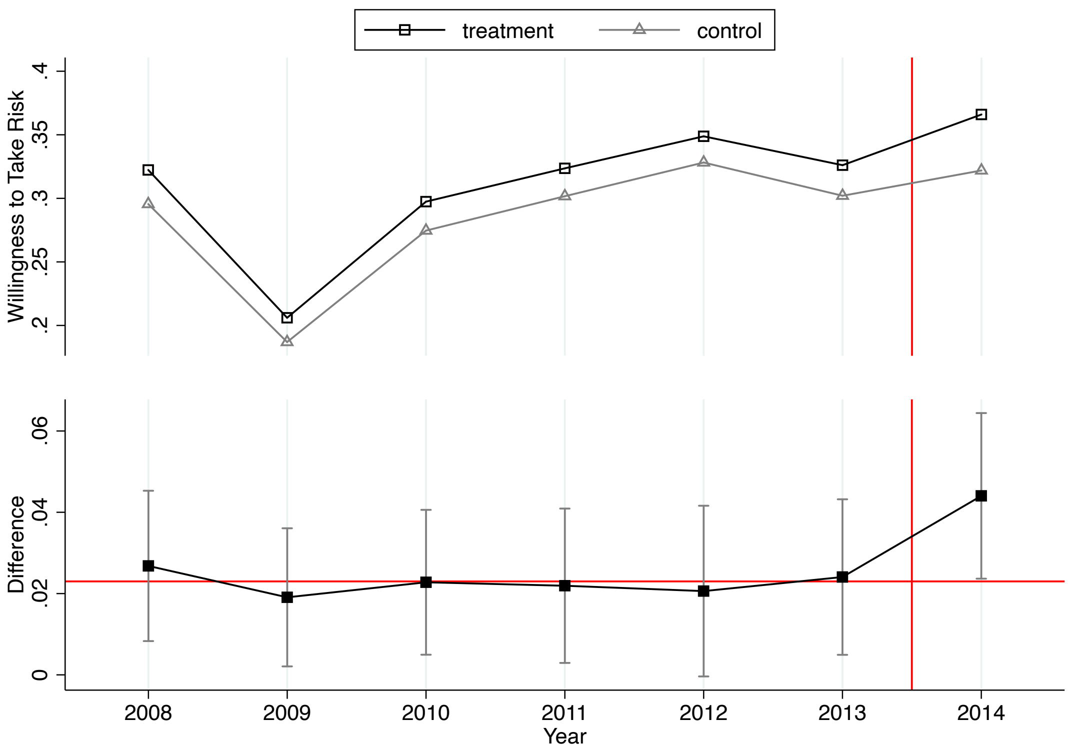
Source: SOEP v33, own calculations. *Notes:* Sample zero includes both treatment and control group in non-treatment periods. Sample one includes treatment group in the treatment period. The median level of life satisfaction is seven.

Figure 3.3: Trends in Google Searches Related to the Crimea Crisis in 2014



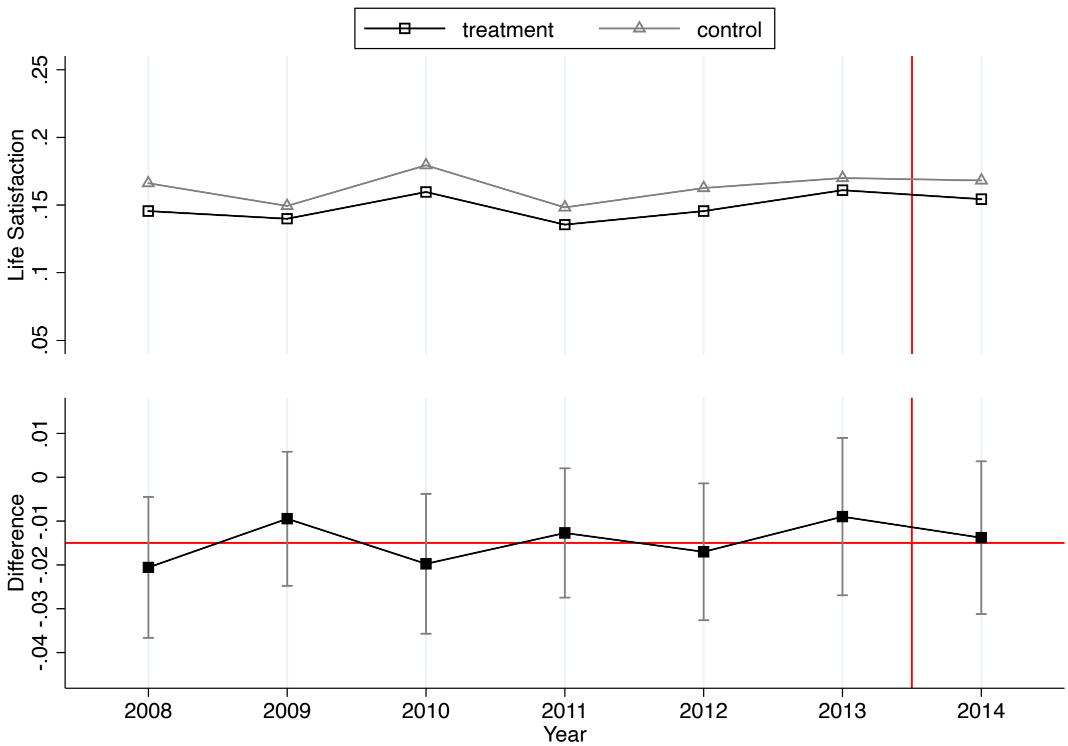
Source: The trend index represents search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of zero means that there was not enough data for this term (see https://trends.google.com).

