The London School of Economics and Political Science

Essays in Applied Microeconomics and Microeconometrics

Claudio Andrea Zeno Schilter

A thesis submitted to the Department of Economics of the London School of Economics and Political Science for the degree of Doctor in Philosophy

London, December 2019

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 40,000 words.

Statement of Conjoint Work

I confirm that chapter 2 was jointly co-authored with Karun Adusumilli and Friedrich Geiecke and I contributed 30% of this work.

Statement of Inclusion of Previous Work

I can confirm that chapter 3 is based on the result of previous study (for a Master of Research award) I undertook at the London School of Economics and Political Science.

Statement of Use of Third Party for Editorial Help

I can confirm that section 1.1 was copy edited for conventions of language, spelling, and grammar by the LSE Language Center and 24x7 Editing.

Acknowledgments

I am deeply grateful to my supervisor Maitreesh Ghatak, to Gharad Bryan, and to Tom Kirchmaier, for their invaluable guidance, support, and advice. I have also greatly benefited from the feedback of faculty members at the LSE and beyond, including Nava Ashraf, Konrad Burchardi, Jonathan de Quidt, Greg Fischer, Tatiana Komarova, Matt Levy, Stephen Machin, Alan Manning, Rachael Meager, Guy Michaels, Taisuke Otsu, and Balázs Szentes. Moreover, I am indebted to Roger Pegram from the Greater Manchester Police for his interest and support.

I would also like to thank my fellow PhD students inside and outside of our amazing office 4.06, including Shan Aman-Rana, Andrés Barrios Fernández, Alexia Delfino, Weihan Ding, Dita Eckardt, Eui Jung Lee, Benedetta Lerva, Ria Ivandic, Felix Koenig, Niclas Moneke, and Vincenzo Scrutinio. I have learned so much from you and the PhD would not have been the same without you. A very special thanks goes to my coauthors, Karun Adusumilli and Friedrich Geiecke, for our discussions, all your advice, and your friendship throughout the PhD.

Special mention must also be made of my friends outside of academia, including Ioannis Andreadis, Estelle Baeriswyl, Fabienne Berchtold, Philippe Gassmann, Dominic Ketelaar, Raphael Lingg, Silvio Lustenberger, Thomas Schmid, Simon Steiner, Dario Stocker, Nishant Tharani, and Linus Zweifel. Many of you have helped me directly at some point of my PhD journey, but most importantly, all of you kept me sane and happy over all these years.

This thesis would not have been possible without the constant love and support from my wonderful parents.

Last but certainly not least, thank you Yvonne, for your love, help, encouragement, patience, for lifting me up when needed, for reading and listening to so many research ideas, and for all your other support during my PhD years.

Abstract

First, I investigate the change in hate crime targeting race or religion after the Brexit vote. My results reveal a substantial and transitory increase in such hate crime following the vote. The focus of my analysis is the considerable spatial heterogeneity of this increase. Areas with a greater increase in hate crime are characterized by both a greater immigrant share and higher income proxies. Issues of multiple hypothesis testing and model selection limit the use of classic methods; therefore, I apply and adapt recent machine learning methods to uncover patterns in the spatial heterogeneity.

I then focus on the question how to utilize data from randomized control trials to obtain an optimal dynamic treatment rule. Consider a situation wherein individuals arrive sequentially - for example when becoming unemployed - to a social planner. Once each individual arrives, the planner decides instantaneously on a treatment assignment - for example job training - while taking into account the characteristics of the individual and the remaining capacity to offer training. In order to decide optimally, expectations over the dynamic process of unemployment patterns are required. Reinforcement learning methods can be used to solve this dynamic optimization problem and the resulting algorithm has a number of desirable properties.

Finally, I study the creation of not-for-profit firms. Reputation is key for high-quality producers when quality is only observed after the time of purchase. For companies that potentially enter several markets, I show that the concern for reputation affects both the optimal organizational form and the decision which markets to enter. Specifically, a market with poor customers that would be ignored in isolation can be served for signaling purposes. The optimal organizational forms in that case are a not-for-profit firm used for signaling in the "market for the poor" and an associated for-profit firm in the "market for the rich".

Contents

List of Tables

List of Figures

Chapter 1

Hate Crime after the Brexit Vote: Heterogeneity Analysis based on a Universal Treatment

1.1 Introduction

Several countries have reported an increasing number of hate crimes, including the United States, Italy, and England.[1] The harm from such violence is not limited to the moment of the act. The hate-based motivation carries the threat of repeated targeting of the victims and their families.[2] Such violence affects a substantial part of the population and challenges policy makers around the globe.[3]

This paper studies hate crime targeting race or religion in the context of the United Kingdom European Union membership referendum ('Brexit vote'). On June 23 2016, the UK decided to leave the EU, defying most polls (e.g. Lord Ashcroft, 2016). News reports and politicians have associated an upsurge in hate crime with the Brexit vote (e.g. BBC, 2017a; Time, 2017; Al Jazeera, 2017; Financial Times, 2017).[4] The broad press-coverage and governmental reports (e.g. Home Office, 2017) indicate that insight into underlying mechanisms is needed. Moreover, analyzing this shock to hate crime contributes to a better understanding of hate crime in general.

This paper not only confirms that the Brexit vote led to an increase in racial or religious hate crime, but uses regional data to investigate the spatial heterogeneity of this increase. As an agnostic basis to evaluate potential mechanisms, I use the the temporal structure of the effect as well as the spatial heterogeneity combined with area-characteristics.

The key findings show that the Brexit vote led to a substantial increase in racial or religious hate crime for approximately six weeks after the vote and had no effect before. In July 2016, the magnitude of the increase was 21% (550 hate crimes) in Greater London and Greater Manchester, with considerable spatial heterogeneity captured by borough-level[5] census and vote data. The average increase is statistically insignificant at 14% in the tercile of boroughs with the lowest predicted effect, but significant at 28% in the tercile with the highest. I find that mainly proxies of the migrant share, income, and wealth are strongly (positively) associated with a higher increase in hate crime after the vote. These findings are robust across different methods and different from the heterogeneity of the increase in hate crime observed

[1] Levin & Reitzel (2018), Monella (2018).

[2] See Craig-Henderson & Sloan (2003) or McDevitt et al. (2001), who outline the psychological costs of hate crime.

[3] See Hall (2013) or Gerstenfeld (2017), who highlight a large number of policy efforts in various countries.

[4] Others dispute the connection between the vote and the rise in hate crime (e.g. Daily Mail, 2016; Spectator, 2017).

[5] A borough is an administrative division. In London and Manchester, the average borough-population is 275,000.

after terror attacks (see also Ivandic et al., 2018).

The main empirical challenge concerns model selection. Scores of economic models from several fields are potentially applicable. Being completely agnostic, more than 10^{16} linear models of the spatial heterogeneity can be formed with the 68 variables that characterize the areas in my main dataset. Since OLS is designed to test but not select a model, standard OLS is not a suitable method. In addition, the issue of seed dependency arises.[6] Finally, the Brexit vote was a unique event by which all regions were treated simultaneously, resulting in the lack of a control group.

To address these challenges, I use a unique dataset and apply and adapt state-of-the-art methodology. While definitive causal statements are not possible, this method results in an agnostic basis on which potential mechanisms can be evaluated. Namely, I do not choose a specific model (out of the 10^{16} possible models) ex-ante but obtain a model as a result, complete with parameter estimates and significance statements.

The dataset has two key components. The first is racial or religious hate crime data from Greater London at the borough-month level. This was only recently made publicly available by the Metropolitan Police. The second is detailed confidential high-frequency data on hate crime from the Greater Manchester Police. Joining the two datasets at the borough-month level results in panel data of 42 boroughs over 88 months, which I am the first to construct and employ. The panel structure with a monthly frequency allows me to conduct a thorough heterogeneity analysis, in particular allowing for spatial heterogeneity in the short term effect.[7]

Contrasting a number of possible mechanisms with the key results of the analysis, my preferred interpretation is that the Brexit vote affected hate crime mainly through information-updating. Namely, individuals updated their information about society's attitude towards immigrants.[8] One specific channel is that the information update led to an increase in hate crime in those boroughs where the expected social cost of offending had been high and the opportunities for offending (i.e. the presence of victims) were plentiful. The former is in line with the increase being pronounced in wealthier areas (see e.g. Mayda, 2006) and the latter with it being more pronounced in areas with a higher migrant share. An alternative channel is that people who were more surprised by the Brexit vote's outcome had a larger information-update which resulted in a larger increase in hate crime (for details see also Albornoz et al., 2018). In that channel, potential offenders operate locally and their beliefs about the national attitudes are biased towards the regional attitudes. A proxy for the latter is the regional Brexit vote, which is correlated with measures of wealth and migrant share.

Conversely, the results provide evidence against a number of alternative mechanisms. A first prominent example are mechanisms building on the unemployment share (e.g. Falk et

[6] Some methods rely on a single data split, usually in a training and a test sample. While the split is random, different seeds lead to different splits and, since the number of observations is finite, potentially to different results.

[7] Using respective time series hate crime data, I show that the aggregated time series of racial or religious hate crime in these two metropolitan areas mirrors that of England and Wales. However, the spike after the Brexit vote appears to be more pronounced in London and Manchester. While the difference is insignificant, it points towards my results since the migrant share is higher in London and Manchester.

[8] Immigration was a key topic, see e.g. Goodwin & Milazzo (2017) or Meleady et al. (2017).

al., 2011, Krueger & Pischke, 1997). Opportunity cost theories of hate crime are also unfit to explain why the increase in hate crime was larger in boroughs with higher wages.[9] Moreover, the transience of the increase in hate crime excludes mechanisms based on altered fundamentals, expected or actual, as these did not revert back after six weeks. One example is the exchange rate, which partially embeds expected changes in other fundamentals and dropped for a sustained period after the vote (see Douch et al., 2018). Furthermore, the lack of an increase in hate crime prior to the vote is evidence against aggressive media coverage (alone) having an effect. The Brexit coverage featured immigration as a key topic long before the vote (see Moore & Ramsay, 2017).

In terms of methodology, this is one of few papers to employ the recent advances in machine learning to obtain valid inference on heterogeneous treatment effects.[10] After evaluating the magnitude and temporal structure of the 'Brexit effect', I proceed in four steps to analyze the spatial heterogeneity. First, I use an adapted version of Chernozhukov et al.'s (2018b) approach to analyze the magnitude of spatial heterogeneity captured by the 'candidate variables', i.e. the variables from the census and vote data. Second, I apply conditional post-selection lasso (CPSL; see Lee et al., 2016; Tibshirani et al., 2016) to obtain a linear prediction model of that heterogeneity. Third, I propose a novel splitting based estimation method which allows for a quasi-linear model with interactions. The standard method to analyze treatment heterogeneity is to interact treatment with the variable of interest. As a complementary final step, I run multiple regressions with each candidate variable individually and adjust the result for multiple hypothesis testing. Model selection remains a problem in this last step, but it allows me to test the individual correlations of each candidate variable with the increase in hate crime, and benchmark the results of the previous methods against these correlations. To the best of my knowledge, the application of Chernozhukov et al.'s (2018b) approach and also the use of CPSL to analyze heterogeneous treatment effects is novel. I provide an adaption and application of these methods to a common situation in economics, where a single event leads to universal treatment.

I evaluate the overall magnitude and temporal structure of the 'Brexit effect' using standard methods. At the month-borough level, I find the effect to be predominantly present in July 2016, the month following the vote, which is relevant for the subsequent heterogeneity analysis.

To evaluate the spatial heterogeneity, I first measure the abnormal hate crime in July 2016 as the difference between the observed and the predicted number of hate crimes.[11] For that measure, I show the spatial heterogeneity that is captured by the candidate variables to be significant (building on Chernozhukov et al., 2018b).[12] What remains to be explained is the

[9] The opportunity cost of time of potential offenders is key in such theories (see e.g. Medoff, 1999).

[10] Other examples include Bertrand et al. (2017), or Knaus et al. (2017).

[11] This measure also allows for the use of simple machine learning methods to predict the change in hate crime for each borough, which can be of direct interest for policing and policy. If the vote and census data is sufficiently explanatory, the spatial heterogeneity in predicted changes is more informative than the heterogeneity in the measure itself due to potential over-fitting on noise in the measure (see e.g. Hastie et al., 2009).

[12] Due to the fact that every borough is treated simultaneously, the overlap condition fails, prohibiting me to use propensity score matching (as done by Chernozhukov et al., 2018b). Instead, I use the measure mentioned above. This measure could potentially be comprised of noise with an arbitrary spatial heterogeneity. Evidence from permutation inference refutes this concern: compared to results from the same procedure, but using the

cross-borough heterogeneity in the single time period July 2016, using a subset of candidate variables. Consequently, the CPSL method is directly applicable, despite the fact that it was not designed for analyzing heterogeneous treatment effects. As a result, a parsimonious linear prediction model of the conditional average treatment effect (CATE) is obtained with valid inference for its parameters. In brief, the CPSL method first uses the standard lasso method to select a model.[13] For the variables contained in the chosen model, confidence intervals are obtained and the lasso parameter estimates are adjusted.

In addition, I develop an ad hoc multiple splitting estimation method (building on Chernozhukov et al., 2018b; Athey & Imbens, 2016, 2017; Rinaldo et al., 2018). Since CPSL theoretically assumes i.i.d. data, I first use the splitting based approach to confirm the linear model resulting from CPSL. While some assumptions are stronger, my approach allows me, for example, to correct for heteroskedasticity.[14] Moreover, interactions of candidate variables can be included to obtain a quasi-linear model instead.[15] The splitting based estimation uses lasso on a random half of the sample to select a model, and the other half to estimate that model. For a single sample split, this leads to valid inference and overcomes the issues of multiple hypothesis testing and model selection. However, in my setting with a finite number of boroughs, different models are obtained depending on the split. This seed dependency challenge is overcome by repeating the process 1000 times. Unfortunately, aggregating the estimates across iterations results in theoretically ambiguous bias. A series of simulations using this paper's data suggest that this bias is negligible in the current context.

The splitting based estimation and the CPSL method arrive at highly similar linear prediction models, providing evidence in favor of their validity in this setting. When the splitting based estimation is used to obtain a quasi-linear model (allowing for interactions and squared terms), the result is again similar. For the absolute effect, it is even the case that the same interaction-free model is chosen again.

In sum, my proposed approach for analyzing the spatial heterogeneity not only proves to be robust, but also provides interesting findings that are arguably free from model selection bias, which would debilitate any heterogeneity analysis relying on standard methods.

Related Literature

This paper contributes to the literature on the economics of hate crime, which is the intersection between three literature strands: the economics of crime, the economics of conflict, and the economics of taste-based discrimination and racism. In addition, parts of the specific

other months as placebo treatments, July 2016 is clearly a tail event.

[13]The lasso (least absolute shrinkage and selection operator, Tibshirani, 1996) is related to OLS, but penalizes the magnitude of the estimated coefficients. Less important coefficients are shrunken to zero. The remaining variables can be interpreted as a data-driven model choice. The caveat that lasso-based methods cannot guarantee the selected variables to be the true data generating process due to correlations in the candidate variables is limited. The true variables are likely unobserved, meaning that the realistic (and achieved) objective is finding an insightful descriptive model.

[14]The key assumption of CPSL is i.i.d. data, although Tibshirani et al. (2018) suggest that the approach could be valid more broadly. The key assumption of my splitting based approach is that the theoretically ambiguous bias is negligible.

[15]More precisely, it allows for interactions under the condition that the non-interacted candidate variables must also be included in the model. This condition makes the interactions interpretable. This is enforced by using hierarchical lasso (see Bien et al., 2013, Lim & Hastie, 2015) instead of regular lasso to obtain the model.

Brexit literature are related to this paper as well. Finally, methodologically, this paper builds on the literature devoted to obtaining valid inference from machine-learning-based methods.

Both the taste-based discrimination and the crime literature date back to seminal work by Becker (1957, 1968). Even when discriminatory tastes are fixed, whether or not people publicly act upon them depends on the expected costs of the action, including social costs. Related to the current paper, Bursztyn et al. (2017) build on this argument and show that the willingness to publicly act in a xenophobic way changes quickly when information about the preferences of society regarding migrants becomes available through a controversial vote. They focus on the Trump election, and while their laboratory setting allows for a better analysis of the precise mechanisms, it is impossible to develop strong statements pertaining to real life decisions, especially violence.

Related papers from the conflict literature include Esteban & Ray (2011), Esteban et al. (2012), Caselli & Coleman (2013), and Mitra & Ray (2014). More specifically related, Blair et al. (2017) use machine learning methods (namely lasso, random forests, and neural networks) to forecast local violence. Their setting allows for the predictions to be tested. Lasso is shown to be the best performing method, which has inspired the current paper to make use of lasso-based methods.

Regarding the effect of the Brexit vote on hate crime, Devine (2018) also finds that the Brexit vote led to a significant increase in racial or religious hate crime. He relies on time series intervention models and does not address the heterogeneity of the effect. In a paper complementary to mine, Albornoz et al. (2018) develop a detailed model, particularly focusing on effect-heterogeneity. The key difference to my paper is that, regarding the heterogeneity of the increase in hate crime, they focus exclusively on the Brexit vote shares, while I use data-driven models.

Further regarding the Brexit vote, but related to the vote outcome rather than crime, Becker et al. (2017) analyze the spatial variation in the vote with candidate variables similar to this paper. Other than the role of income proxies, which are positively associated with both the remain vote share and the increase in hate crime, the important variables are different.[16]

Finally, this paper is related to the literature about finding a (interpretable) model for the conditional average treatment effects (CATE) (see Chernozhukov et al., 2018b, for a recent review). In addition to the literature mentioned above when outlining the methods, Wager and Athey (2018) specifically address heterogeneous treatment effects. They use a 'causal forest', but emphasize that the method relies on large samples. Furthermore, forests are considerably harder to interpret than lasso results.

The remainder of this paper is structured in the following way. Sections 2 and 3 describe the data and analyze the overall effect of the Brexit vote on racial or religious hate crime. Section 4 outlines how the spatial heterogeneity of the effect is measured and shows that it is

[16] Fetzer (2018) suggests that austerity measures were pivotal for the vote. These measures have predominantly concerned public sector employees and receivers of social benefits, starting from the year 2011. In the current study, 2011 census data is used which includes the share of public sector workers, social housing renters, and further proxies for the share of people receiving benefits. These measures are not of central importance for the post-vote increase in hate crime.

substantial. The methodology used to model the heterogeneity is outlined in section 5, and its results are presented subsequently in section 6. Section 7 discusses the implications of the results for potential economic mechanisms at play as well as policy considerations. Finally, section 8 concludes the paper.

1.2 Datasets and Summary Statistics

Regarding crimes, panel data from the police forces of two of the three largest metropolitan areas of England is used: Greater London and Greater Manchester. For the Greater Manchester Police (hereafter GMP), I use extensive confidential data from April 2008 to July 2017. This data contains all incidents the GMP has attended. Among other things, the type of offense, its severity, time, and precise location is recorded.[17] A subset of these incidents classify as crimes, which is the focus of this analysis. In principle, every crime has at least one offender. However, it is only rarely the case that the offender is known (which results in limited and selected data on offenders). It is moreover recorded whether or not an incident is classified as a hate incident/crime, and if so categorized respectively (hate targeting race, religion, disability, sexual orientation, etc). This paper focuses on racial or religious hate crime, which is by far the most common type of hate crime. Moreover, most racial or religious hate crimes are racial hate crimes.

Regarding the Metropolitan Police of Greater London, the monthly total of racial or religious hate crimes for each of its 32 boroughs is available, from April 2010 to April 2018 (see Metropolitan Police, 2018).

In addition, time series data for 38 aggregated police forces from England and Wales is provided by the Home Office (2017).[18] Monthly data is available from April 2013 to August 2017, and daily data from April 2016 to August 2017. Time series data is not suitable for the heterogeneity analysis, but it can provide insight with regard to the overall response to the Brexit vote (see also Devine, 2018).

In my main analysis regarding the spatial heterogeneity, I combine the data from London and Manchester and focus on racial or religious hate crimes per borough-month and population million.[19] Summary statistics about this data are presented in Table 1.1 and further visualizations of the aggregate data (including details available for Manchester only) are provided in section 1.3. Table 1.1 shows the variance in hate crimes is considerable and already suggests

[17] While the data is unusually detailed, information about victims, which is highly sensitive and confidential, could not be obtained.