Figure 3.4: Trends in Risk Attitudes by Treatment Status



Source: SOEP v33, own calculations. *Notes:* The treatment group consists of individuals who are interviewed during or after March in 2014, the control group consists of individuals who are interviewed before March in 2014. The red horizontal line depicts the pre-treatment average difference between treatment group and control group. Confidence bands in the lower panel stem from robust standard errors with adjustment for 1,872 clusters at the level of primary sampling region.

Figure 3.5: Trends in Life Satisfaction by Treatment Status



Source: SOEP v33, own calculations. *Notes:* The treatment group consists of individuals who are interviewed during or after March in 2014, the control group consists of individuals who are interviewed before March in 2014. Life satisfaction refers to having an above median life satisfaction. The red horizontal line depicts the pre-treatment average difference between treatment group and control group. Confidence bands in the lower panel stem from robust standard errors with adjustment for 1,872 clusters at the level of primary sampling region.

3.9.2 Tables

Table 3.1: Descriptive Statistics of Risk Attitudes and Controls

	2008-2013		2014		
	Control (C)	Treatment (T)	Control	Treatment	$\Delta_C - \Delta_T$
RA	0.281	0.304	0.322	0.366	0.021**
LSAT	0.163	0.148	0.168	0.154	0.001
Female	0.530	0.525	0.530	0.526	0.001
Age	54.990	51.612	58.440	55.142	0.080***
East	0.301	0.236	0.301	0.234	-0.002
Abitur	0.217	0.255	0.220	0.259	0.002*
Married	0.653	0.681	0.657	0.690	0.004
N	27,583	31,832	4,585	5,293	-

Source: SOEP v33, own calculations. RA refers to having an above median willingness to take risk. LSAT refers to having an above median level of life satisfaction. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level. Robust standard errors with adjustment for 1,872 clusters at the level of primary sampling region are used.

Table 3.2: Risk Attitudes: Fixed Effects Regressions

	(1)	(2)	(3)
$treat * year2014$	0.022***	0.022***	0.018**
	(0.008)	(0.008)	(0.009)
Married		0.002	0.002
		(0.010)	(0.010)
East		0.023	0.023
		(0.029)	(0.029)
Individual FE	✓	✓	✓
Year FE	✓	✓	✓
Month FE			✓
R^2	0.010	0.011	0.011
N	69,293	69,293	69,293

Source: SOEP v33, own calculations.
Notes: The outcome variable is a binary indicator for having an above median willingness to take risk. Robust standard errors with adjustment for 1,872 clusters at the level of primary sampling region are reported in parentheses. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level.

Table 3.3: Life Satisfaction: Fixed Effects Regressions

	(1)	(2)	(3)
$treat * year2014$	0.001	0.001	0.008
	(0.008)	(0.008)	(0.008)
Married		0.018**	0.018**
		(0.009)	(0.009)
East		-0.027	-0.027
		(0.023)	(0.022)
Individual FE	✓	✓	✓
Year FE	✓	✓	✓
Month FE			✓
R^2	0.001	0.001	0.001
N	69,293	69,293	69,293

Source: SOEP v33, own calculations.
Notes: The outcome variable is a binary indicator for having an above median life satisfaction. Robust standard errors with adjustment for 1,872 clusters at the level of primary sampling region are reported in parentheses. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level.