[18] There are in total 43 regular police forces in England and Wales, plus the transport police. The 38 forces that did provide the data are: Avon and Somerset, Bedfordshire, British Transport Police, Cambridgeshire, Cheshire, City of London, Cleveland, Cumbria, Devon and Cornwall, Durham, Dyfed-Powys, Gloucestershire, Greater Manchester, Gwent, Hampshire, Hertfordshire, Humberside, Kent, Lancashire, Lincolnshire, Merseyside, Metropolitan Police, North Wales, North Yorkshire, Northamptonshire, Northumbria, Nottinghamshire, South Wales, South Yorkshire, Staffordshire, Surrey, Sussex, Thames Valley, Warwickshire, West Mercia, West Midlands, West Yorkshire, Wiltshire.

[19] Borough-level population estimates are available for the relevant years up to and including the key year 2016 (Office for National Statistics, 2017). Thereafter (i.e. for the 7 months in 2017), the 2016 population data is used.

that while there are more hate crimes in July in general, there were even more in July 2016.

	Manchester	London	Total
Min. Hate Crimes per Month	9	0	0
Max. Hate Crimes per Month	417	595	595
Mean Hate Crimes per Month	94	118	112
[N_{Man} = 880; N_{Lon} = 2816; N_{Tot} = 3696]	(52)	(66)	(64)
Mean July Hate Crimes	110	149	140
[N_{Man} = 80; N_{Lon} = 256; N_{Tot} = 336]	(59)	(86)	(82)
Mean July 2016 Hate Crimes	138	245	219
[N_{Man} = 10; N_{Lon} = 32; N_{Tot} = 42]	(65)	(110)	(111)

Note: Standard deviations in parentheses. Hate crimes are measured in terms of the monthly number of racial or religious hate crimes per million of borough population. The mean total number of monthly hate crimes per borough is 27 for Manchester, 31 for London, and 30 in total. 42 boroughs (10 and 32) are observed over 88 months.

Table 1.1: Racial or Religious Hate Crime Summary Statistics

While there are strict guidelines on the side of the police to minimize the influence of reporting, the definition of hate crime is subjective. According to a general agreement of the major agencies involved, a hate crime is defined as: “Any crime which is perceived by the victim or any other person to be motivated by hostility or prejudice based on a person’s race” (Home Office, 2017). There are therefore both the criminal actions by the offenders as well as the reporting behavior of the victims and witnesses that potentially affect the recorded hate crimes. While the fact that only reported crimes enter the data is a common issue of research in crime, the ramifications of this in the specific context of this paper will still feature in the discussion (see section 1.7.1). Crucially, further data from the Home Office (2017) regarding ‘hate’ and ‘non-hate’ crimes of specific offense-types supports the claim that (considerable) relabelling of non-hate crimes to hate crimes did not occur.

The Brexit vote data on the counting area level is publicly available from the Electoral Commission (2016). This data includes all 382 local authority districts in the UK, 42 of which are covered by the operating area of the GMP and the London Metropolitan Police. In the urban context, these districts are boroughs. The overall outcome of the Brexit vote is commonly considered to be surprising. For example, data from a survey in May 2016 by Lord Ashcroft (2016) suggests that only 35% expected the leave campaign to win.

Finally census data is used from the 2011 census (Office for National Statistics, 2016). The census is taken every decade, so the 2001 census is considerably outside of the period for which crime data is available. The census contains a large number of correlated variables. Avoiding very high correlations while also aiming to lose as little information as possible, I have selected 67 variables.[20]

[20] These variables are fractions of people with certain characteristics living in the area of interest, namely: male, single, same sex civil partnership, divorced, Christian, Buddhist, Hindu, Jewish, Muslim, Sikh, other religion, not stating religion, aged younger than 16, between 16 and 29, between 30 and 64, over 64, living in a one person household, lone parents, social housing renter, white, South Asian, other Asian, black, Arab, of mixed ethnicity, born in an old EU state (joined prior to 2000), born in a new EU state, born in the rest of

The variables from the vote and census data are the 'candidate variables' I consider in the main analysis regarding the heterogeneity of the increase in racial or religious hate crime after the Brexit vote. Summary statistics about them (using the average across candidate variables) are outlined in Table 1.2. The statistics confirm that there is cross-borough variation in the vote and census data and show to what extent the variables are correlated.

	Manchester	London	Total
Average Borough-Mean	0.168	0.168	0.168
	(0.224)	(0.205)	(0.208)
Average Borough-Minimum	0.138	0.111	0.108
	(0.200)	(0.172)	(0.170)
Average Borough-Maximum	0.203	0.246	0.253
	(0.247)	(0.246)	(0.255)
Average Absolute Correlation	0.46	0.36	0.39
	(0.27)	(0.23)	(0.24)

Note: Standard deviations in parentheses. The average refers to the averaging across the 68 census and vote variables. Mean, minimum, and maximum refer to a comparison across boroughs (32 in London, 10 in Manchester). The absolute correlation refers to the pairwise correlation across the 68 census and vote variables.

Table 1.2: Candidate Variables Summary Statistics

1.3 Empirical Analysis of the Aggregate Effect

Every region in the UK has been treated simultaneously by the Brexit vote. Event time therefore coincides (up to a constant) with real time. This makes the analysis of its effects challenging, especially its short term effect. Including dummy variables for a short period after the vote in a regression of racial or religious hate crime on several controls is prone to produce confidence intervals that are not valid for the effect of the Brexit vote. Asymptotically, validity is achieved, but this is asymptotic with respect to the length of the 'short period'. Weather, for example, could affect crime, news from across the globe, and countless other factors. Trying to control for many things can help, but is doomed to virtually never achieve evidence that is truly informative about the effect of interest.

Permutation tests can address this issue, and also visual inspection results in evidence for a strong 'Brexit effect'.[21] The approaches indicate that racial or religious hate crime has increased after the Brexit vote for a period of approximately six weeks. No permanent effect

the world, born in the UK, not speaking English, immigrated within the last 2 years, providing unpaid care, of bad health, disabled, fully deprived, not deprived, having central heating, having no qualifications, and being economically active. In addition, of those that are economically active, the share of unemployed, self employed, working from home, and working part time. Finally, the fraction of 2 bedroom flats as well as the share of people working in each of the 18 main industries (highest level of aggregation), and in each of the four social grades: AB, C1, C2, and DE.

[21] The use of difference-in-difference as well as synthetic control methods also support this insight. They faces some challenges in the setting at hand and are discussed in Appendix A.1 and A.2.

could be detected. The effect is visually obvious for the combined 38 police forces and London. For Greater Manchester, it is less pronounced. In sum, the 'Brexit effect' seems to exist overall, but there are signs for the importance of spatial heterogeneity.

1.3.1 Descriptive Data Visualization

Visual inspection of the time series of racial or religious hate crime leave little doubt that the increase fallowing the Brexit vote was more than coincidence. In Figure 1.1 'Brexit' indicates July 2016 and a clear, unprecedented high. Following a synthetic control approach leads qualitatively to the same result (see Appendix A.1), and also provides evidence that this is a hate-crime specific phenomenon. The combined London and Manchester data form the basis of my analysis in the following sections of this paper, as only there I possess the relevant data for the heterogeneity analysis.[22] Figure 1.1 shows that it is qualitatively similar to the overall data from the 38 forces of England and Wales, the spike after the vote possibly being more pronounced in London and Manchester (which is in line with my results in section 1.6).

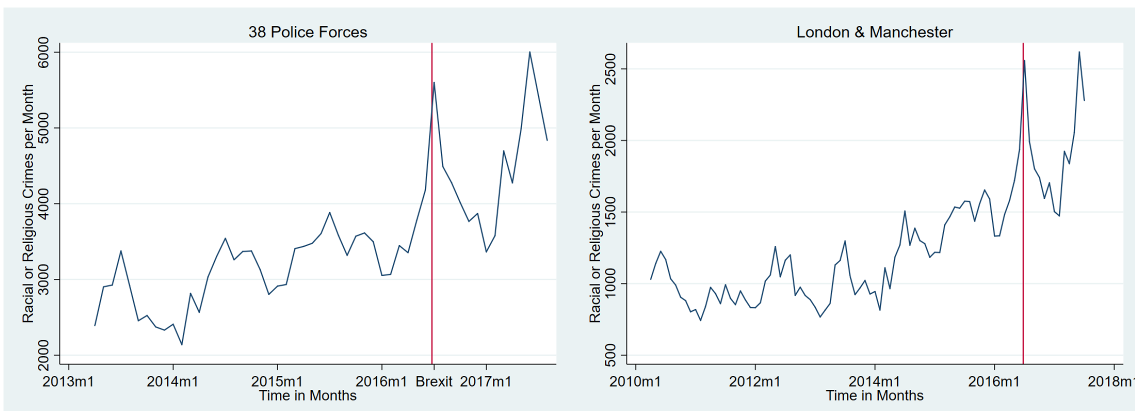


Figure 1.1: Racial or Religious Hate Crime over Time: England and Wales vs London and Manchester

There is a second clear spike in Figure 1.1, which corresponds to June 2017. This is arguably a consequence of the terror attacks in Manchester (22$^{\text{nd}}$ of May), and London (3$^{\text{rd}}$ of June and 19$^{\text{th}}$ of June). Another attack took place in London on the 22$^{\text{nd}}$ of March, potentially explaining the increase before the second peak. Such major terror attacks on English soil did not appear in the rest of the time period at hand (the previous major incidence was the '7/7 attack' in July 2005). Terror attacks are not the focus of this paper, but Ivandic et al. (2018) use the same Manchester dataset to analyze the effect of terror on hate crime (and find especially strong effects after attacks on English soil).

While the 'Brexit effect' appears rather strikingly in the 38 combined forces, especially in London, it is less pronounced for Manchester. This is a first sign for the importance of spatial

[22] As discussed previously, different hate crime data sources are used. To the best of my knowledge, they are not in conflict and can be added (just as the 38 police forced were added by the Home Office). Still, force-specific differences are imaginable. In the heterogeneity analysis, I will specifically allow for the possibility of a differential effect of Brexit by force, but this dummy is not picked up. Regarding the main analysis, I include borough fixed effects, which implies force dummies. In Figure 1.1, the data was simply added.

heterogeneity. Moreover, at least for Manchester where I possess the data needed to make this distinction, the 'Brexit effect' was almost exclusively driven by racial (but not religious) hate crime. The respective figures can be found in Appendix A.6.

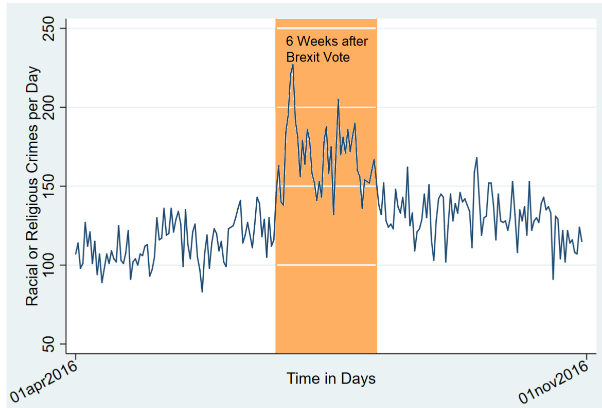


Figure 1.2: Concentration 6 Weeks after the Vote and No Pre-Vote Effect Visible (38 Police Forces)

In terms of temporal structure, the monthly data in Figure 1.1 shows an increase in crime numbers mainly in July 2016. Figure 1.2 uses daily data for the 38 forces. There is no visible effect prior to the vote. After the vote, the effect seems to be pronounced for approximately six weeks. Figure 1.3 uses the weekly data for Manchester and compares the average effect of periods of different lengths (but all starting the day after the vote where the results were made public, the 24$^{\text{th}}$ of June).[23] This average effect is ranked compared to all other possible (overlapping) periods[24] of the same length starting at another week (488 weeks are in the data). After 6 weeks, adding additional weeks to the duration is clearly harming the average effect's ranking. This is evidence against the effect lasting for longer than 6 weeks. The 'Summer of Terror' 2017 makes it difficult to visually evaluate long-term effects. However, especially in the longest time series, that for London, there does not seem to be an obvious long-term effect of the Brexit vote (see Appendix A.6).

Comparing the behavior of racial or religious hate crime to some key factors of the environment, no obvious alternative explanation can be found. If anything, Figure 1.4 shows that July 2016 was a little dryer than other Julys in England, and that the immigration of EU job-seekers was at a high just before the Brexit vote.[25] The latter is arguably related to the topic at hand, but the increase was rather gradual, and resembles in no way the clear spike after the vote. The increase happened before the vote, and thereafter, there was a decrease that was steady but not immediately dropping.

[23]The data from Manchester is suitable for such an exercise due to the long pre-vote period. This lacks the daily data of the 38 forces.

[24]Periods are blocks of weeks. This visualization is related to the permutation tests discussed in section 1.3.2.

[25]In the graphs with quarterly data, the red line indicates the third quarter 2016, starting on the first of July.

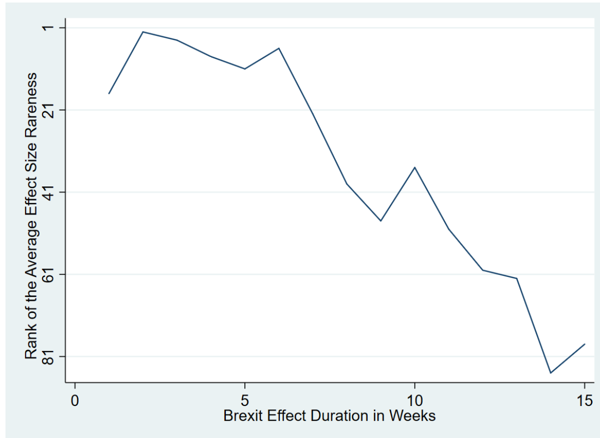
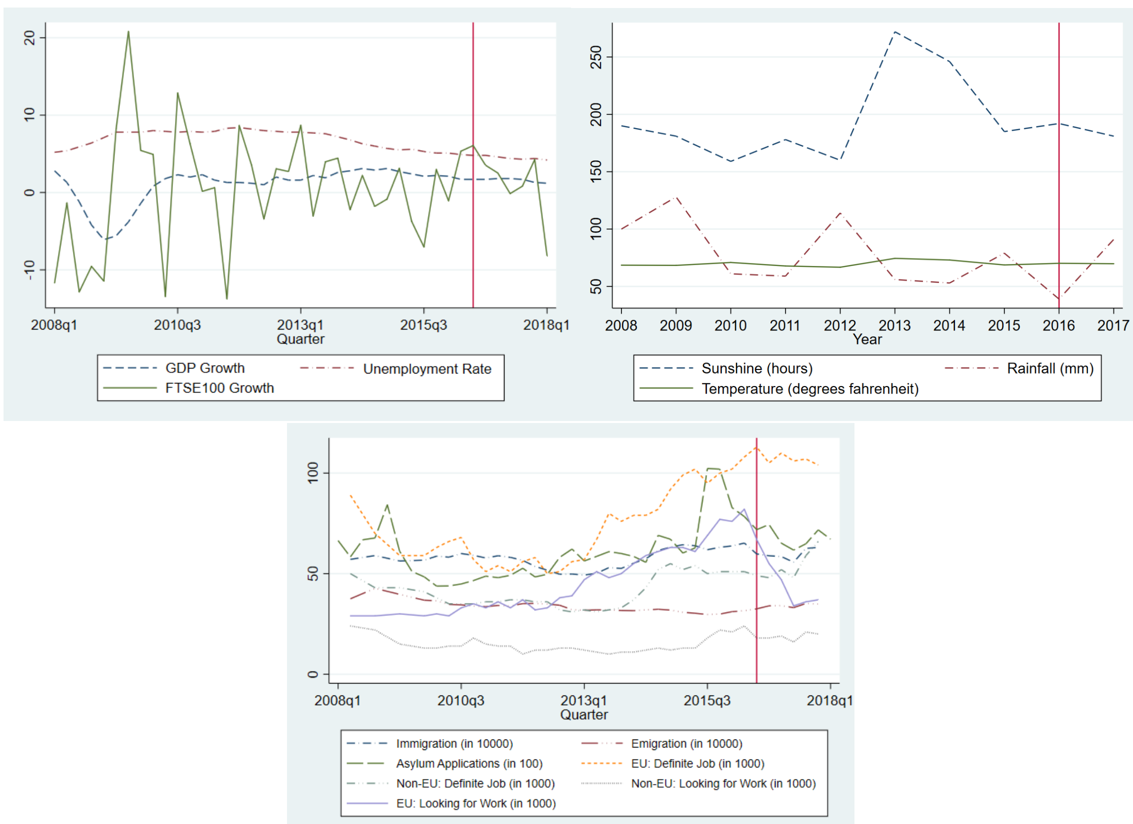


Figure 1.3: Rarity of Spike Depending on Duration Confirms 6 Week Effect-Duration (Manchester)



Note: Weather data for England, other data for the UK. Data sources: MetOffice (2018), Office for National Statistics (2018a, 2018c), London Stock Exchange (2018), Home Office (2018).

Figure 1.4: No Comparable Spike in Migration, Economic Activity, or July Weather

In terms of the type and severity of the crimes that are flagged as racial or religious hate crime, the differences are marginal. As shown in Figure 1.5 for Manchester (for which I possess the relevant data), if anything, violence against the person is relatively more common in the period after the vote, presumably at the expense of criminal damage.[26] In terms of

[26]Other crime types that have been flagged as racial or religious hate crimes are: sexual offenses, theft offenses, possession of weapons, miscellaneous crimes against society, and fraud. Individually, they all make

severity, there is no significant difference in the most severe crimes, which are arguably the most important ones both in terms of their frequency and effect on the victim. There is a small increase in crimes of the lowest severity. Finally, in terms of how the police was informed about the crime, there are again only minor differences. Calls over the radio (i.e. directly from police officers) seem to have decreased and miscellaneous methods have increased.

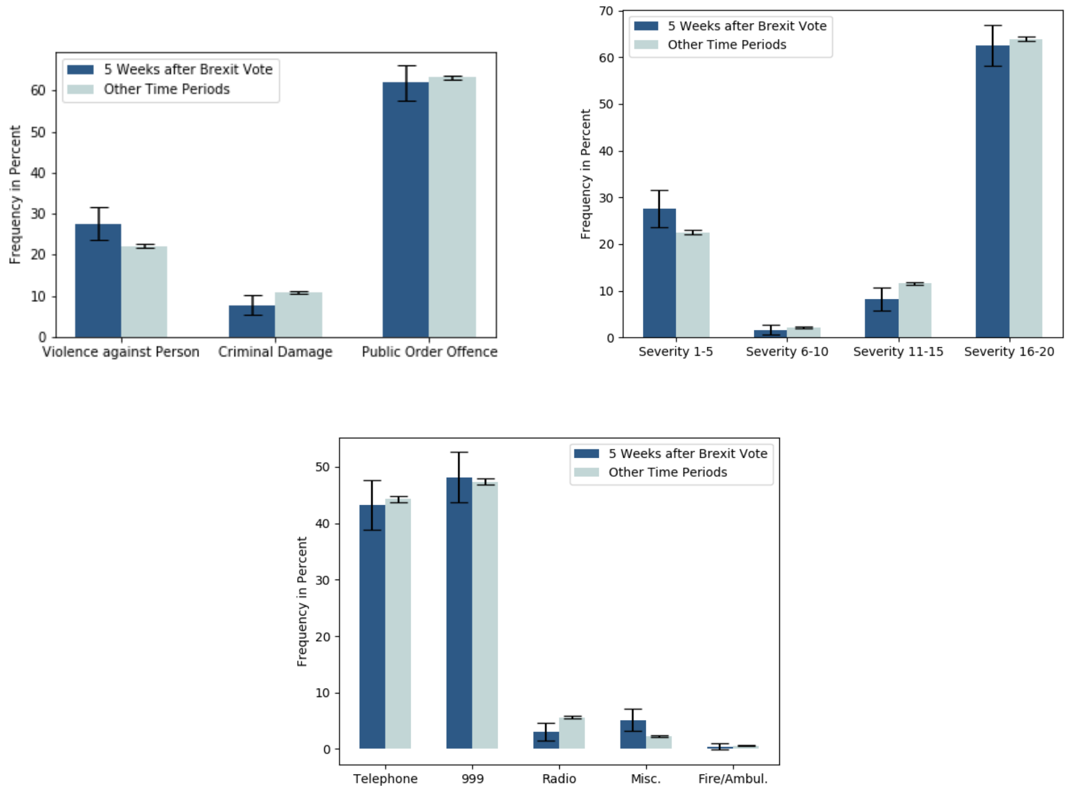


Figure 1.5: Only Small Differences in Method of Call, Crime Types, and Severity (Manchester)

1.3.2 Permutation Tests

The idea of permutation tests dates back to Fisher (1935). Using permutation tests for dependent data, which is generally implied by time series data, has evolved later and blocks of data can be used in a way related to block bootstrapping (see Kirch, 2007).[27] I consider July 2016 as potential short term effect, and the period thereafter as potential long term effect. As the data is at the monthly level, the duration is therefore only one period (and the remainder of the data respectively) and hence no blocks are necessary.[28] A common use in current economics for permutation tests in the time dimension is in the setting of synthetic

up less than 1.7% of all racial or religious hate crime cases.

[27]The major difference between block bootstrapping and block permutation tests is that in the latter, each possible block is drawn exactly once (as opposed to random draws with replacement).

[28]Blocks were used in the related visualization in Figure 1.3.

control studies, where they are also called 'in-time placebos' (see Appendix A.1 and Abadie et al., 2015).

The general idea is to assign a placebo treatment to time periods where in fact no treatment has occurred, and compare the placebo treatment estimates to the true treatment estimates. This is informative about the likelihood that the treatment-estimate was caused by shocks that coincided with the treatment date, rather than the treatment itself. Assuming that no other rare event took place simultaneously that (strongly) affected hate crime, and under the stronger assumption that frequency and distribution of such shocks was constant, the p-value can be interpreted as probability of the treatment (Brexit vote) having a causal effect. To weaken the latter assumption, I use time, time squared, and month of the year dummies as controls. Following Freedman and Lane (1983; confirmed as appropriate in the comparative study of Winkler et al., 2014), I use the residuals of a regression of monthly crimes on the mentioned controls.

	(1)	(2)	(3)	(4)
	RR Hate Crime	Log(RR Hate Cr.)	RR Hate Crime	Log(RR Hate Cr.)
July 2016	550**	0.21**	784**	0.13**
	[373]	[0.07]	[399]	[0.02]
Post July 2016	-48	-0.05	20	0.01
Data	London & Manchester	London & Manchester	England & Wales	England & Wales
Observations	88	88	53	53

Note: Permutation inference on time series data (88 and 53 months). Lower consonance interval boundary of significant parameters in brackets (i.e. effect size at a 95% benchmark of the placebo treatments). Controls: Time, time squared, month of the year. Mean racial or religious hate crime for London & Manchester: 1251 (log: 7.09); for England & Wales: 3495 (log: 8.13). Using (flawed) classic robust standard errors, all July 2016 dummies are significant at the 1% level, and none of the post July 2016 dummies are significant.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.3: Overall Effect of the Brexit Vote on Hate Crime Considerable but Transitory

The results are presented in Table 1.3. There is a significant temporary, but no permanent effect. The point estimate for the temporary effect for London and Manchester is 550 racial or religious hate crimes in absolute terms and 21% in relative terms (a difference-in-difference estimation leads to a highly similar result, see Appendix A.2). The lowest possible p-values depend on the number of observations and are $\frac{1}{88}$ (achieved for London and Manchester), and $\frac{1}{53}$ respectively. While classical confidence intervals are not defined for permutation inference, 'consonance intervals' are (see e.g. Kempthorne & Folks, 1971). The lower bound of the consonance interval is obtained in the following way. First, the regression is run with all 87 (52 for England & Wales) placebo months and all 87 placebo-coefficients are saved for both the short term ("July 2016") and long term ("Post July 2016") effect. Next, the 95th percentile of the placebo-coefficients is computed and subtracted from the coefficients of the real regression. The resulting difference is the lower bound of the consonance interval. For the long-term effect, the effect of the real regression is not higher than the 95th percentile of the placebo estimates (and also not lower than the 5th percentile). Consequently, no lower bound of the consonance interval is reported in Table 1.3.

The effects are qualitatively alike in both datasets, i.e. both London and Manchester as well as England and Wales as a whole experience an increase in the short term but no long-term effect. The relative magnitude of the short-term effect appears to be larger in London and Manchester. However, the lower bound of the short term effect's consonance interval for London and Manchester is not higher than the point estimate for England and Wales. Either there is indeed no significant difference or the number of datapoints is too low to detect it. In section 1.6 (regarding the heterogeneity within London & Manchester), a key finding is the (short-term) effect to be more pronounced in areas with more recent immigrants. If this finding is extrapolated to all of England and Wales, it would make sense for the short-term effect to be more pronounced in London and Manchester as the share of recent immigrants is higher compared to the rest of England and Wales. In sum, while the subsequent heterogeneity analysis is limited to London and Manchester and external validity remains a concern, there is at least no obvious discrepancy coming from the times-series data of hate crimes.

1.4 Spatial Heterogeneity: Challenge, Measure, and Relevance

The previous section was already suggestive of the importance of spatial heterogeneity. This section measures it, shows the existence of relevant spatial heterogeneity that is captured by the candidate variables, and discusses the challenges that a standard OLS approach cannot address.

1.4.1 Standard Method and Challenges

If the model of heterogeneity is unknown, the standard OLS method is not applicable. It is not designed for model selection. The standard OLS method is illustrated in the following example. In the first step, the researcher chooses one or more candidate variables, say the remain-share in the Brexit vote. Only this choice allows for the estimation of a regression such as the following:

$$crime_{at} = \alpha + B_t\gamma + B_tR_a\beta + T_{at}\delta + A_a + e_{at} \tag{1.1}$$

In the above regression, B_t represents a dummy for the selected period after the Brexit vote (July 2016), A_a is the area fixed effect, R_a stands for the share of remain votes, T_{at} controls for time effects, and $crime_{at}$ is the number of racial or religious hate crimes per million of borough population per month. Due to the area fixed effect, the level of R_a does not enter the regression. The parameter of interest, β, shows how the post-vote period was differentially affecting hate crime depending on the remain vote share. Choosing months as time dimension (t) and boroughs as areas (a) allows me to use both the London and the GMP data. I moreover choose time, time squared, and month-of-the-year dummies as well as their interactions with the borough fixed effect as time controls T_{at}.[29]

[29]The parameter of interest is estimated to be highly significant at 2.85. The term is also economically large.

As any other candidate variable (R_a) could have been chosen by the researcher, a multiple hypothesis problem arises. For example, even absent of any true heterogeneity 5% of all variables will in expectation be significant for explaining heterogeneous treatment effects (when tested at the 5% level of significance). This means not only that a single researcher that tries to find a significant effect will be successful if sufficiently many variables are tested, but also the collective of researchers might be unknowingly doing just that.

Nevertheless, at least if there are only few possible variables, this is the standard method to test individual effects. Therefore, one possibility is to run a separate regression for each candidate variable and record each $\hat{\beta}$ with its p-value. I will use this as a benchmark to the subsequently introduced machine-learning-based methods. To alleviate the multiple hypothesis problem, I use the FDR correction method by Benjamini and Yekutieli (2001).[30] Controlling for borough-specific time effects helps to satisfy the parallel trends assumption for the various regressions. Assuming that it is not the case that both ε_{at} is serially correlated and the relevant lags are correlated with the candidate variable (e.g. R_a), these tests indicate which variables are individually correlated with the increase in hate crime after the Brexit vote.

However, this ignores the issue of model selection. Out of the 68 possible candidate variables, a researcher could choose a subset of variables (several variables of interest and/or controls). The number of possible models is larger than 10^{16}. Consequently, standard OLS methods cannot be adjusted or used to choose a specific model.

This paper builds on lasso-based methods in order to choose the most informative model in terms of prediction performance. A fully theoretically founded economic model is infeasible to use. It would need to be sufficiently established that it is credibly chosen ex-ante. No such model (or set of models) exists in the current setting. The most informative model is a feasible alternative. If the resulting model is approached from a specific economic perspective and a subset of the resulting variables is interpreted as controls, the implicit criterion for controls is to include those that are relevant in terms of predictive performance.

However, simply applying lasso to a version of regression 1.1, where R_a is replaced with all candidate variables, is hardly constructive. The standard lasso does not produce confidence intervals or p-values. Moreover, the path of the lasso to select variables follows the correlation with the dependent variable. Since $B_t R_a$ is zero in 87 out of 88 months, the correlation of this term with the dependent variable will be low. Therefore, any such interactions are almost guaranteed not to be part of the selected model. I propose instead to focus on the one month where this interaction is not zero.

Boroughs with a high remain vote have a remain vote share of around 60%. The estimated crime-increase for July 2016 in a 60%-remain borough is 60.7 hate crimes per million of population. Boroughs with a low remain vote have a remain-share of around 40%, which results in an estimated increase of only 3.8. Standard errors were clustered at the borough level. The regression table can be found in Appendix A.6.

[30] Early papers about multiple hypothesis testing focus on the family-wise error rate (FWER), for example the Bonferroni correction (see Dunn, 1961). However, with many potential hypotheses, the power of FWER methods becomes minuscule. False discovery rate (FDR) methods are often regarded as an improvement (see Austin et al., 2014). The standard method is the Benjamini and Hochberg (1995) FDR correction. However, this relies on the test statistics being non-negatively dependent of each other. This assumption is strong in the current setting due to the combination of the correlations among the candidate variables with the correlations of the candidate variables with the treatment effect. Benjamini and Yekutieli (2001) have developed a more conservative FDR adjustment that is also robust to negatively dependent test statistics.

1.4.2 Obtaining a Measure: Detrending, Deseasonalizing, and Demeaning

To measure the abnormal crime after the Brexit vote, I use the following regression:

$$crime_{at} = \alpha + B_t\gamma + T_{at}\delta + A_a + \varepsilon_{at} \tag{1.2}$$

Time controls (T_{at}), area (boroughs), period (months), and Brexit definition (B_t; July 2016) are equivalent to the specification above, but contrary to regression 1.1, no dimension of heterogeneity is included. Based on section 1.3, the treatment is assumed to be only present (or relevant) in one period: July 2016. I use the heterogeneity in $\varepsilon_{a,July16}$ as heterogeneity of interest.[31] Consequently, the analysis of this measure ($\varepsilon_{a,July16}$) is fully cross sectional.[32] To obtain a measure of abnormal crime, the constant $\hat{\gamma}$ is added, which is not relevant for the heterogeneity analysis.

The related question about the spatial heterogeneity in the relative instead of the absolute increase in racial or religious hate crime is addressed accordingly using regression 1.3:

$$log(crime_{at}) = \alpha^l + B_t\gamma^l + T_{at}\delta^l + A_a + \varepsilon^l_{at} \tag{1.3}$$

There are several other ways how the abnormal crime in July 2016 could be measured. Using all months but July 2016, all months but June, July, and August 2016, or only the months before July 2016 to then predict the (counterfactual) number of hate crimes per borough in July 2016 all result in similar or even identical models chosen by the lasso. The same is true for including the lagged dependent variable as additional regressor or instead of the time-squared regressor.[33] Details can be found in Appendix A.4. All of the above use a panel-OLS as predictor. Consequently, the maintained model assumption in this paper is how time trends, seasonalities, and borough fixed effects are controlled for. The remainder of the model is chosen by lasso-based methods (see section 1.5). The panel-OLS arguably adds reasonable structure to support out of sample predictions. A random forest, for example, performs inferiorly.[34]

A disadvantage of the first proposed (and used) measure is that it suffers from attenuation bias if only one heterogeneity-variable is used (see Lemma 1) and from unknown bias with

[31] The number of observations in this cross section (42 overall, reduced to 21 in the splitting approaches) is not uncommon in the relevant literature. In two key papers for the remainder of this paper, Lee et al. (2016) use simulations with 25 observations, and Tibshirani et al. (2016) such with 50 observations. Also Hebiri and Lederer (2013) use only 20 observations in their simulations. However, this finite number implies that seed dependency is a serious issue.

[32] It is not the case that the rest of the data is disregarded though. It is used to detect time trends, seasonalities, and borough fixed effects in regression 1.2, and moreover for permutation tests in section 1.4.4.

[33] A final approach would be not to generate a separate measure but instead use a partially penalized lasso in a version of regression 1.1 where R_a is replaced with all candidate variables and all control variables are not punished. The result is again similar. However, it is not obvious how such an approach can be paired with the CPSL method in section 1.5.1 (circumnavigating the problem via differentially weighted parameter standardization does not provide a satisfactory result). Given this disadvantage, I generally refrain from using this approach. The exception is section 1.6.3, which includes an illustrative comparison of the this approach with the here proposed measure.

[34] Using all months except for June, July, and August 2016, leaving out one additional month for growing the forest or running the panel OLS, the OLS produces a lower average MSE for predicting the left out month (done once for each month in the data).

more than one heterogeneity-variables. In retrospect, it would have been preferable to use one of the other ways to obtain the (counterfactual) number of hate crimes per borough in July 2016, for example those not using July 2016 or those not using June, July, and August 2016. Fortunately, the resulting models are the same in terms of the chosen variables (see Appendix A.4) and the simulation-results in section 1.5.2.2 are promising even with this current method.

The current method can be interpreted as panel-OLS results using a different (but closely related) dependent variables than regression 1.1: hate crime free from the part that is explained by the control variables in the least square sense. This is the case as all candidate variables I consider are constant in the period I study.[35] Consequently including a borough fixed effect makes including levels redundant. Again, the assumption is needed that it is not the case that both ε_{at} (ε^l_{at}) is serially correlated and the relevant lags are correlated with the candidate variables.

Lemma 1
Take the regression $y = x\beta + Z\gamma + \nu$. The OLS estimator $\hat{\beta}$ obtained from the regression $M_Z y = x\beta + \varepsilon$ is attenuated compared to that obtained from $y = x'\beta + Z\gamma + \nu$.[36]

Proof: see Appendix A.3.

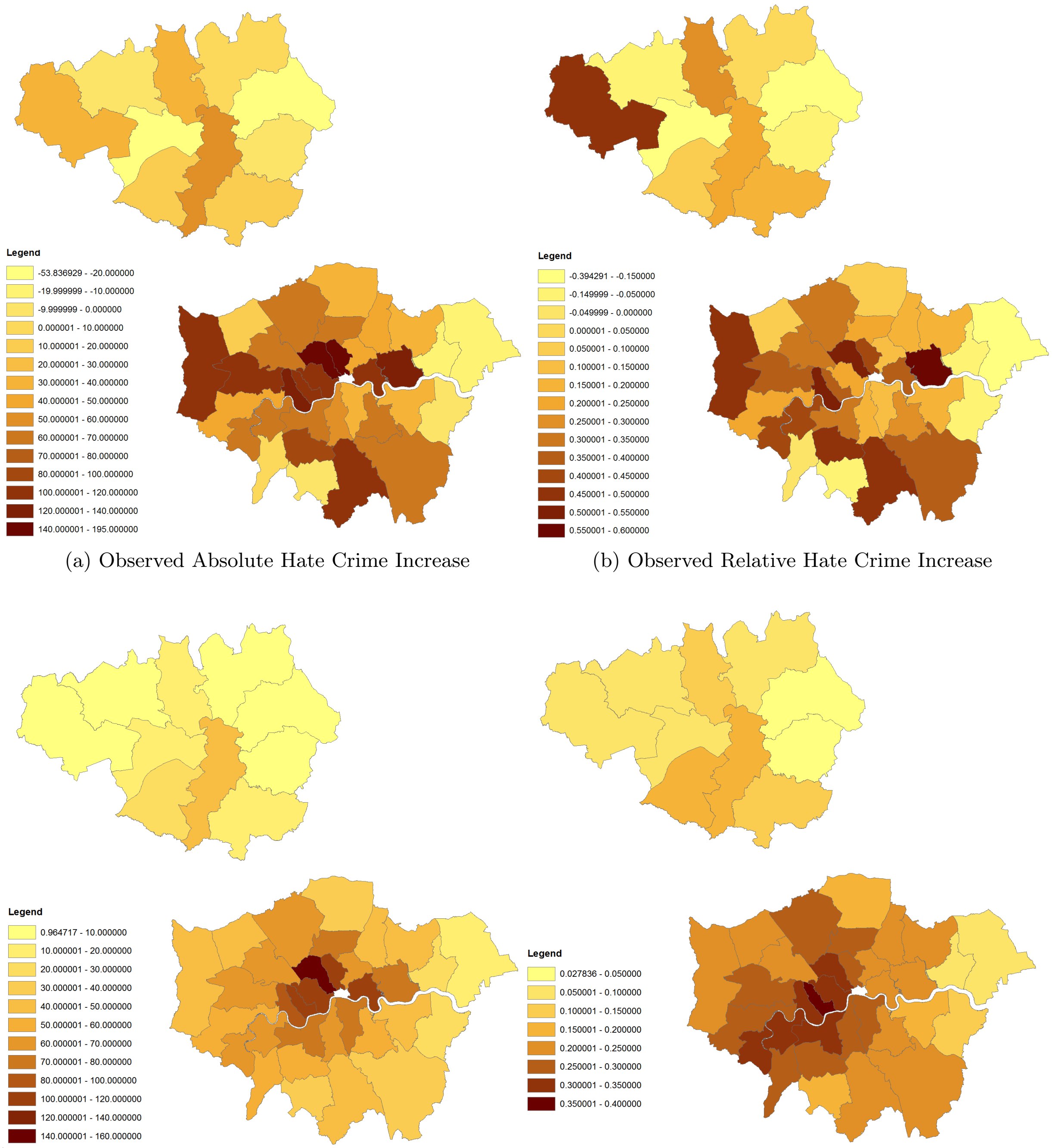
In the current setting, the interaction between a candidate variable and the treatment dummy take the role of x in Lemma 1 (provided there is only one candidate variable as Lemma 1 does not generalize to the multi-variable case), and all controls take the role of Z (i.e. all regressors in regression 1.2). I propose to only use the residuals of the Brexit period (July 2016). This is only a minor adjustment. All x entries in the above notation are 0 in any other period. Therefore, only the estimate of the intercept is potentially affected by this, which is of no concern in my setting. In section 1.6.3, both the here proposed measure as well as a standard regression (alike regression 1.1) are used. The differences are small (and insignificant).

1.4.3 Mapping the Spatial Heterogeneity

A first use of the measure is to obtain predictions at the borough level. This is directly relevant for policymakers and the police. However, plainly mapping $\varepsilon_{a,July16} + \hat{\gamma}$ and $\varepsilon^l_{a,July16} + \hat{\gamma}^l$ of the regressions 1.2 and 1.3 is problematic since the contained noise is displayed as much as the signal. In line with the remainder of the paper, I use a lasso with all candidate variables to visualize the predicted changes in racial or religious hate crime in July 2016 on maps. The cross validation used along with the lasso avoids over-fitting and hence extracts signal from the measure (to the extent that the candidate variables capture the spatial heterogeneity).

The results are shown in Figure 1.6. Two aspects stand out. First, the prediction quite visibly imposes a structure on the effect. While this might partly be due to missing additional prediction variables, it is also due to the extraction of signal from noise. Second, even in the

[35]The census data is decennial and the Brexit vote data unique.
[36]*As usual, $M_Z = I - Z(Z'Z)^{-1}Z'$.*



(a) Observed Absolute Hate Crime Increase

(b) Observed Relative Hate Crime Increase

(c) Predicted Absolute Hate Crime Increase

(d) Predicted Relative Hate Crime Increase

Note: Top: Greater Manchester. Bottom: Greater London. Increase refers to the difference in observed or predicted hate crimes per million of borough population versus the borough-level detrended, deseasonalized, and demeaned value. Method used for the predictions: Lasso with 69 candidate variables (vote and census data, plus a dummy for Manchester).

Figure 1.6: Borough-Level Hate Crime Increase in July 2016

predicted values, the heterogeneity is considerable. This heterogeneity is now analyzed more formally.