Table 3.4: Robustness Checks: Estimated Coefficients for Risk Attitudes

	(1)	(2)	(3)	(4)	(5)	(6)
$treat * year2014$	0.022***	0.018**	0.038**	0.016***	0.021**	0.018**
	(0.008)	(0.008)	(0.015)	(0.006)	(0.008)	(0.009)
Excluding Age<25 in 2014	✓					
Logit Regressions (APE)		✓	✓			
Unbalanced Panel Regressions				✓		
January-April 2014					✓	
February-March 2014						✓
Individual FE	✓		✓	✓	✓	✓
Treatment Group FE		✓				
Year FE	✓	✓	✓	✓	✓	✓
R^2	0.010	0.010	0.011	0.011	0.010	0.010
N	68,576	69,293	40,815	99,307	60,544	51,695

Source: SOEP v33, own calculations.
Notes: Robust standard errors with adjustment for 1,872 clusters at the level of primary sampling region are reported in parentheses. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level. Columns (2) and (3) show the average marginal effects from Logit regressions. Individual fixed effects for column (3) are incorporated via a conditional Logit estimation.

Table 3.5: Robustness Checks: Estimated Coefficients for Life Satisfaction

	(1)	(2)	(3)	(4)	(5)	(6)
$treat * year2014$	0.001	0.001	0.002	-0.001	0.002	0.004
	(0.008)	(0.007)	(0.024)	(0.006)	(0.008)	(0.009)
Excluding Age<25 in 2014	✓					
Logit Regressions (APE)		✓	✓			
Unbalanced Panel Regressions				✓		
January-April 2014					✓	
February-March 2014						✓
Individual FE	✓		✓	✓	✓	✓
Treatment Group FE		✓				
Year FE	✓	✓	✓	✓	✓	✓
R^2	0.001	0.001	0.004	0.000	0.001	0.001
N	68,576	69,293	25,605	99,307	60,544	51,695

Source: SOEP v33, own calculations.
Notes: Robust standard errors with adjustment for 1,872 clusters at the level of primary sampling region are reported in parentheses. Statistical significance is denoted by *** at the 1 percent, ** at the 5 percent, and * at the 10 percent level. Columns (2) and (3) show the average marginal effects from Logit regressions. Individual fixed effects for column (3) are incorporated via a conditional Logit estimation.

Chapter 4

Conclusion

Today, in spring 2020, we experience a new profound crisis. It is the crisis caused by measures undertaken to slow down the spread of the new coronavirus, Covid-19. What can one learn from this thesis about how this crisis will affect economic attitudes?

During the so-called refugee crisis in 2015, German residents became on average more anti-immigrant. Chapter 2 argues that this is very likely because of increasing group conflict over scarce resources, because of social identity concerns, and due to the political economy of hatred. Just as in 2015, the corona crisis is likely to increase the conflict over scarce resources such as social benefits and therefore bears the potential of heating group conflicts. But the dividing lines of potential conflicts are much less clear than 2015, when a new group identity of incoming refugees entered the conflict. Moreover, social identity is not linked to the new crisis. However, new conspiracy theories arise and become increasingly popular since the beginning of the corona crisis, and the political economy of hatred may lead some factions to exploit this in the future.

Russia's annexation of the Crimea led to a modest increase in the willingness to take risk among Germans. Chapter 3 argues that this increase shows an adaption behavior of humans to their environment. The perceived risk of living in Europa increased because of Russia's acts, and there was little German residents could do about it. Thus, they learned to live with it. In some sense, the crisis caused by the corona pandemic is the exact opposite to the Crimea crisis analyzed in the third chapter. Although it is not clear at this point how much the spread of the virus will change the riskiness of life, a lot of measures are undertaken to control the risk. Moreover, unlike the Crimea crisis, today's crisis did already impact on economic outcomes throughout the world. Therefore, the realization effect discussed in the second chapter will very likely induce a relevant fraction of the population to become more risk averse. However, different sub-groups in the society are affected very differently by the crisis, and strongly heterogeneous effects are to be expected.

Chapter 2 and 3 focused on attitudes related to very salient topics during the respective crises. The refugee crisis was concerned with non-Germans coming to Germany, so the attitudes towards immigration were affected. Russia's annexation increased Germans' perceived risk of war, so attitudes towards risk were affected. But what is the salient topic during the corona crisis? Today's crisis consists of concerns about health, solidarity towards vulnerable individuals, and working out a worldwide problem in international cooperation. But these concerns blend in with worries about the economy, restrictions of basic rights, and many more. The scope of this crisis is broader and larger than during the so-called refugee crisis in 2015 and the Crimea crisis in 2014. Learning about attitudinal changes induced by the corona crisis will hence be a challenging but very interesting task. The main line of this thesis is that future crises like the corona pandemic will change humans' economic attitudes. And although it may be difficult to forecast in which way attitudes will be affected, Chapter 1 shows how future research can eventually learn about the effects.