1.4.4 Significance of the Captured Heterogeneity

The spatial heterogeneity that is captured by the candidate variables is the fundamental basis for the models obtained in the following sections. This section shows the heterogeneity is significant and most of it can be attributed to the Brexit vote.

For a first illustration, I use a measure analogue to that proposed in section 1.4.2 but constructed for each month in the data (not only July 2016). Figure 1.7 depicts the variance of this measure across the 42 boroughs of Greater Manchester and Greater London for each month. In other words, regression 1.2 with B_t re-defined for each month in the data was run 88 times (for each month in the data) and the variance across boroughs of the residual for which $B_t = 1$ (i.e. its averaged squared value) is depicted in Figure 1.7. Striking is not only the magnitude of the spike after the Brexit vote, but also the lack of a comparable spike when racial or religious hate crime increased a year later (arguably) due to the terror attacks.

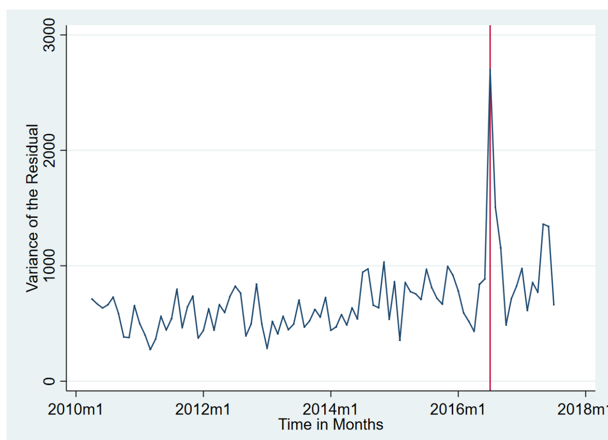


Figure 1.7: Variance in Detrended Deseasonalized Racial or Religious Hate Crime across Boroughs

To formally analyze the *captured* heterogeneity, I build directly on the approach proposed by Chernozhukov et al. (2018b) which only requires minimal assumptions. Their method uses propensity score matching which allows (asymptotically) for statements about the true underlying CATE. This is impossible in my setting. All regions were treated simultaneously and all candidate variables are time-constant. Consequently, the overlap condition, which is necessary for propensity score matching, fails. Instead, I will use a simpler version of Chernozhukov et al.'s (2018b) approach, using the previously introduced measure that encompasses the 'Brexit effect' but also noise in July 2016. Consequently, Chernozhukov et al. (2018b) denote such an approach as 'Naive Strategy' and 'not Quite Right'. However, I do not to stop at this point but use appropriate permutation tests to provide evidence that July 2016 leads to a tail outcome (compared to results using each other month as placebo). This provides strong evidence that the measure does contain relevant signal and that the Brexit vote is at least partly responsible for the observed heterogeneity. More specifically, permutation inference suggests that with 90% certainty, at least 71% of the captured heterogeneity is due to the Brexit vote.

The key technique used is repeated splitting. The data is randomly split in two equally large parts that Chernozhukov et al. (2018b) denotes 'auxiliary' and 'main sample'. In the auxiliary sample, any machine learning method can be used to produce predictors that will be used for the main sample.[37] Given the focus on lasso-based methods in this paper (consistent with the findings of Blair et al., 2017), I choose to use a simple lasso, using 3-fold[38] cross validation to determine the lasso penalty parameter commonly denoted λ.

Using the machine learning algorithm trained on the auxiliary sample, predicted treatment effects for the main sample are obtained. These are a function of the candidate variables and can be used to conduct the following analyses:

1) Divide the main sample in K bins according to the predicted treatment. The difference in hate crime between the top and bottom bin reflects the importance of the heterogeneity captured by the candidate variables. While Chernozhukov et al. (2018b) use K=5, I have a smaller sample and use 3 bins.[39]

2) Regress the dependent variable on a constant and the demeaned treatment prediction. The estimate of the latter is a second way to assess the importance of the heterogeneity that is captured by the candidate variables.[40]

To address seed dependency of the specific split used, the above is repeated 100 times. The median of the values is used as point estimate. Regarding the p-values, the median is used as well, but they are corrected by doubling them. While this guarantees that the median approach does not lead to a larger share of false positives, it is rather conservative. Further details can be found in Chernozhukov et al. (2018b).

Regarding the permutation tests, I exploit the fact that I observe 87 other months in the data. The above procedure is repeated another 87 times, using a regression analogue to regression 1.2 (regression 1.3 respectively), but changing the definition of the B_t dummy and choosing the respective residuals each time. As a result, 87 complete sets of placebo results are obtained.

[37] Chernozhukov et al. (2018b) outline a method to find the best method and tuning parameters to maximize the correlation between the true and the estimated proxy CATE. This is again infeasible in my setting. Chernozhukov et al. (2018b) stress that while using the "best" method is helpful, their approach only requires a useful proxy of a treatment prediction.

[38] 3-fold cross validation was chosen since after the split, 21 observations remain, so 3 is a divisor of the sample size. Moreover, this section focuses on tercile comparisons, so using 3-fold cross validation seems consistent.

[39] Chernozhukov et al. (2018b) also propose to compare the difference in means of all candidate variables of the observations in the top versus those in the bottom bin. The results of doing that can be found in Appendix A.9. The issue of multiple hypothesis testing due to having many variables is not resolved though. I therefore propose to again use Benjamini-Yekutieli (2001) FDR adjustments. The results of such a binning approach could be of policy interest and provides a robustness check to the method outlined in section 1.4.1. The splitting is useful as it is not advisable to bin using the true crime outcome that includes the error term (see Abadie et al., 2018, for more details).

[40] Since the measure proposed in section 1.4.2 is used, the constant should result in a zero estimate. This can be used as an additional check (successful in this paper).

The results of the first measure of heterogeneity are shown in Table 1.4. After sorting the boroughs into three bins (terciles) according to their predicted increase in crime (which is only a function of the candidate variables), it becomes visible that the top tercile has experienced a considerably higher increase in racial or religious hate crime (as measured by $\hat{\gamma} + \hat{\varepsilon}_{a,July2016}$ and $\hat{\gamma}^l + \hat{\varepsilon}^l_{a,July2016}$ from the regressions 1.2 and 1.3 respectively). For the top bin, the increase is significant at 92 crimes per population-million. For the bottom bin it is insignificant at 10. The difference is large and significant at 82 crimes per population-million. However, the significance here has two issues. First, it could be that there are some differences every month due to various other causes (which is again why this is 'not Quite Right', see above). Second, the significance here is based on 42 datapoints, of which only 21 are used in every iteration (to guarantee that the bins are indeed only based on a function of the candidate variables). De facto, however, there are 88 months times 21 (or 42) observations, which can exploited.

	(1)	(2)	(3)	(4)
	$\varepsilon(Hate\,Crimes\,per\,Pop.\,Mio.)$		$\varepsilon(log(Hate\,Crimes\,per\,Pop.\,Mio.))$	
	Top Tercile	Bottom Ter.	Top Tercile	Bottom Ter.
July 2016 Mean	91.8***	9.8	0.278***	0.142
Tercile Difference	81.99***		0.136	
Permutation Sign.	**		**	
Perm. 90% Benchmark	76.30 [93%]		0.12 [86%]	
Perm. 95% Benchmark	69.50 [85%]		0.09 [67%]	
Candidate Variables	69		69	
Observations	42		42	
Placebos (Perm. Test)	87		86	

Note: $\varepsilon(.)$ indicates that the borough-level dependent variable (racial or religious hate crimes per population million or the log of it) was first detrended, demeaned and deseasonalized using 88 months of data. 42 boroughs repeatedly sampled and classified to terciles. Terciles according to (out of sample) predicted values. Method used for the predictions: Lasso with 69 candidate variables (vote and census data, plus a dummy for Manchester). July 2016 indicates how in July 2016 (the month after Brexit), this value was higher than mean, trend, or season would suggest. Permutation inference uses other months than July 2016 as placebos. The benchmarking refers to subtracting the 90th/95th percentile of the placebo values. Percentages in brackets indicate how much of the heterogeneity is attributed to the Brexit vote if July 2016 had spatial noise equal to the 90th/95th percentile (using 88 months of data). The one month where one of the boroughs experienced 0 hate crimes was not used as a placebo for the relative case as the logarithm is not defined.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.4: Captured Heterogeneity Large and Mainly Attributable to the Brexit Vote

Permutation inference solves both these issues. Analogue to section 1.3.2, a lower bound of the consonance interval is obtained by subtracting the 90th (or 95th) percentile of the difference obtained in the 87 placebo-months from the difference obtained in July 2016. After this subtraction, 93% (85%) of the effect (i.e. of 82 crimes per population-million) is still there. This is considerable. It means that, assuming other forces affecting hate crime were within the first 9 deciles of their distribution, this 'Naive Strategy' (see above) results in a measure that is mainly due to the 'Brexit effect'. It provides the rational for basing this paper's heterogeneity analysis on this measure.

The mean number of monthly racial or religious hate crimes in the three years before the vote across all Greater London and Greater Manchester is 85 per population-million. The heterogeneity in Table 1.4 (with a difference of 82) is therefore large, which could be expected from the Figures 1.6 and 1.7.

Regarding the relative effect, the difference is not significant using the standard measure. While this is likely due to a lower heterogeneity, note that the difference is significant using the (here arguably superior) permutation inference - i.e. the heterogeneity in the relative effect is unusually large in July 2016 compared to the other months in the sample. The fact that 86% (67%) of the effect remains after subtracting the 90th (95th) percentile of the placebo measures indicates again that the difference is mainly due to the 'Brexit effect' (assuming no other uncommonly considerable other factor). Consequently, while the relative effect demands a more cautious treatment, an analysis of it appears justified. It is moreover the case that understanding this smaller relative heterogeneity is crucial given the absolute crime numbers underlying it.

The results regarding the second measure of heterogeneity are highly similar (see Appendix A.4).

1.5 Heterogeneity Models: Methodology

Having established that the spatial heterogeneity captured by the available candidate variables is substantial, this section outlines two methods to obtain parsimonious linear and quasi-linear prediction models to describe this heterogeneity. The selection of the candidate variables is lasso-based, but the aim is not only to obtain a model, but also estimate it with valid confidence intervals for the parameters.

1.5.1 Obtaining Linear Models: Conditional Post-Selection Lasso

The conditional post-selection lasso concept (see Lee et al., 2016; Tibshirani et al., 2016) was designed for valid inference after model selection (by lasso or lars[41]), but not specifically heterogeneous treatment effects. Due to the plain setup following the proposed measure (see section 1.4.2), however, it is directly applicable to the problem at hand.

In a first step, the model is selected using a regular lasso.[42] In a second step, the estimates are adjusted and confidence intervals are generated by conditioning on the fact that the chosen model was selected in that way. This is based on a truncated Gaussian test, which builds on the assumption of i.i.d. Gaussian errors. However, Tibshirani et al. (2018) show that asymptotically, the errors do not need to be Gaussian. They even report simulations with heteroskedastic errors where this approach still produces valid results.

[41] Least angle regression, see e.g. Efron et al. (2004).

[42] In principle, forward stepwise, lars, or lasso have been demonstrated to be feasible (Tibshirani et al., 2016). All of which are closely interlinked penalized regression methods (see Efron et al., 2004). Using the arguably most common type and following Lee et al. (2016), I choose lasso.

The first step implies choosing a penalty term, the hyperparameter commonly labeled λ.[43] I employ repeated 10-fold cross validation (following Kim, 2009),[44] and use the common rule to 'err on the side of parsimony' (Hastie et al., 2009), i.e. choose the most parsimonious model within one standard error.[45]

The main reason to use repeated cross validation is the overarching topic of (virtual) seed independency, which is violated by single 10-fold cross validation. Unlike the methods outlined in section 1.4.4 and 1.5.2, the conditional post-selection lasso approach does not involve any other form of repeated lasso-itarations. In line with the previous section, I also use repeated 3-fold cross validation (in Appendix A.4), which leads to virtually identical results in my setting. I always standardize the independent variables.

For the current setting, the performance of the lasso with correlated candidate variables is important. Hebiri and Lederer (2013) show that in such a case, penalty terms (λ) based on common theoretical considerations lead to suboptimal predictions, but not such based on cross validation (which is used throughout this paper). Another concern is the variable selection. Zhao and Yu (2006) outline that the irrepresentability condition must be satisfied for the true variables of the data generating process to be selected. This condition is almost certainly violated by the correlations at hand. In the current setting (and frequently in non-simulation cases), however, it is doubtful that the truly data generating variables are in my data. Therefore, the resulting estimated models in section 1.6.1 and 1.6.2 do not necessarily point to variables of the true data generating process, but instead to useful predictors. In order to still be able to compare the results across different methods, I have reduced the number of candidate variables by removing those that are the most correlated (in general above 0.9), while still arguably representing all relevant census categories.

1.5.2 Obtaining Linear & Quasi-Linear Models: Splitting Based Estimation

This section is inspired by Athey and Imbens (2017), Chernozhukov et al. (2018b), and Rinaldo et al. (2018). I propose an alternative approach which is more flexible than that outlined in section 1.5.1. Avoiding seed dependency comes at the cost of theoretically ambiguous bias. After developing the approach, I show the extent of this concern in simulations. Within my sample and the results I observe, the bias seems small and the coverage high.

[43] To answer the simpler question which single variable explains the heterogeneity the best, I also choose another λ such that the number of chosen regressors is one. Theoretically, the lasso approach can select and drop variables. If only one variable is selected though, it is impossible that a variable is dropped for one other variable (only a combination of variables can make the first variable redundant). Consequently, this will be a uniquely defined variable, which is the one for which the correlation with the dependent variable is the highest. The results of this exercise are displayed in Appendix A.9.

[44] Specifically, I use 1000 repetitions, so around 10% of the possible splits are used, which is optimal in simulations by Kuhn & Johnson (2014) (although the difference to using e.g. 100 repetitions is marginal).

[45] While this is standard, it makes little sense to do so in the other lasso applications in this paper. In case of the previous section, the model never becomes visible but only its prediction is used. As outlined in the next section, in case of the repeated splitting, there is already a tendency to shorter models due to the selection procedure, hence adding more parsimony does not seem warranted. In Appendix A.4, I also list the results from using the plain minimum of the cross validation (i.e. discarding this rule for parsimony).

1.5.2.1 Empirical Approach

The key concept of this approach is the following. First, the sample is split into two halves. Second, one half is used for model selection using lasso. Third, the other half is used to estimate the model resulting from the previous step. Fourth, the first three steps are repeated 1000 times. Finally, the (estimated) models of the 1000 iterations are aggregated to a single model with parameter estimates.

The approach combines repeated splitting similar to section 1.4.4 with the logic of the 'honest' approach developed by Athey and Imbens (2016, 2017). Contrary to Athey and Imbens (2016) though, I assume a (quasi) linear structure and use lasso, not regression trees. Intuitively, the lasso selection in one half of the data provides a model to which I can commit in the other half. This is related to the ex-ante commitment through a pre-analysis plan (see e.g. Olken, 2015).

Regarding model selection, I use two specifications. In the first, I use standard lasso (on standardized data) to find a linear model just like with CPSL in the previous section.[46] In the second specification, I use the hierarchical lasso approach of Bien et al. (2013) to find a quasi-linear model. This approach allows me to include interactions between candidate variables in the model, but guarantees that the level of each chosen interaction-variable is also included, allowing the result to be interpretable. This second specification is not only interesting for the current application but also shows that the splitting based estimation is more flexible in the model-selection part than standard CPSL (which focuses on standard lasso). Aside from hierarchical lasso, several other model selection techniques are imaginable for the splitting based estimation.[47]

Regarding estimating the model, I use OLS. This means not only following most of the standard economics literature, but also that the common standard error corrections are possible (contrary to the CPSL approach). In other applications, other estimation techniques might be more appropriate and the splitting based estimation approach is flexible in the sense that other techniques can simply replace the OLS estimation.

Omitting the last two steps and splitting the sample only once delivers consistent and unbiased estimates. However, then the approach is seed dependent. Consequently, the last two steps are necessary since seed dependency matters (see Appendix A.7) and since is impossible for a single researcher, and especially for the profession of social sciences as a whole, to commit ex post to a specific random sample split.

The last step involves aggregating 1000 different models.[48] Rinaldo et al. (2018) point out that a theoretical foundation to aggregate different models is missing in the current literature. I suggest the following procedure (inspired by Chernozhukov et al., 2018b). Select the most

[46] I employ 10-fold cross validation for choosing the hyperparameter λ, reporting the 3-fold cross validation results in Appendix A.4. The exception are the computationally intensive hierarchical lasso models that allow for interactions and squared terms. For those I only use the computationally less intensive 3-fold cross validation.

[47] The recent One Covariate at a Time Multiple Testing (OCMT) procedure (see Chidik et al., 2018), or the Leave-Out-COvariates (LOCO) approach (see Rinaldo et al., 2018), for example, could be interesting alternatives.

[48] Note that this is different in section 1.4.4, where each iteration produces estimates for the same parameters.

common model (subject to constraints, see below) and use the median estimates of those iterations that resulted in the selected model. Contrary to Chernozhukov et al. (2018b), I refrain from doubling the p-values. This is rather conservative and the subsequent simulations do not indicate that it would be required for this method and setting.

While this procedure virtually eliminates seed dependency (see Appendix A.7), and to some extent the common criticism of sample splitting methods that half of the data is not used for inference, it introduces bias. Since all splits influence the choice which model is selected (and hence which splits are generating the final estimates), there is no longer complete independence between the sub-sample used for model selection and the sub-sample used for inference. The resulting net bias remains unclear. Intuitively, there are both forces to amplify and attenuate the estimators.[49] It is a key concern what kind of net-bias can be expected for the specific data, questions, and results at hand. The simulations in the subsequent section are designed to address this concern.

Another use of the simulations is to compare constraints regarding selecting the most common model across the 1000 iterations. The constraint concerns the length of the models. Longer models have the issue of featuring more possible combinations of variables. Especially since my candidate variables are correlated, the probability that a one-variable model is chosen multiple times is much higher than the probability that the exact same five-variable model (for example) is chosen multiple times. Choosing merely similar but not identical (long) models in the last step of the splitting based estimation bears the problem that there is no obvious choice within those models and estimators. This is the reason why I propose choosing the most common model across the 1000 iterations subject to a constraint. The constraint is that the model has to include at least L variables. I allow L to take the following values: one, two, the floor of the mean of the model lengths minus one standard deviation, and the respective ceiling.[50] The simulations in the following section are consulted to make a choice that is appropriate for the specific data, questions, and results at hand.

[49] On the one hand, there is at least one force to attenuate the true effect. If the sample is split such that many of the observations that are driving the effect are in the sub-sample for model selection, the model is safe to be selected but they are lacking in the sub-sample where inference is conducted, resulting in attenuated estimates. Conversely, if these observations are in the sub-sample for inference, a different model is likely to be chosen in the model-selection stage, resulting in these splits being dropped. On the other hand, there is at least one countering force. In the sample as a whole, those variables are chosen which have the highest correlation due to signal plus noise. Therefore, such noise that produces amplification bias is prone to be found in the chosen estimators.

In addition, the confidence intervalls might also be incorrect. One reason for that is that they are standard OLS intervals, based on half of the observations (inference sample only). The fact that other splits of the data resulted in the same variable(s) being correlated with the outcome variable implies that, de facto, more than half of the observations were used for the estimates. Consequently, the confidence intervals would be too large.

[50] As in section 1.5.1, I also show which single variable explains the heterogeneity best. This collapses to finding the candidate variable that is most correlated with the dependent variable in the mining sample. Comparing this to the equivalent question in section 1.5.1, the advantage of this approach is to be potentially less affected by outliers (a general concern of standard econometric methods, see Young, 2018). However, this approach has the downside of potential bias (relying on simulation evidence). The results are shown in Appendix A.9.

1.5.2.2 Simulations

I construct the dependent variable in the following way: $y = X\beta + \epsilon$. The true effects (β) and truly relevant regressors (X) vary for the different simulations, while the noise parameter ϵ is always a random variable with a standard Normal distribution.[51] The number of relevant regressors and the magnitude of the true effects is motivated both by simulations in the related literature (namely Lee et al., 2016; Tibshirani et al., 2016; Reid et al., 2017), as well as the characteristics of the results obtained in section 1.6.[52]

Simulations are a useful tool in this context to analyze whether (1) the true model is found (which is ambitious since the irrepresentability condition is violated), and (2) whether the obtained $\hat{\beta}$ is consistent for $\widetilde{\beta}$, where $\widetilde{\beta} = argmin(y - X\beta + \epsilon)^2 \; s.t. \; \left[\beta = \hat{\beta} \; if \; \hat{\beta} = 0\right]$; i.e. whether the conditional correlations are measured consistently conditional on using the model that is "found".

As shown in Table 1.5, I consider the following data generating processes (DGP): $y = 2X_1 + \epsilon$; $y = 2X_2 + 2X_3 + \epsilon$; $y = 2X_1 + X_4 + X_5 + \epsilon$; $y = 2X_1 + 2X_3 + 2X_6 + 2X_7 + \epsilon$. Further complementary simulations that concern only the best single variable (instead of the best model) are shown in Appendix A.6 and also use $y = \epsilon$; $y = X_1 + \epsilon$.

I use the real dataset to guarantee that the correlation structure of the candidate variables is identical to that in the actual regressions in section 1.6. In order to have the most relevant correlation structure for the results, an important part of the simulations considers effects driven by those regressors that are detected in section 1.6 and such that are highly correlated with them. Specifically, variable X_1 is chosen rather clearly in section 1.6. The variables X_2 and X_3 are highly correlated with X_1 and with each other. They are included to test to what extent this is a concern. The third DGP follows again the results obtained in section 1.6. The final DGP contains again X_1, as well as one highly correlated (X_3) and two random other variables. The common magnitude of the β in the true model (signal strength) of two follows the aforementioned literature and the results in section 1.6.

As outlined in section 1.5.2.1, it is possible to restrict the model length (L). I compare the following restrictions: $L \geq 1$ ("1Var"), $L \geq 2$ ("2Var"), $L \geq \lfloor mean_N(Z) - \sqrt{var_N(Z)} \rfloor$ ("Floor"), and $L \geq \lceil mean_N(Z) - \sqrt{var_N(Z)} \rceil$ ("Ceiling"). Indexing by N denotes the mean/variance across the $N = 1000$ iterations of obtaining the splitting estimator, not across the repetitions of this process in the $S = 1000$ simulation iterations.

[51] As shown in Appendix A.6, the residual from the regressions 1.2 and 1.3 appear approximately normally distributed (to the extent this can be judged from 42 observations).

[52] My setting consists of 42 observations and 68 vote and census variables (in total 69 candidate variables if a dummy for Greater Manchester is added). Lee et al. (2016) simulate with 25 observations and 50 candidate variables, and choose 5 variables to carry signal. Their case is slightly different, but in their setting, the signal is 2. The simulations in Tibshirani et al. (2016) have 50 observations, 100 candidate variables and 2 truly active variables. Again the signal strength is not directly comparable, but some of their simulations have signals of up to 5. Reid et al. (2017) outline a more general approach. They propose using n^α truly active variables, with α taking values between 0.1 and 0.5 (i.e. between 1 and 6 variables for 42 observations). They take 1000 observations and suggest using signal strengths between 3 and 6. The characteristics of the simulation results (not the parameters, but the frequency distribution within the 1000 splits of one iteration of my proposed estimator) start to diverge strongly from the characteristics of the results in section 1.6 when many variables with strong signals are introduced. In the interest of proximity to my setting, I remain on the lower end in terms of number of variables and signal strength compared to the mentioned literature.

True DGP	Constraint	Bias: Pos Param.	Bias: Neg Param.	Param. in CI	Length	Freq.	True DGP	Int. Only
$2X_1$	1 Var	-8.1% of CI Length	NA	97.6%	2.3	1.55%	99.7%	0
$2X_1$	2 Var	-8.7% of CI Length	NA	97.9%	2.7	1.20%	99.6%	0
$2X_1$	Floor	-6.4% of CI Length	NA	99.0%	4.1	0.97%	97.6%	0
$2X_1$	Ceiling	-4.6% of CI Length	NA	99.5%	5.2	0.53%	96.1%	0
$2X_2+2X_3$	1 Var	-5.9% of CI Length	NA	98.9%	4.0	0.95%	99.9%	0
$2X_2+2X_3$	2 Var	-5.9% of CI Length	NA	98.9%	4.0	0.95%	99.9%	0
$2X_2+2X_3$	Floor	-4.3% of CI Length	NA	99.1%	5.4	0.65%	99.8%	0
$2X_2+2X_3$	Ceiling	-2.6% of CI Length	NA	99.3%	6.3	0.40%	99.5%	0
$2X_1+X_4+X_5$	1 Var	-0.2% of CI Length	-4.3% of CI Length	99.7%	3.8	0.66%	56.9%	1.6%
$2X_1+X_4+X_5$	2 Var	-0.3% of CI Length	-4.4% of CI Length	99.6%	3.9	0.65%	58.1%	1.6%
$2X_1+X_4+X_5$	Floor	-0.6% of CI Length	-1.4% of CI Length	99.6%	7.1	0.33%	59.9%	1.6%
$2X_1+X_4+X_5$	Ceiling	-0.2% of CI Length	-0.8% of CI Length	99.0%	8.5	0.22%	58.2%	1.6%
$2X_1+2X_3+2X_6+2X_7$	1 Var	-0.1% of CI Length	-0.5% of CI Length	99.6%	11.2	0.15%	79.3%	0
$2X_1+2X_3+2X_6+2X_7$	2 Var	-0.1% of CI Length	-0.5% of CI Length	99.6%	11.2	0.15%	79.3%	0
$2X_1+2X_3+2X_6+2X_7$	Floor	0.6% of CI Length	-0.1% of CI Length	99.3%	13.2	0.12%	78.1%	0
$2X_1+2X_3+2X_6+2X_7$	Ceiling	0.7% of CI Length	-0.1% of CI Length	99.4%	13.8	0.11%	76.1%	0

Note: Each simulated 1000 times. X_1: share of recent immigrants; X_2: share of remain votes in the Brexit referendum; X_3: share of people in a same-sex civil partnership; X_4: share of people with no qualifications; X_5: share of people not stating a religion; X_6: share of people working from home; X_7: share of people working in the information and communication industry (industry code J). Constraint refers to the imposed minimum number of variables in the model (floor: the floor of the mean model length minus one standard deviation; ceiling accordingly). Int. Only refers to the share of models that were intercept-only, i.e. no predictor was chosen. Possible predictors: the 68 candidate variables from the census and vote data. Heteroskedasticity robust errors used. If instead i.i.d. Normal errors are used, the bias increases slightly (the absolute magnitude remains $<$ 11% of the CI length in any case) and the coverage is closer to 95% (92% to 98%). The results in section 1.6 are robust to either case.

Table 1.5: Simulation Results of the Splitting Estimation: Bias Small, Coverage High

Table 1.5 summarizes the results for different DGPs and constraints. While I generally average biases across both models and regressors, I make the distinction between biases of positive and negative parameters.[53] Due to that, column 3 and 4 are informative about the overall bias being potentially attenuating or amplifying. Column 5 indicates how often $\widetilde{\beta}$ is contained in the 95% interval of $\hat{\beta}$, and column 6 states $mean_S(L_{\hat{\beta}})$. Column 7 shows $mean_S(\frac{C}{1000})$, where C is the number of times the most frequent model (subject to the relevant restriction) was chosen within each simulation iteration. As the true DGP becomes longer, the mean frequency approaches 0.1%, i.e. C=1. This is problematic for the key argument of seed independency (and highlights a limitation of the general applicability of the splitting based approach). The eighth column indicates what fraction of the true DGP is on average encompassed in the selected approximation model, i.e. what fraction of the non-zero elements of β are also non-zero in $\hat{\beta}$.[54] As mentioned above, this is not necessarily expected to be large due to the violated irrepresentability condition. Finally, the last column indicates how often the weakest constraint ($L \geq 1$) was binding across the simulation iterations.

The results show that the suggested approach is likely to be applicable to the current setting. The magnitude of the bias is rather low, and more than 95% of the truly relevant parameters are contained in the 95%-confidence intervals (around 95% if heteroskedasticity robust S.E. are not used). If anything, the confidence intervals seem too conservative. Between the constraints regarding the model selection, that requiring at least one variable seems to have the most desirable attributes.

1.6 Heterogeneity Models: Results

Both approaches outlined in the previous sections lead to the same main result. It is boroughs with a wealthy and/or immigration-heavy population that have experienced a more pronounced increase in hate crime. The borough-share of recent immigrants (for the absolute 'Brexit effect'), and that of people with formal qualifications (for the relative 'Brexit effect') are of key predictive importance. Models that allow for interactions generally support this result. The model regarding the absolute 'Brexit effect' even remains unchanged when including interactions would be possible.

1.6.1 Linear Heterogeneity Models

The results of finding linear (prediction) models of the post-vote increase in hate crime are summarized in Table 1.6. Panel A reports the result regarding the absolute effect, and panel B that regarding the relative effect. The two approaches outlined in the sections 1.5.1 and 1.5.2 arrive at highly similar results. As a benchmark, the standard lasso is unsurprisingly

[53] Positive parameters are such for which the mean of the parameter of 1000 repeated OLS regressions of the relevant approximation model with 42 observations randomly drawn from the DGP's $\widetilde{\beta}$ is larger than 0.01. Negative parameters are smaller than -0.01 respectively. Other forms of grouping are imaginable. Given the homogeneity in the type of candidate variables and the concern for amplification versus attenuation bias, I believe this grouping in positive and negative parameters to be reasonable and informative.

[54] In the third model, for example, X_4 is almost always absent in the chosen models while the other two 'true' regressors are usually included.

considerably lower since the lasso penalty term (λ) leads to attenuation bias.[55] Calculating standard errors is generally problematic for the standard lasso (see e.g. Kyung et al., 2010). I refrain from doing so as the lasso estimates are solely displayed as a comparison.

Regarding the result for the absolute increase, the chosen model contains the share of recent immigrants, the share of people that do not state a religion, and the share of people with no[56] formal qualifications.[57] The share of recent immigrants is significant and refers to immigrants that have arrived within two years before participating in the 2011 census. Further data provides strong evidence that (recent) immigration areas are virtually the same in 2011 and 2016.[58] The parameter regarding the share of people without any qualifications is not significant; the p-values are at 0.17 and 0.11 for column 1 and 2 respectively. Finally, the share of people not stating a religion (which is different from people stating not to have a religion) appears to be informative as well. However, the asterisks in parentheses indicate that the significance disappears once heteroskedasticity robust errors are used (which is possible with the splitting based approach, but not CPSL).[59]

Following the previous simulations of the splitting based approach, I chose the model-selection constraint to include at least one variable (column 2). This constraint is not binding (the intercept-only model is selected only twice out of 1000 splits, not shown in Table 1.6).[60] As indicated in the last row of panel A, the resulting model was chosen 16 out of the 1000 splits. The fact that this number is not higher illustrates the problem of the lasso to obtain different models in only slightly different settings. However, the approach presented in this paper to use 1000 splits (i.e. similar settings) and then to aggregate alleviates this problem to some extent. This problem is related to the issue of seed-dependency, in the sense that different splits result in different but similar settings. As shown in Appendix A.7, running lasso on a single split (i.e. one setting) reaches dramatically different models across different splits. Aggregating it over 1000 splits leads to a considerable improvement in consistency. It is also the case that a frequency of 1.6% is similar to the successful simulation in section 1.5.2.2.

[55] This is especially true for the relative case. Because a one-variable model was chosen, the penalty parameter λ is arguably more important, and the lasso estimates suffer from more attenuation bias (which appears to be the dominant bias here).

[56] This is the only direct measure of qualifications I include (i.e. no differentiation between different qualifications is included; see footnote 20).

[57] Results using different forms of cross validation are qualitatively similar, especially regarding the importance of recent immigrants, and can be found in Appendix A.4.

[58] The Office for National Statistics (2018b) provides some annual immigration data. The correlation of the borough-share of immigrants in 2011 with that of 2016 is 0.97 for Greater London and Greater Manchester. Direct annual data on recent immigrants is not available. However, the numbers of migrants that first register with a general practitioner (relative to the borough population) in 2011 and 2016 are correlated with a coefficient of 0.94. Moreover, the numbers of migrants registering in their borough to obtain a national insurance number (necessary to work) in 2011 and 2016 (relative to the respective borough population) are correlated with a coefficient of 0.97.

[59] In general, the effect of using HC errors is small, this being the only parameter whose significance is affected by it. The reason why significance in parentheses is used is to illustrate that the two methods obtain the same result, even regarding (qualitative) significance, under the same conditions (i.e. not correcting for heteroskedasticity).

[60] In fact, even the ‘2 variables’- and the ‘floor’-constraint were not binding, and hence result in the identical model. The ‘ceiling’-constraint did bind and leads to additionally including the share of people working in electricity, gas, steam and air conditioning supply (industry code D). This results in the additional coefficient being small with a large standard error (highly insignificant), and the other coefficients being hardly affected.

	(1)	(2)	(3)	Indep. Var. Means
		A) $\varepsilon(Hate\,Crimes\,per\,Pop.\,Mio.)$		
Recent Immigrants	1237***	1150*	977	0.024
No Religion Stated	641***	715$^{(**)}$	405	0.080
No Qualifications	-235	-274	-114	0.194
Mean H.C. / Pop. Mio.	219	219	219	
Frequency	NA	1.6%	NA	
		B) $\varepsilon(log(Hate\,Crimes\,per\,Pop.\,Mio.))$		
No Qualifications	-2.55*	-2.67***	-0.37	0.194
Mean Log(H.C. / Pop. Mio.)	4.57	4.57	4.57	
Frequency	-	2.0%	-	
Method	CPSL	Splitting	Lasso	
Candidate Var.	69	68	69	
Observations	42	42	42	

Note: Method-chosen models from 68(69) candidate variables. $\varepsilon(.)$ indicates the dependent variable is detrended, deseasonalized, and demeaned on the borough level using 88 months of data. Cross sectional analysis across 42 boroughs in July 2016 (the month after Brexit). Hate Crimes (H.C.) is short for racial or religious hate crimes. Recent Immigrants: share of people that have arrived in the UK within 2 years of the 2011 census. No Religion Stated: share of people not stating a religion. No Qualifications: share of people without formal qualifications. CPSL assumes i.i.d. errors by construction. Splitting allows for robust errors, parentheses indicate lost significance due to using heteroskedasticity robust errors. Significance not defined for plain lasso (which serves as benchmark only). Splitting cannot use the Manchester dummy variable as candidate since it has a constant value (0) for more than half of the sample.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (for CPSL and Splitting)

Table 1.6: Best Linear Model for the Absolute/Relative Increase in Hate Crime

Finally, the model is the same as the one resulting from CPSL. The fact that the lasso agrees with the conditional post-selection lasso (CPSL) in terms of selected variables is guaranteed by construction.

Another benchmark to the two approaches outlined in this paper is the single split estimator. This method is alike Athey and Imbens' (2016) causal tree, but uses a lasso in half of the sample instead of growing a tree (as suggested in passing by Athey and Imbens, 2017). As discussed, the issue with that is its seed dependency and the here proposed splitting based approach is stable across different seeds (dropping and/or adding at most one insignificant variable in case of the full model as only main difference, see Appendix A.7).

In terms of magnitude, the result shows that boroughs that have one percentage point more recent immigrants experienced a 'Brexit effect' that was 12 racial or religious hate crimes higher per million borough-population in July 2016. The fraction of recent immigrants differs across boroughs: the fraction is 0.7% at the 25$^{\text{th}}$ percentile and at the 75$^{\text{th}}$, it is 3.3%.[61] The predicted difference in the 'Brexit effect' between the 25$^{\text{th}}$ and the 75$^{\text{th}}$ percentile is therefore approximately 31 crimes per million of population, controlling for the other two variables in the model.[62] The average number of racial or religious hate crimes per million in the three years before the Brexit vote ranges across boroughs from 49 to 278, the mean across boroughs being 117. As aforementioned, the mean across Greater London and Greater Manchester overall is 85. The estimate of the July 2016 dummy in regression 1.2 is 52 (see Appendix A.6 for regression output).

Regarding the result for the relative increase, the chosen model contains only the share of people with no formal qualifications.[63] The constraint to include at least one variable in the splitting based method was binding. The low variance across boroughs caused the splitting based approach to choose the intercept-only model in 225 of the 1000 splits. As outlined in section 1.4.4, the spatial heterogeneity is smaller in the relative than the absolute case. This provides evidence that the areas where most of the 'Brexit effect' has occurred in absolute terms are also areas with a generally high number of hate crimes. Indeed, running a regression with no borough fixed effect but instead the selected candidate variables (listed in Table 1.6) on the pre-Brexit-vote period demonstrates that the fraction of recent immigrants is indeed highly correlated with the number of racial or religious hate crime (see Appendix A.6). In that regression, the coefficient for the share of people without formal qualifications positive though. This is in line with the findings about the relative 'Brexit effect'.

Regarding the magnitude of the effect, moving across boroughs from the 75$^{\text{th}}$ to the 25$^{\text{th}}$ percentile of the share of people without formal qualifications, the increase in racial or religious hate crime is around 18 percentage points higher. As a benchmark, the average increase across boroughs estimated with the July 2016 dummy in regression 1.3 is 20.6% (see Appendix A.6).

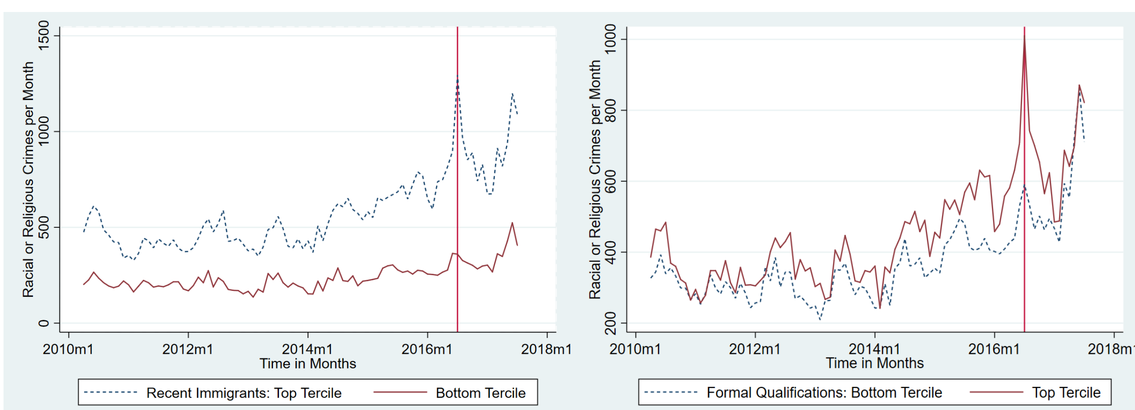


Figure 1.8: Heterogeneity in the Brexit Effect - Terror Spikes in 2017 Different from Brexit

The findings are visualized in Figure 1.8. There are considerably more racial or religious

[61] The 25th percentile of the fraction of people with no qualifications is 16.3%, and that of those not stating a religion is 6.4%. The 75th percentile is 8.5% and 23.0% respectively.

[62] This is already substantial, and without controls it is almost twice as large (see Appendix A.9 for the result of the best single variable).

[63] Single split comparisons, comparisons using different forms of cross validations, and results for the splitting selection rule imposing at least two (not one) variables can be found in Appendix A.4 and A.7.

hate crimes in the top tercile of boroughs with respect to share of recent immigrants, and the spike after the Brexit vote is pronounced. In the bottom tercile, that spike is virtually absent. The subsequent spike around the terror attacks indicates that it is hardly the case that there was a lack of opportunities or victims in low-immigration boroughs - and also that the heterogeneity regarding the increase of hate crime after terror attacks is different. Comparing the top and bottom tercile of boroughs with respect to the fraction of qualified people shows that while the numbers of crimes before and after the Brexit vote are comparable, the 'Brexit effect' is dramatically different.

As mentioned previously, these findings are regarding a parsimonious prediction model. A number of observed and unobserved factors are likely to drive the true data generating process. While no included factor is a better single variable explanation of the absolute and relative effect respectively, a combination of variables can certainly not be excluded to have an effect; especially highly correlated ones. The candidate variables that are most correlated with the variables chosen in the above models are listed in Appendix A.8. The lack of formal qualifications, for example, is highly correlated with indicators of the social grade.

1.6.2 Quasi-Linear Heterogeneity Models

As outlined in section 1.5.2, I also use the hierarchical lasso method of Bien et al. (2013) in the mining part of the splitting based estimation in order to obtain a model that is allowed to include well interpretable interactions. As this increases the number of possible models to an even higher number, I have increased the number of iterations to 5000. The results are shown in Table 1.7.

The key insight from these results is that the same linear prediction model as in the previous section is obtained in the absolute case. This is striking given that with (hierarchical) interactions, there are now more than 10^{49} possible model choices; more than half the estimated number of atoms in the known universe. The restriction to have at least one variable is always binding. The model of the relative effect is different than before. The share of people without formal qualifications is still a key predictor, but its effect seems to be different for different boroughs. The minimum fraction of male people is 48.0%, the maximum 52.1%; but given the large confidence intervals of the parameters, this results in little insight. In sum, Table 1.7 provides further robustness-evidence regarding the absolute model and further reason for a cautious interpretation of the relative effect (but support for the share of people with formal qualifications being important).

1.6.3 Individually Interacted Candidate Variables

Using each candidate variable individually implies a model with no control variables. This tests the simple null hypothesis that, considered individually, the given candidate variable is correlated with the increase in hate crime after the Brexit vote. It complements the previous

	(1)	(2)	Indep. Var. Means
	A) $\varepsilon(Hate\,Crimes\,per\,Pop.\,Mio.)$		
Recent Immigrants	1426*	1309*	0.024
No Religion Stated	595	641	0.080
No Qualifications	-214	-211	0.194
Mean H.C. / Pop. Mio.	219	219	
Frequency	0.8%	1.0%	
	B) $\varepsilon(log(Hate\,Crimes\,per\,Pop.\,Mio.))$		
No Qualifications	20.69	3.50	0.194
Male	8.64	-2.46	0.493
Male * No Qualifications	-48.47	-6.14***	0.096
Mean Log(H.C. / Pop. Mio.)	4.57	4.57	
Frequency	1.2%	1.1%	
Square Terms Allowed	Yes	No	
Candidate Var.	68	68	
Observations	42	42	

Note: Method-chosen models from 68 candidate variables, their first order interactions, and, where indicated, their squared values. $\varepsilon(.)$ indicates the dependent variable is detrended, deseasonalized, and demeaned on the borough level using 88 months of data. Cross sectional analysis across 42 boroughs in July 2016 (the month after Brexit). Hate Crimes (H.C.) is short for racial or religious hate crimes. Recent Immigrants: share of people that have arrived in the UK within 2 years of the 2011 census. No Religion Stated: share of people not stating a religion. No Qualifications: share of people without formal qualifications. Heteroskedasticity robust errors used.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: Best Quasi-Linear Model with Interactions using Splitting Based Estimation

results and shows the link to approaches that are more standard. Moreover, it allows me to use both the measure proposed in section 1.4.2 and the standard approach.

To assess the economic significance and compare the results to the previous findings, I weigh the results by a measure of variance. In line with the other sections of this paper, I use the difference between the mean of the highest tercile of a given variable across the 42 boroughs and the mean in the lowest tercile. I refer to it as tercile span. The estimated coefficient is then multiplied with this tercile span. The Tables 1.8 and 1.9 show the 5 variables with the highest effect as well as interesting cases where the null hypothesis could not be rejected.[64] Regarding the p-values, I follow the procedure described in section 1.4.1 and use Benjamini-Yekutieli (2001) FDR adjustments.

[64]A complete list of all candidate variables can be found in Appendix A.9.
[65]Skilled working class: Main income from skilled manual work
[66]Water supply, sewerage, waste management and remediation activities

Variable	Estimate * Tercile Span	Variable	Estimate * Tercile Span
Recent Immigrants	85.0***	Recent Immigrants	99.1***
Social Grade C2[65]	-77.8***	Social Grade C2	-90.8***
Mixed Ethnicity	77.1***	Mixed Ethnicity	90.0***
Industry Code E[66]	-76.8***	Industry Code E	-89.6***
Born in the UK	-76.5***	Born in the UK	-89.2***
Econ. Active	11.8	Econ. Active	13.7
Unemployed	1.1	Unemployed	1.3
Using Residuals (ε)		*Using Full Regression*	

Note: Estimate refers to the individual estimated effect of the variable on racial or religious hate crime per borough population million. Tercile span is the difference between the mean of the highest tercile of a given variable across the 42 boroughs and the mean of the lowest tercile. Using residuals denotes a cross sectional analysis across 42 boroughs in July 2016 (the month after Brexit) with the dependent variable detrended, deseasonalized, and demeaned at the borough level using 88 months of data. Full regression denotes directly using panel data of 42 boroughs and 88 months and interacting the variable with a dummy for July 2016. Benjamini-Yekutieli (2001) FDR adjusted p values.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.8: Top 5 Individually Important Variables & Unemployment (Absolute Increase)

I conduct this approach both following regression 1.1 (replacing the remain vote share with each candidate variable), and also regressing the residual from regression 1.2 on the candidate variable (and the respective analogues when using logarithmic crime as dependent variable).[69] The former is the standard procedure, while the latter uses the proposed measure analogue to all previous results. As expected, the differences are small and none are statistically significant. The standard approach produces amplified coefficients relative to the residual based regression as proven theoretically in Lemma 1 (see section 1.4.2).

The most important variable is unsurprisingly that chosen by the CPSL method and the magnitude is similarly large. In terms of interesting variables for which I cannot reject the null hypothesis, the share of unemployed people stands out given its prominence in the hate crime literature. As it is related, I also report the coefficient regarding the share of economically active people. Both are clearly insignificant in either case. Regarding the relative case, only few variables are significant. This follows from the heterogeneity in the relative effect being smaller.

The CPSL approach can also be used for finding the best one-variable model (see Appendix A.9). In that case, the penalty parameter in the first step (the lasso) is set such that exactly one variable is chosen. The first variable chosen by the lasso is the one that, after standardization, is the most correlated with the dependent variable. Since the results in the Tables 1.8 and 1.9 are weighed by the tercile span (which is similar to standardization as the variable is divided by a measure of its variance), it is not surprising that the same variable appears as the most important one. Independent of the method, the disadvantage of focusing on the single most

[67] Water supply, sewerage, waste management and remediation activities

[68] Information and communication

[69] For the former, the whole panel is used and I cluster at the borough level (which includes robustness to heteroskedasticity) as described in section 1.4.1. Regarding the latter, I use accordingly heteroskedasticity robust errors, using only the cross section of residuals for July 2016 (see section 1.4.2).

Variable	Estimate * Tercile Span	Variable	Estimate * Tercile Span
No Qualifications	-0.29**	No Qualifications	-0.34**
Industry Code E[67]	-0.27	Industry Code E	-0.32
Industry Code J[68]	0.27*	Industry Code J	0.31*
Recent Immigrants	0.25*	Recent Immigrants	0.30*
Remain Vote	0.25	Remain Vote	0.30
Econ. Active	0.10	Econ. Active	0.12
Unemployed	-0.05	Unemployed	-0.06
Using Residuals (ε)		*Using Full Regression*	

Note: Estimate refers to the individual estimated effect of the variable on log(racial or religious hate crime per borough population million). Tercile span is the difference between the mean of the highest tercile of a given variable across the 42 boroughs and the mean of the lowest tercile. Using residuals denotes a cross sectional analysis across 42 boroughs in July 2016 (the month after Brexit) with the dependent variable detrended, deseasonalized, and demeaned at the borough level using 88 months of data. Full regression denotes directly using panel data of 42 boroughs and 88 months and interacting the variable with a dummy for July 2016. Benjamini-Yekutieli (2001) FDR adjusted p values.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.9: Top 5 Individually Important Variables & Unemployment (Relative Increase)

important variable is the arbitrary choice of a one-variable model.

1.7 Mechanisms and Policy Implications

The findings of this paper provide a useful basis to discuss the possible economic mechanisms behind the increase in racial or religious hate crime after the Brexit vote. The empirical part remained agnostic with respect to theoretical considerations. This allows me to asses, specify, and reframe various existing theories. The obtained results are also of direct interest for policing and policy.

1.7.1 Economic Mechanisms

The main channel outlined in this section is information updating through the Brexit vote. The idea is that the Bexit vote resulted in an information-update regarding the expected opinions or norms in society. In a recent paper, Bursztyn et al. (2017) show that information about vote results can significantly affect whether people publicly commit xenophobic actions, which is in line with a model of expected social norms. The driving force in Bursztyn et al.'s (2017) mechanism is that the update through an election result can quickly alter what is considered to be socially acceptable behavior (which is crucial due to a cost of publicly acting in a socially unacceptable way). Bursztyn et al. (2017) do not address spatial heterogeneity directly. The most direct application of that paper regarding spatial heterogeneity in the current Brexit-case could be to consider local election results. This is, however, not in line with the empirical evidence as the increase in hate crime was not more pronounced in areas with a high leave vote (but the opposite is the case, see section 1.6.3).

A first mechanism uses the driving force of Bursztyn et al.'s (2017) approach but is in line with the positive correlation of the increase in hate crime with the remain vote. It builds on the fact that the outcome of the vote was according to pre-vote polls more surprising in remain voting areas (see e.g. Lord Ashcroft, 2016). The key concept of this mechanism is that the larger the surprise of the overall outcome of the Brexit vote (which was to leave the EU), the larger the information update, and hence the larger the effect on hate crime. The underlying idea is that people are influenced by those that live nearby and have a biased opinion regarding the norms in the population as a whole. In a paper closely related and simultaneous to the current paper, Albornoz et al., 2018 outline this mechanism in detail and provide further evidence in favor of it.

The results presented in the previous sections are in line with this mechanism. While the remain vote-share is not among the main heterogeneity variables (see section 1.6), it is correlated with them. Regarding the temporal structure, the mechanism is silent. Arguably, it does not seem justified to reject it due to the transitory nature of the increase in hate crimes. Potential offenders receive further updates about the social norms with passing time or after the action of offending. Moreover, the salience (and hence the 'salience weight') of the 'Brexit signal' decreases over time (see Bordalo et al., 2012).

Another similar mechanism builds again on the driving force of Bursztyn et al.'s (2017) approach and the overall outcome of the Brexit vote. In general, (potential) offenders have to weigh the social and other costs against the benefits of their action. Prior to the vote, some areas are in an equilibrium with relatively high social costs (social sanctions/'push-back'/reactions from bystanders of the crime) but also high benefits, while others are in an equilibrium with lower costs and benefits. Regarding benefits, if the aim is to oust an opposing group, an argument from the conflict literature is that violence targeted at wealthier areas is effective because access to economic opportunities (e.g. jobs) or assets (e.g. flats) is obtained (see e.g. Mitra & Ray, 2014). This is in line with immigration-critical arguments of immigrants taking away jobs or flats. Regarding social costs, Mayda (2006) finds that in wealthy countries such as the UK, skilled people's preferences are more immigration friendly. This evidence is in line with theoretical considerations that skilled people benefit more from immigrants and have a lower risk to lose their job as a consequence of immigration (Mayda, 2006; Borjas, 1995). Consequently, social norms are likely more protective of immigrants in wealthier areas. It is then not surprising that, controlling for the share of immigrants, there is less hate crime in wealthier areas before the Brexit vote (see Appendix A.6).[70] The Brexit vote then provided an information shock about social norms, which had a larger effect in areas where the social costs were higher.

This mechanism specifically focuses on the association of wealth and income proxies (the share of people with formal qualifications) with a higher increase in hate crime. The association with a larger share of recent immigrants seems less surprising and in line with an opportunity channel.[71] The mechanism is again in line with the results found in the previous sections. An

[70]The share of qualified people is highly correlated with the share of immigrants (correlation: 0.64).

[71]More potential victims represent more opportunities for post-Brexit hate crime. While I have no access to victim data, this finding suggests that immigrants are victims of the additional hate crime after the Brexit vote. This is in line with anecdotes of immigrants thats were victims of racial hate crimes and incidents after the

advantage of this mechanism is that it does not require me to assume that offenders offend in the same region (i.e. borough) they live.[72] While this assumption is rather common in the literature, it becomes stronger the smaller the considered regions are, and in case of this paper, the regions are rather small (smaller than in Falk et al., 2011; Krueger & Pischke, 1997; Medoff, 1999, but comparable to Albornoz et al., 2018).

In related literature, a number of other candidate-mechanisms are outlined that are, in principle, applicable to the current setting. While it is impossible to discuss all of them, I list some prominent 'candidate mechanisms' and show how the results of the previous sections contrast with them.

A first class of mechanisms state that a change in fundamentals can lead to an equilibrium with a different level of conflict (see e.g. Esteban & Ray, 2011; Esteban et al., 2012; Caselli & Coleman, 2013). The UK (rather surprisingly) voting to leave the EU has led to a change in fundamentals, both expected and actual. A common example in the economics literature on the Brexit vote is that the exchange rate was affected (see e.g. Douch et al., 2018), which arguably implies expectations about future trade and other economic factors. However, the data does not support mechanisms whose core is a permanent change to an equilibrium with a higher level of conflict. Unlike hate crime, the exchange rate (as well as other fundamentals) was affected rather permanently and did not revert back after six weeks.

A second candidate aspect of the Brexit vote is that it provided a signal on the basis of which people coordinated to riot. However, information-induced riots are considerably shorter-lived than six weeks (see e.g. Glaeser, 1994). There is also no evidence for a number of consecutive gatherings or riots, and the daily time series data of England and Wales show increased numbers of hate crime for virtually every day for weeks after the vote (see Figure 1.2).

Third, the aggressive rhetoric in the media or from campaigning is another potential aspect of the Brexit vote to affect hate crime (for a brief review, see e.g. Gerstenfeld, 2017). However, there was broad coverage and strong rhetoric against immigration several weeks prior to the vote (Moore & Ramsay, 2017). The absence of a pre-vote effect in the hate crime numbers is evidence against this theory.[73]

Fourth, the relative deprivation theory states that the unemployed (or their children) feel deprived and are consequently more likely to follow extreme ideologies and develop violent predispositions (Falk et al., 2011; Siedler, 2011). The unemployment share has consequently been a major variable in previous hate crime analyses (e.g. Falk et al., 2011; Krueger & Pischke, 1997). I find that the unemployment rate is neither individually significantly correlated with the post-vote increase in hate crime, nor does it appear as a key predictor in any model. Despite the relatively high correlations between candidate variables in general, that of the

Brexit vote (e.g. in BBC, 2017b or Independent, 2016, the interviewed victims are Polish, Italian, Romanian, and German - and they report that the offenders specifically mentioned their nationality and the Brexit vote).

[72] Since most hate-crime offenders are not caught, this is difficult to evaluate. At least in Manchester, for which I have the detailed data, most offenders are not caught. Looking at the caught offenders is hardly helpful as it is plausible that those offenders that live further away from the location of the crime are less likely to be caught.

[73] Although it cannot be ruled out as a necessary pre-condition.

unemployment share with the share of recent immigrants, people not stating a religion, and people lacking formal qualifications is only 0.19, 0.06, and 0.46 respectively.[74] The key caveat is that offenders may not commit crimes in the borough they live. Nevertheless, the relative deprivation theory is not a suitable model to explain the spatial heterogeneity in the hate crime increase after the Brexit vote. Even if the crimes were committed by unemployed people, the theory is silent about where this happens if it is not where the unemployed live.

Fifth, according to the opportunity cost theory, people with a low opportunity cost of time are more likely to commit hate crimes (Medoff, 1999). The share of people with a low wage is consequently in the focus of some hate crime studies (e.g. Krueger & Pischke, 1997; Medoff, 1999). I find that income proxies are positively related to the post-vote increase in hate crime. Therefore, the opportunity cost theory is also unfit to explain the observed spatial heterogeneity.

Finally, the data in this paper naturally covers only reported hate crimes. Consequently, changed reporting is another candidate mechanism. First, hate crimes could have been reported after the Brexit vote that were previously not reported at all. Second, victims could have been more sensitive about the fact that they are victims of a hate crime instead of a regular crime.[75] Regarding the second issue, there are five types of crime which are recorded either as racially or religiously aggravated or non-aggravated.[76] The hate crime data used previously includes all types of crime, but these five types are sufficiently often racially or religiously aggravated that the Home Office (2017) keeps specific record of aggravated and non-aggravated crimes of these types. As shown in Appendix A.5, there is no drop in the non-aggravated crimes when the respective aggravated crimes increase after the Brexit vote. This is evidence against mere relabeling of crimes. Moreover, there are strict guidelines on the side of the police to minimize the influence of reporting and subjectivity. Regarding the first issue, I focus on crimes, not incidents. Minor incidents that are reported to the police as hate crime do not enter my data as these would be classified as incidents, but not crimes (see section 1.2). Incidents that do qualify as crimes and were not at all reported previously are a more serious issue. However, if I assume that crimes of the highest severity are equally often reported before and after the vote, there is evidence against this mechanism: the share of most severe crimes among hate crimes did not change after the Brexit vote (see Figure 1.4).

[74]Testing individual correlations represents a comparison without controls, but remains unaffected by correlations between candidate variables. The linear prediction model's key predictors do not have the former problem, but due to these correlations, it cannot be guaranteed that the unemployment rate does not feature the true model. Therefore, it cannot be ruled out that the unemployment rate had any effect. However, given the rather low pairwise correlations, a more complex correlation construct would be necessary. The unemployment rate is certainly not the predominant key factor, and there is evidence against it being important at all.

[75]Theoretically, the reporting issue has at least one more dimension, which is the reporting behavior of the police. As shown in Figure 1.3, the way in which the police was called to a crime has, if anything, changed to the police being less often called over the radio. Crucially, that implies that fewer crimes were directly reported by police officers themselves in the period after the Brexit vote, which is evidence against different reporting on the police-side.

[76]These are: assault with injury, assault with no injury, harassment, public distress, and criminal damage.

1.7.2 Policy Implications

The question regarding policy implications involves that of external validity. I have argued that the key feature of the Brexit vote regarding hate crime is that it was a public information shock about society's preferences regarding immigrants. While immigrants were not the only focus of the Brexit vote, it was one of only few key themes (see Moore & Ramsay, 2017). The preferences of those eligible to vote were then publicly revealed. Arguably, referendums are generally more focused than elections. However, certain elections do have a strong focus on immigration (for example) of the winning party or candidate, similar to the Brexit vote, and have also led to increases in hate crime (e.g. the Trump election, see Bursztyn et al., 2017; Mueller & Schwarz, 2018; but potentially also the latest Italian election, see Monella, 2018). In that respect, there are two key policy implications.

The first concerns policing. For the police, it is of paramount importance to be prepared for increases in hate crime of such high magnitudes. Handling racial or religious hate crimes requires tact, experience, and expertise. Refresher courses for dealing with hate crime for officers in the most affected locations, or distributing expert officers accordingly are two examples how an expected rise in hate crime can be addressed better than an unexpected one (see Gerstenfeld, 2017). In case of a similar event, for example another referendum that is tied closely to immigrants, the best predictor for the absolute rise in racial or religious hate crime is the share of recent immigrants and the best predictor for the relative increase is the share of people with formal qualifications;[77] and this finding could arguably be extended beyond London and Manchester, in particular to other British urban areas. Terror attacks also led to sharp increases in hate crimes, but in different places.

The second policy implication concerns politics, namely holding certain referendums. While many referendums target neither de jure nor de facto immigrants, or a specific group of people in general, the Brexit vote shows that some did and it is arguably not unlikely that others will follow. Taking into account the psychological trauma of hate crime victims (see e.g. Levin, 1999; Craig-Henderson & Sloan, 2003; McDevitt et al., 2001), the effects of such referendums on hate crime should be taken into account. Consequently, appropriate accompanying and preventive measures should be taken in the respective areas to counteract the potential increase in hate crime in the weeks after the vote. In practice, examples of such measures include training or mediation in schools or youth centers, or psychological and legal victim assistance (see Gerstenfeld, 2017). Campaigns to encourage reporting hate crimes seem particularly suitable. Paired with investigating hate crimes thoroughly, this increases the expected legal cost for offenders, counteracting the decrease in the expected social cost.

Moreover, while a shock is fundamental to analyze the role of expected social norms, a gradual change in these expected norms has the same effect in the mechanisms discussed above. Consequently, if there are signs of such a gradual change in expected norms or beliefs of others, my findings help again to guide policy efforts to the most relevant locations. In that respect, long term policies are more relevant, such as creating permanent training or mediation centers or schemes. Reporting can be addressed more fundamentally with more time available.

[77] Alternatively, using the calculated predictions of the full model leads to an even better prediction.

The confidence in the police can be strengthened, for example, or a strong local awareness of alternative reporting possibilities can be built (see Gerstenfeld, 2017). The latter ideally also result in police investigations and hence increase the expected punishment for offenders.

1.8 Conclusion

The key concept in the criminology literature with respect to increases in hate crime is arguably that of 'trigger events': an event pushes certain potential offenders over the threshold to commit a hate crime (see King & Sutton, 2013). In the classic economic formulation, a crime is committed when its perceived benefits exceed the costs (see Becker, 1968), and a 'trigger event' can affect either factor. In this framework, my results help to narrow down what the trigger implied by the Brexit vote was, in which areas triggered people acted, and what underlying mechanisms were at play.

Regarding the trigger itself, the evidence points towards the Brexit vote being a public information shock regarding society's attitude towards immigrants. While this cannot be stated with certainty, several prominent alternative mechanisms can be ruled out. The temporal structure, for example, is neither in line with an 'inflammatory rhetoric' mechanism, nor one of changed economic fundamentals.

The spatial heterogeneity of the increase in hate crime was substantial, hence analyzing in which areas triggered people acted is important. As a first step, I show that the heterogeneity captured by census and vote data is considerable.

In a second step, I derive linear (prediction) models. These provide predictors with valid confidence intervals. Both conditional post-selection lasso and the novel ad hoc splitting based estimation lead to the same result: the share of recent immigrants is the best predictor for the absolute increase in hate crime, and the fraction of people with formal qualifications is the best predictor for the relative increase. The former, combined with a lower heterogeneity in the relative effect, is in line with an opportunity channel. One mechanism consistent with the latter is that the information update from the Brexit vote resulted in lower expected social sanctions of hate crime, and that these sanctions are more important to prevent hate crime in wealthier areas. The predictions and predictor estimates are also important to guide future qualitative and quantitative research to better understand the effect of the Brexit vote on hate crime, and potentially hate crime more generally. Moreover, they are of key importance for policing and preventive measures.

In a third step, allowing for quadratic terms and interactions provides evidence that interactions do not play an important role in best predicting the increase in hate crime after the Brexit vote.

Finally, in a methodologically more standard fashion, each candidate variable is interacted with the treatment dummy in separate regressions. This provides insight into individual correlations, but ignores the model selection problem. Nevertheless, the results are in line with the previous findings. Interestingly, the unemployment rate is neither economically nor statistically significant.

This four-step procedure of applying state-of-the-art machine learning methods to a universal unique treatment proves to be reasonable for the current setting, in part due to the simulation evidence from using the real data. The first (and fourth) step can be directly applied to other settings. Whether or not conditional post-selection lasso can be used directly depends on the setting. The current setting allows me to analyze the heterogeneity in a cross sectional setup, and use the remaining panel data for detrending, demeaning, and deseasonalizing the dependent variable at the borough level. The key reasons why this is possible are that the effect is temporary, and because including a borough fixed effect is reasonable. The third step, using the proposed splitting based estimation, is theoretically ambiguous and relies on simulations. Further theoretical research regarding the bias and consistency properties of this method is needed and seems worthwhile given the promising simulation-performance. In my simulations, the method performs best with few variables in the true data generating process. Many variables carrying a strong signal are more problematic for the method, as then the frequency with which the same model is chosen multiple times becomes prohibitively low.

Overall, this paper provides evidence on an agnostic basis to evaluate possible mechanisms at play after the Brexit vote. This generates insight relevant for policy and academia alike. While this insight does not provide all final answers regarding the specific mechanisms at play, my findings help to identify those regions where the effect was particularly strong and also those where the effect was virtually absent. Moreover, the fact that the result depicts variables that are correlated and predictive of a larger increase can guide future studies. Either these variables or something strongly correlated with them is likely of paramount importance. More research is needed; examples include qualitative studies of the offenders' motivation and how they interpret public information shocks differently in different circumstances, or what prevents increased offending in certain regions.

Chapter 2

Dynamically Optimal Treatment Allocation using Reinforcement Learning

2.1 Introduction

Consider a situation wherein a stream of individuals arrive sequentially - for example when they get unemployed - to a social planner. Once each individual arrives, the planner needs to decide what kind of action or treatment assignment - for example whether or not to offer free job training - to provide to the individual, while taking into account various institutional constraints such as limited budget and capacity. The decision on the treatment has to be taken instantaneously, without knowledge of the characteristics of future individuals, though the planner can, and should, form expectations over these future characteristics. Once an action is taken, the individual is assigned to a specific treatment, leading to a reward, i.e. a change in the utility for that individual. The planner does not observe these rewards directly since they may be only realized much later, but she can estimate them using data from some past observational studies. The action of the planner does, however, generate an observed change to the institutional variables, such as budget or capacity. The planner takes note of these changed constraints, and waits for the next individual to arrive. The process then repeats until either time runs out or terminal constraints are hit, e.g. when budget or capacity is depleted.

The above setup nests many important applications including selecting and supporting people or projects with a fixed annual or quarterly budget or capacity. Indeed, we contend that this is a common situation across governmental and non-governmental settings. Job training, development aid, or any form of credit are only three examples that illustrate the importance of reaching the best allocation decisions. In this paper, we show how one can leverage randomized control trial or observational data and employ Reinforcement Learning to estimate the optimal treatment assignment rule or policy function in this dynamic context that maximizes ex-ante expected welfare.

If the dynamic aspect can be ignored, there exist a number of methods to estimate an optimal policy function that maximizes social welfare, starting from the seminal contribution of Manski (2004), and further extended by Hirano and Porter (2009), Stoye (2009, 2012), Chamberlain (2011), Bhattacharya and Dupas (2012) and Tetenov (2012), among others. More recently, Kitagawa and Tetenov (2018), and Athey and Wager (2018) have proposed using Empirical Welfare Maximization (EWM) in this context. While these papers address the question of optimal treatment allocation under co-variate heterogeneity, the resulting treatment rule is static in that it does not change with time, nor with current values of institutional constraints such as budget and capacity. Moreover, such rules assume that the population of individuals under consideration is observable, which is not the case in our setup since individuals arrive sequentially. We also impose the requirement that the institutional constraints (e.g. limited budget) hold ex-post. This too is in contrast to previous work which either required the constraints to hold on average (e.g. Bhattacharya and Dupas, 2012), or employed a purely random allocation when the constraints were not met (e.g. Kitagawa and Tetenov, 2018). Note that the latter essentially implies using a different welfare criterion than the standard ex-ante expected welfare criterion.

On the other hand, there also exist a number of methods for estimating the optimal treatment assignment policy using sequential or 'online' data. This is known as the contextual bandit problem, and a number of authors have proposed online learning algorithms for this purpose including Dudik et al. (2011), Agarwal et al. (2014), Russo and van Roy (2016) and Dimakopoulou et al. (2017). The central concern in bandit problems is the tradeoff between exploration, for estimating the optimal treatment rule, and exploitation, for applying the best current policy function. However, the eventual policy function that is learnt is still static in the sense of not changing with institutional constraints or time. By contrast, we do not attempt the question of exploration; indeed, the data we use is 'offline' in the sense of being observed historically. However dynamics arise in our context due to inter-temporal tradeoffs: the social planner has to weigh the utility gain from treating an individual arriving today against the possibility that she may run out of budget or capacity for treating a more deserving individual in the future.

Perhaps the closest set of results to our work is from the literature on dynamic treatment regimes. We refer to Laber et al. (2014) and Chakraborty and Murphy (2014) for an overview. Dynamic treatment regimes consist of a sequence of individualized treatment decisions for health related interventions. Like our paper this involves solving a Bellman equation. However the number of stages or decision points is quite small, typically between 1 and 3. By contrast, the number of decision points (which is inversely related to frequency of arrivals) in our setting is very high, and in fact we will find it more convenient to take the limit as the arrivals become continuous and formulate the model as a differential equation. Additionally, in our paper, we aim to find the optimal policy within a pre-specified class of policy rules. As explained by Kitagawa and Tetenov (2018), one may wish to do this for computational, legal or incentive compatibility reasons. We thus propose algorithms to solve for the optimum in a restricted policy class. This is typically not a concern in dynamic treatment regimes. Finally, the data requirements are very different. Dynamic treatment regimes are estimated from Sequential, Multiple, Assignment Randomized Trials (SMART) (Murphy, 2005; Lei et al., 2012), where participants move through two or three stages of treatment, which is randomized in each stage. By contrast we only make use of a single RCT. Each individual in our setup is only exposed to treatment once. The dynamics are faced not by the individual, but by the social planner.

In this paper we propose techniques for estimating an optimal policy function that maps the current state variables of observed characteristics and institutional constraints to probabilities over the set of actions. We treat the class of policy functions as given. Then for any policy from that class, we can write down a Partial Differential Equation (PDE) that characterizes the evolution of the expected value function under that policy, where the expectation is taken over the distribution of the individual covariates. Using the data, we can similarly write down a sample version of the PDE that provides estimates of these value functions. The estimated policy rule (within the candidate class) is the one that maximizes the estimated value function at the initial time period. By comparing the two PDEs, we can uniformly bound the difference in their corresponding solutions, i.e. the value functions. We then use this to in turn bound the welfare regret from using the estimated policy rule relative to the optimal policy in the candidate class. We find that the regret is of the (probabilistic) order $n^{-1/2}$; this is the same rate as that obtained by Kitagawa and Tetenov (2018) in the static simultaneous allocation case. And also as in Kitagawa and Tetenov (2018), the rate depends on the complexity of the policy function class being considered, as indexed by its VC dimension. While treatment choice can be approached as a classification problem, we do not base our approach on this literature. Due to our setting being dynamic, we instead build on the optimal-control literature. Hence our theoretical results are based on exploiting the properties of the PDEs using novel techniques that may be of independent interest. Our techniques also allow for potentially non-differentiable value functions through the notion of viscosity solutions to PDEs.

In terms of computation, we approximate the PDEs with suitable dynamic programming problems by discretizing the number of arrivals. We then propose a modified Reinforcement Learning algorithm that can be applied on the latter and that achieves the best value in a pre-specified class of policy rules. Previous work in this literature in economics has used Generalized Policy Iteration (see e.g. Benitez-Silva et al., 2000). While this method works well with discrete states, there are two major drawbacks: First, the algorithm becomes cumbersome even with a few continuous states.[78] Second, and more importantly, it does not allow for restricting the solution to a pre-specified class of policy rules. In this paper, we propose a modified Reinforcement Learning (RL) algorithm to solve this problem.[79] We adapt the Actor-Critic algorithm (e.g. Sutton et al., 2000; Bhatnagar et al., 2009) that has been applied recently to great effect in applications as diverse as playing Atari games (Mnih et al., 2015) and medical diagnosis (De Fauw et al., 2018). This appears to be a novel approach to the solution of Hamilton-Jacobi-Bellman type PDEs. In addition to boasting of strong convergence properties, our algorithm is also parallelizable, which translates to very substantial computational gains.

We also outline the computational and numerical properties of our algorithm. On the computational side, we prove that it converges to a well defined optimum. This is based on the convergence of stochastic gradient descent, and we are able to directly employ theorems from the RL literature to this effect. On the numerical approximation side, we use results from the theory of viscosity solutions to provide conditions on the level of discretization so that the numerical error from this is negligible compared to the statistical error in the regret bounds.

[78] Continuous states may be handled through discretization or parametric policy iteration. The former is typically slower and suffers from a strong curse of dimensionality (see Benitez-Silva et al., 2000, Section 2.5); while the latter requires numerical integration which is also very demanding with more than a few states. Also, there is no proof of convergence for parametric policy iteration, and it is known that it fails to converge in some examples.

[79] We refer to Sutton and Barto (2018) for a detailed comparison of recent RL algorithms with policy iteration.

Finally, we also incorporate non-compliance in our dynamic environment. Under instrument monotonicity, the only population subgroup for whom the social planner can affect a change in welfare are the compliers. For always-takers and never-takers the social planner would always end up treating the former and never treating the latter. Hence we can still perform welfare maximization using identified quantities, i.e. the local average treatment effect, and the conditional probabilities of being in each of the sub-groups given the covariates.

We illustrate the feasibility of our algorithm using data from the Job Training Partnership Act (hereafter JTPA). We incorporate dynamic considerations into this setting in the sense that the planner has to choose whether to send individuals for training as they arrive sequentially. The planner faces a budget constraint, and the population distribution of arrivals is also allowed to change with time. We consider policy rules composed of 5 continuous state variables (3 individual covariates along with time and budget), to which we add some interaction terms. We then apply our Actor-Critic algorithm to estimate the optimal policy rule.

2.2 An Illustrative Example: Dynamic Treatment Allocation with Budget or Capacity Constraints

To illustrate our setup and methods consider the following example: A social planner wants to provide training to unemployed people. The planner starts with a fixed budget that she can use to fund the training. Individuals arrive sequentially when they get unemployed, and the planner is required to provide an instantaneous decision on whether to allocate training to each individual as he/she arrives. The planner makes a decision based on the current budget and the characteristics of the individual. Some individuals benefit more from the treatment than others, so the planner has to decide whether to provide training to the current individual, or to to hold off for a more eligible applicant at the risk of losing some utility due to discounting. To help the planner decide, she can draw on information from a historical Randomized Control Trial (RCT) on the effect of training, along with data on past dynamics of unemployment. We assume in this section that the waiting time between arrivals is drawn from an exponential distribution that does not vary with time, and also that the cost of training is the same for all individuals. This allows us to characterize the problem in terms of Ordinary Differential Equations (ODEs), which greatly simplifies the analysis. We consider more general setups, leading to Partial Differential Equations (PDEs), in the next section.

Formally, let x denote the vector of characteristics on individual, based on which the planner makes a decision on whether to provide training ($a = 1$) or not ($a = 0$). The current budget is denoted by z. Once an action, a, has been chosen, the planner receives a reward, i.e. a change in social welfare, of $Y(a)$ that is equivalent to the potential outcome of the individual under action a. We shall assume that $Y(a)$ is affected by the covariates x but not the budget. Define $r(x, a) = E[Y(a)|x]$ as the expected (instantaneous) reward for the social planner when the planner chooses action a for an individual with characteristics x. Since we only consider additive welfare criteria in this paper, we may normalize $r(x, 0) = 0$, and set

$r(x,1) = E[Y(1) - Y(0)|x]$. Note that we can accommodate various welfare criteria, as long as they are utilitarian, by redefining the potential outcomes.

If the planner takes action $a = 1$, her budget is depleted by c, otherwise it stays the same. The next individual arrives after a waiting time Δt drawn from an exponential distribution Exponential(N). Note that N is the expected number of individuals in a time interval of length 1 (one could alternatively use this as the definition of N itself). We shall use N to rescale the budget so that $c = 1/N$. With this, we reinterpret the budget as the fraction of people in a unit time period that can be treated. Each time a new individual arrives, the covariates for the individual are assumed to be drawn from a distribution F that is fixed but unknown. The utility from treating successive individuals is discounted exponentially by $e^{-\beta \Delta t}$. Note that the expected discount factor is given by $E[e^{-\beta \Delta t}] = 1 - \frac{\tilde{\beta}}{N}$ where $\tilde{\beta} = \beta + O(N^{-1})$. For simplicity, we shall let $\tilde{\beta} = \beta$.

The planner chooses a policy function $\pi(a|x,z)$ that maps the current state variables x, z to a probabilistic choice over the set of actions:

$$\pi(.|x,z) : (x,z) \longrightarrow [0,1].$$

The aim of the social planner is to determine a policy rule that maximizes expected welfare after discounting. Let $v_\pi(x,z)$ denote the value function for policy π, defined as the expected welfare from implementing policy $\pi(\cdot|x,z)$ starting from the state (x,z). In other words,

$$v_\pi(x,z) = E\left[\frac{1}{N} \sum_{i=1}^{\infty} e^{-\beta T_i} r(x_i,1)\pi(1|x_i,z_i)\mathbb{I}(z_i > 0) \middle| x,z \right],$$

where the expectation is joint over the times of arrival $T_i := \sum_{j=1}^{i} \Delta t_j$, covariates $x \sim F$ and z_i which evolves according to the distribution of x and the randomization of the policy $\pi(.)$. It is more convenient to represent $v_\pi(z,t)$ in a recursive form as the fixed point to the equations[80]

$$\begin{aligned} v_\pi(x,z) &= \frac{r(x,1)}{N}\pi(1|x,z) + \left(1 - \frac{\beta}{N}\right) E_{x' \sim F}\left[v_\pi\left(x', z - \frac{1}{N}\right)\pi(1|x,z) + v_\pi(x',z)\pi(0|x,z) \right] \\ &\quad \text{for } z > 1/N \\ v_\pi(x,0) &= 0. \end{aligned}$$

We can obtain more insight into the model if we integrate out x. This gives us the integrated value function, which we also call the h-function:

$$h_\pi(z) = E_{x \sim F}[v_\pi(x,z)].$$

Define $\bar{\pi}(a|z) = E_{x \sim F}[\pi(a|x,z)]$ and $\bar{r}_\pi(z) = E_{x \sim F}\left[r(x,1)\pi(1|x,z)\right]$. We can also obtain $h_\pi(.)$

[80] We assume for simplicity that z is always in multiples of $1/N$.

as the solution to the recursive equations

$$h_\pi(z) = \frac{\bar{r}_\pi(z)}{N} + \left(1 - \frac{\beta}{N}\right)\left\{h_\pi\left(z - \frac{1}{N}\right)\bar{\pi}(1|z) + h_\pi(z)\bar{\pi}(0|z)\right\} \text{ for } z > 1/N, \tag{2.1}$$
$$h_\pi(0) = 0.$$

In practice the value of N is very large, so that budget is almost continuous. In such cases it is more convenient to work with the limiting version of (2.1) as $N \to \infty$. To this end let us subtract $\left(1 - \frac{\beta}{N}\right) h_\pi(z)$ from both sides of equation (2.1), multiply both sides by N and take the limit as $N \to \infty$. We then end up with the following Ordinary Differential Equation (ODE) for the evolution of $h_\pi(.)$:

$$\beta h_\pi(z) = \bar{r}_\pi(z) - \bar{\pi}(1|z)\partial_z h_\pi(z), \qquad h_\pi(0) = 0, \tag{2.2}$$

where ∂_z denotes the differential operator with respect to z, and $h_\pi(0) = 0$ is the initial condition for the ODE.[81] (2.2) is similar to the well known Hamilton-Jacobi-Bellman (HJB) equation. The key difference however is that (2.2) determines the evolution of $h_\pi(.)$ under a specified policy, while the HJB equation provides an equation for the evolution of the value function under an optimal policy. Both the discrete and continuous forms (2.1) , (2.2) are insightful: we use the discrete version for computation, while the ODE version is more convenient for our theoretical results.

The social planner's decision problem is then to choose the optimal policy π^* that maximizes the ex-ante expected welfare $h_\pi(z_0)$, over a pre-specified class of policies Π, where z_0 denotes the initial value of the budget:

$$\pi^* = arg \max_{\pi \in \Pi} h_\pi(z_0).$$

How should the planner choose Π? Let us first look at the first best policy function:

$$\pi^*_{FB}(1|x, z) = \mathbb{I}\left\{r(x, 1) + \partial_z h_{\pi^*_{FB}}(z) > 0\right\}.$$

While $\pi^*_{FB}(.)$ maximizes the planner's welfare, it suffers from two main drawbacks: First, estimation of $\pi^*_{FB}(1|x, z)$ is computationally expensive when the dimension of states is very large. Second, $\pi^*_{FB}(1|x, z)$ is highly non-linear in x, z, and the social planner may prefer policies that are simpler for legal, ethical or incentive compatibility reasons. For instance, if the policy function is highly non-linear in z, individuals may rationally decide to arrive at slightly different times when the budget is different. Kitagawa & Tetenov (2018) provide various arguments as to why it would be useful to restrict the complexity of $\pi(1|x, z)$ with respect to x in the static case. Ultimately, the choice of Π depends on computational and policy considerations of the planner. For our theoretical results we take this as given and consider a class Π of policies indexed by some (possibly infinite dimensional) parameter θ.

[81] The Picard-Lindeöf theorem guarantees a unique solution to (2.2) as long as $\bar{r}_\pi(z)$ and $\bar{\pi}(1|z)$ are continuous.

Our results on computation are only slightly more restrictive in that we require $\pi_\theta(.)$ to be differentiable in θ. This still allows for rich spaces of policy functions. A rather convenient one that we use in our empirical results is the class of exponential soft-max functions. To describe this, let $f(x, z)$ denote a vector of basis functions of dimension k over the space of (x, z). The soft-max function takes the form

$$\pi_\theta(1|x, z) = \frac{\exp(\theta' f(x, z))}{1 + \exp(\theta' f(x, z))}$$

for some parameter θ. In this case, the set of all possible policy functions considered is the parametric class $\Pi \equiv \{\pi_\theta(.|s) : \theta \in \Theta\}$, for some compact set Θ. An advantage of the softmax form is that it can be used to approximate any deterministic policy arbitrarily well. Indeed, it can approximate $\pi^*_{FB}(.)$ if one chooses the dimension k of the basis to be large enough. At the same time, it permits greater freedom than a deterministic policy since in problems with significant function approximation, the best approximate policy is usually stochastic. Other choices of policy classes are also possible, e.g. multi-layer neural networks. In Section 2.6.3 we additionally discuss how one could work with deterministic classes of policy rules such as the linear and generalized eligibility rules considered in Kitagawa and Tetenov (2018).

In what follows, we specify the policy class as $\Pi \equiv \{\pi_\theta(.) : \theta \in \Theta\}$ and denote $h_\theta \equiv h_{\pi_\theta}$ along with $\bar{r}_\theta \equiv \bar{r}_{\pi_\theta}$. The social planner's problem is then

$$\theta^* = arg \max_{\theta \in \Theta} h_\theta(z_0). \tag{2.3}$$

Clearly (2.3) is not feasible as one does not know $r(x, 1)$, nor the distribution F to calculate $h_\theta(z)$. However the planner does have access to an RCT. Let us assume that the RCT consists of an iid draw of size n from the distribution F. The empirical distribution F_n of these observations is thus a good proxy for F. Let W denote the treatment assignment in the RCT data. We also let $\mu(x, w) = E[Y(w)|X = x, W = w]$ denote the conditional expectations for $w = 0, 1$, and $p(x)$, the propensity score. We can then estimate $r(x, 1)$ in many different ways. We recommend a doubly robust method (see Athey and Wager, 2018), e.g.

$$\hat{r}(x, 1) = \hat{\mu}(x, 1) - \hat{\mu}(x, 0) + (2W - 1)\frac{Y - \hat{\mu}(x, W)}{W\hat{p}(x) + (1 - W)(1 - \hat{p}(x))},$$

where $\hat{\mu}(x, w)$ and $\hat{p}(x)$ are non-parametric estimates of $\mu(x, w)$ and $p(x)$ respectively, and Y is the observed outcome variable. In practice, $\hat{\mu}(x, w)$ can be obtained through series regression, or lasso; while $\hat{p}(x)$ can be obtained using a logistic regression or logistic lasso.

Define $\hat{\pi}_\theta(a|z) = E_{x \sim F_n}[\pi_\theta(a|x, z)]$ and $\hat{r}_\theta(z) = E_{x \sim F_n}[r(x, 1)\pi_\theta(1|x, z)]$. Based on the knowledge of $\hat{r}(.)$ and F_n, we can calculate the sample version of the h-function in the discrete case as the solution to the recursive equations:

$$\hat{h}_\theta(z) = \frac{\hat{r}_\theta(z)}{N} + \left(1 - \frac{\beta}{N}\right)\left\{\hat{h}_\theta\left(z - \frac{1}{N}\right)\hat{\pi}_\theta(1|z) + \hat{h}_\theta(z)\hat{\pi}_\theta(0|z)\right\} \text{ for } z > 1/N, \tag{2.4}$$

$$\hat{h}_\theta(0) = 0.$$

Alternatively, in the limit as $N \to \infty$, we have the following ODE:

$$\beta \hat{h}_\theta(z) = \hat{r}_\theta(z) - \hat{\pi}_\theta(1|z)\partial_z \hat{h}_\theta(z), \qquad \hat{h}_\theta(0) = 0. \tag{2.5}$$

Using $\hat{h}_\theta(.)$ we can solve a sample version of the social planner's problem:

$$\hat{\theta} = arg \max_{\theta \in \Theta} \hat{h}_\theta(z_0).$$

Given θ, one could solve for $\hat{h}_\theta$ by backward induction starting from $z = 1/N$ using (2.4). In this simple example this is feasible as long as N is not too large, but note that one would still need to calculate the summations $E_{x \sim F_n}[\pi_\theta(a|x,z)]$ and $E_{x \sim F_n}[r(x,1)\pi_\theta(1|x,z)]$ for all the possible values of z. And even where solving for $\hat{h}_\theta(z_0)$ is feasible, we yet have to maximize this over $\theta \in \Theta$. Such a strategy is computationally too demanding.[82] Therefore in this paper we advocate a Reinforcement Learning algorithm that directly ascends along the gradient of $\hat{h}_\theta(z_0)$ *and* simultaneously calculates $\hat{h}_\theta(z_0)$ in the same series of steps. This makes the algorithm very efficient. We describe this in greater detail in Section 2.4.

In the remainder of this section, we briefly outline the theory behind our approach. The derivations here are informal, but provide intuition for our formal results in Section 2.5.

2.2.1 Regret Bounds

We would like to know how $\hat{\theta}$ compares to θ^* in terms of the regret $h_{\theta^*}(z_0) - h_{\hat{\theta}}(z_0)$. The bound for this depends on the sample size n and the complexity of the space $\Pi = \{\pi_\theta : \theta \in \Theta\}$. One way to determine the complexity of Π is by its Vapnik-Cervonenkis (VC) dimension. In particular, denote by v the VC-subgraph index of the collections of functions

$$\mathcal{I} = \{\pi_\theta(1|\cdot, z) : z \in [0, z_0], \theta \in \Theta\}$$

indexed by z and θ. We shall assume that v is finite. Kitagawa and Tetenov (2018) were the first to characterize the regret in the static setting in terms of the VC dimension of Π. Relative to this, our definition of the complexity differs in two respects. First, our policy functions are probabilistic (but cover deterministic treatment rules as special cases). Second, for the purposes of calculating the VC dimension, we treat z as an index to the functions $\pi_\theta(1|\cdot, z)$, similarly to θ. In other words $\pi_\theta(1|\cdot, z_1)$ and $\pi_\theta(1|\cdot, z_2)$ with the same θ are treated as different functions. This is intuitive since how rapidly the policy rules change with budget is also a measure of their complexity. Note that the VC index of $\mathcal{I}$ is not $\dim(\theta)$ when θ is Euclidean, but is in fact smaller. To illustrate, suppose that x is univariate and

$$\mathcal{I} \equiv \{\text{Logit}(g_1(z) + g_2(z)x) : g_1, g_2 \text{ are arbitrary functions}\}.$$

[82] Alternatively, one could solve ODE (2.5) directly, but this also has the same complexity.

In this case the VC-subgraph index of $\mathcal{I}$ is at most 2. This is because the VC-subgraph index of the class of functions $\mathcal{F} = \{f : f(x) = a + xb \text{ over } a, b \in \mathbb{R}\}$ is 2 since $\mathcal{F}$ lies in the (two dimensional) vector space of the functions $1, x$. The VC-subgraph index of $\mathcal{I}$ is the same or lower than that of $\mathcal{F}$ (since the logit transformation is monotone), hence $v \leq 2$ in this example.

We now show how one can derive probabilistic bounds for the regret $h_{\theta^*}(z_0) - h_{\hat{\theta}}(z_0)$. First, under the assumption of finite VC dimension and other regularity conditions, Athey and Wager (2018) show that for doubly robust estimates of the rewards,

$$E_{x\sim F}\left[\sup_{\theta\in\Theta, z\in[0,z_0]} |\bar{r}_\theta(z) - \hat{r}_\theta(z)|\right] \leq C_0\sqrt{\frac{v}{n}}, \tag{2.6}$$
$$E_{x\sim F}\left[\sup_{\theta\in\Theta, z\in[0,z_0]} |\bar{\pi}_\theta(1|z) - \hat{\pi}_\theta(1|z)|\right] \leq C_0\sqrt{\frac{v}{n}}$$

for some universal constant $C_0 < \infty$. The above implies that the ODEs (2.2) and (2.5) governing the motion of $h_\theta(z)$ and $\hat{h}_\theta(z)$ are very similar, which indicates that $h_\theta(.)$ and $\hat{h}_\theta(.)$ should be uniformly close. Formally, denote $\hat{\delta}_\theta(z) = h_\theta(z) - \hat{h}_\theta(z)$. Now under some regularity conditions (made precise in Section 2.5), it can be shown that $\sup_{\theta\in\Theta, z\in[0,z_0]} |h_\theta(z)| < \infty$. Then from (2.2) and (2.5), we have

$$\partial_z\hat{\delta}_\theta(z) = \frac{-1}{\bar{\pi}_\theta(1|z)}\beta\hat{\delta}_\theta(z) + \frac{\bar{r}_\theta(z)}{\bar{\pi}_\theta(1|z)} - \frac{\hat{r}_\theta(z)}{\hat{\pi}_\theta(1|z)} + \left(\frac{1}{\hat{\pi}_\theta(1|z)} - \frac{1}{\bar{\pi}_\theta(1|z)}\right)\beta\hat{h}_\theta(z); \quad \hat{\delta}_\theta(0) = 0$$

or

$$\partial_z\hat{\delta}_\theta(z) = \frac{-1}{\bar{\pi}_\theta(z)}\beta\hat{\delta}_\theta(z) + K_\theta(z); \quad \hat{\delta}_\theta(0) = 0, \tag{2.7}$$

where

$$E_{x\sim F}\left[\sup_{\theta\in\Theta, z\in[0,z_0]} |K_\theta(z)|\right] \leq M\sqrt{v/n}$$

for some $M < \infty$ by (2.6) and the uniform boundedness of $h_\theta(z)$, assuming that $\bar{\pi}_\theta(z)$ is uniformly bounded away from 0. Rewriting (2.7) in integral form and taking the modulus on both sides, we obtain

$$\left|\hat{\delta}_\theta(z)\right| \leq zM\sqrt{\frac{v}{n}} + \int_0^z \frac{1}{\bar{\pi}_\theta(\omega)}\beta\left|\hat{\delta}_\theta(\omega)\right| d\omega,$$

based on which we can conclude via Grönwall's inequality that

$$\left|\hat{\delta}_\theta(z)\right| \leq M_1\sqrt{\frac{v}{n}}$$

uniformly over all $\theta \in \Theta, z \in [0, z_0]$, for some $M_1 < \infty$ - here, all the inequalities should be interpreted as holding with probability approaching one under F. The above discussion implies

$$h_{\theta^*}(z_0) - h_{\hat{\theta}}(z_0) \leq 2\sup_{\theta\in\Theta, z\in[0,z_0]}\left|\hat{\delta}_\theta(z)\right| \leq 2M_1\sqrt{\frac{v}{n}}$$

with probability approaching one under F. Hence through this derivation, we have shown that the regret declines with $\sqrt{v/n}$, which is the same rate that Kitagawa and Tetenov (2018) derived for the static case.

2.2.2 Discretization and Numerical Error

As we mentioned earlier, we do not recommend using the ODE version of the problem to solve for $\hat{\theta}$. Instead, it is usually much quicker to solve a discrete analogue of the problem as in (2.4). Now in practice N maybe unknown or too large, but in either case we can simply employ any suitably large normalizing factor b_n, and solve the recurrence relation

$$\tilde{h}_\theta(z) = \frac{\bar{r}_\theta(z)}{b_n} + \left(1 - \frac{\beta}{b_n}\right) \left\{ \tilde{h}_\theta \left(z - \frac{1}{b_n}\right) \bar{\pi}_\theta(1|z) + \tilde{h}_\theta(z) \bar{\pi}_\theta(0|z) \right\} \tag{2.8}$$

for $\tilde{h}_\theta(.)$ together with the initial condition $\tilde{h}_\theta(0) = 0$. We are now faced with the issue of choosing b_n so that $\tilde{h}_\theta(.)$ is sufficiently close to $\hat{h}_\theta(.)$ obtained from (2.5).

To answer this, we first note that $\hat{h}_\theta$ and $\partial_z \hat{h}_\theta$ are both Lipschitz continuous uniformly in θ under some regularity conditions (c.f Section 2.5). Lipschitz continuity of $\partial_z \hat{h}_\theta$ implies

$$\hat{h}_\theta(z) = \frac{\bar{r}_\theta(z)}{b_n} + \left(1 - \frac{\beta}{b_n}\right) \left\{ \hat{h}_\theta \left(z - \frac{1}{b_n}\right) \bar{\pi}_\theta(1|z) + \hat{h}_\theta(z) \bar{\pi}_\theta(0|z) \right\} + \frac{B_\theta(z)}{b_n^2},$$

where $|B_\theta(z)| \leq B < \infty$ uniformly over θ and z. Then defining $\tilde{\delta}_\theta(z) = \hat{h}_\theta(z) - \tilde{h}_\theta(z)$, and subtracting (2.8) from the previous display equation, we get

$$\tilde{\delta}_\theta(z) = \left(1 - \frac{\beta}{b_n}\right) \left\{ \tilde{\delta}_\theta \left(z - \frac{1}{b_n}\right) \bar{\pi}_\theta(1|z) + \tilde{\delta}_\theta(z) \bar{\pi}_\theta(0|z) \right\} + \frac{B_\theta(z)}{b_n^2}.$$

Now let $\mathcal{Z}(n) = \{1/b_n, 2/b_n, \dots, z_0\}$. From the previous display equation, it follows

$$\sup_{\theta \in \Theta, z \in \mathcal{Z}_n} |\tilde{\delta}_\theta(z)| \leq \left(1 - \frac{\beta}{b_n}\right) \sup_{\theta \in \Theta, z \in \mathcal{Z}_n} |\tilde{\delta}_\theta(z)| + \frac{B}{b_n^2},$$

which implies $\sup_{\theta \in \Theta, z \in \mathcal{Z}_n} |\tilde{\delta}_\theta(z)| \leq B/b_n$ upon rearrangement. So far $\tilde{h}_\theta(.)$ was only defined for multiples of b_n, but we can extend it to all of $[0, z_0]$ by setting $\tilde{h}_\theta(z) = \tilde{h}_\theta(b_n \lfloor z/b_n \rfloor)$. Combining the above with the (uniform) Lipschitz continuity of $\hat{h}_\theta(\cdot)$, we obtain

$$\sup_{\theta \in \Theta, z \in [0, z_0]} \left| \tilde{\delta}_\theta(z) \right| = O\left(\frac{1}{b_n}\right).$$

Suppose that θ were estimated using (2.8) as

$$\tilde{\theta} = arg \max_{\theta \in \Theta} \tilde{h}_\theta(z_0).$$

Then in view of the previous discussion,

$$h_{\theta^*}(z_0) - h_{\tilde{\theta}}(z_0) \leq 2M_1\sqrt{\frac{v}{n}} + 2\sup_{\theta\in\Theta, z\in[0,z_0]}\left|\tilde{\delta}_\theta(z)\right| = 2M_1\sqrt{\frac{v}{n}} + O\left(\frac{1}{b_n}\right).$$

Hence the numerical error from discretization declines at the rate b_n^{-1}. In particular, as long as b_n is chosen to be substantially bigger than $\sqrt{n}$, this approximation error is dwarfed by the statistical error from the regret bound derived in Section 2.2.1.

2.3 General Setup

In this section, we generalize the setup of Section 2.2 in two ways: First we let the waiting times vary with t so that time now becomes a state variable. Second, we allow for the evolution of z to depend on the previous state variable (x, z, t). Among other things, this allows for situations in which the cost of treatment is different for people with different covariates. We do not aim to provide the most general model, but rather envision this as a template to demonstrate our general results. In what follows, we use the 'prime' notation (e.g. x') to denote one-step ahead quantities following the current one (e.g. x).

The state variables are now

$$s := (x, z, t),$$

where x denotes the vector of characteristics or covariates of the individual, z is the institutional variable (e.g. the current budget or capacity), and t is time. As in Section 2.2, the planner has to choose among actions $a = \{0, 1\}$. The choice of the action is determined by a policy function, $\pi_\theta(a|s)$, indexed by θ, that maps the current state s to probabilities over the set of actions:

$$\pi_\theta(.|s) : s \longrightarrow [0, 1].$$

Once an action, a, has been chosen, the planner receives a reward, i.e. a change in social welfare, of $Y(a)$ that is equivalent to the potential outcome of the individual under action a. As in Section 2.2, we assume that $Y(a)$ is affected by the covariates x but not the institutional variable z or time t. Similarly, we also define $r(x, 1) = E[Y(1) - Y(0)|s]$, together with the normalization $r(x, 0) = 0$.

The waiting time $t' - t$ between individuals is random, and distributed as an Exponential random variable with parameter $\lambda(t)$:

$$N(t' - t) \sim \text{Exponential}(\lambda(t)),$$

where N is a normalization term, defined akin to Section 2.2 as the expected discounted number of individuals that the planner would face. In contrast to Section 2.2, the hazard rate $\lambda(t)$ is now allowed to change with t. Additionally, the previous state s, and action a

determine the future values of the institutional variable as

$$N(z' - z) = G_a(x, t, z),$$

where $G_a(.)$ is some known, deterministic function. For example, in the setup of Section 2.2,

$$G_a(x, t, z) = \begin{cases} -1 & \text{if } a = 1 \text{ and } z > 0 \\ 0 & \text{if } a = 0. \end{cases} \tag{2.9}$$

Our current setup is more general and can allow for situations where the cost of the treatment varies with individual, or with time. Finally, the distribution of the covariates is given by

$$x \sim F,$$

where F is fixed and does not change with t or z. The program starts with an initial value of $z = z_0$ at $t = t_0$, and ends when $z \leq 0$.

It is useful to note here the assumptions implicit in our model. First, the rewards are independent of time. This rules out cases where the effect on an action for each individual may change with the arrival time of the individual. Note that rewards may still indirectly depend on time if the distribution of the covariates changes with time. This assumption is mainly made for convenience. Where estimation of time varying rewards is possible, one can estimate $r(x, t, a)$, and our results go through with some modifications. Note also that we prohibit past values of x from affecting the distribution of x'. This rules out network effects, for instance.

More importantly, our setup rules out situations where individuals can strategically respond to the social planner's policy, for example by arriving at different times. Indeed, the waiting times and distribution of covariates are assumed to be independent of all state variables except time. This is reasonable in some contexts, such as in unemployment dynamics where the date of termination is not under complete control of the individual. Alternatively, this is also reasonable if the policy under consideration is incentive compatible. For instance, if the social planner were to pool individuals by placing them in a queue for a short while, this would reduce the incentive for any single individual to change his/her arrival time.

Finally, we also do not allow the distribution of covariates to vary with time. This is a restrictive assumption, and imposed mainly for ease of exposition and derivations. In Section 2.6.2, we show how one could relax this assumption.

Define $V_\theta(s)$ as the expected discounted value of all normalized future rewards $r(x, a)/N$, starting from state s, and when the planner chooses actions according to π_θ. We let $g_{\lambda(t)}(.)$ denote the probability density function of the exponential distribution with parameter $\lambda(t)$.

$V_\theta(.)$ can be obtained as the fixed point of the recursive equation:

$$\begin{aligned}V_\theta(s) =& \frac{r(x,1)\pi_\theta(1|x,z,t)}{N} \\ &+ \int e^{-\beta\frac{\omega}{N}} E_{x'\sim F}\left[V_\theta\left(x', z + \frac{G_1(x,t,z)}{N}, t + \frac{\omega}{N}\right)\pi(1|x,z,t) + \dots\right. \\ &\left.\dots + V_\theta\left(x', z + \frac{G_0(x,t,z)}{N}, t + \frac{\omega}{N}\right)\pi(0|x,z,t)\right] g_{\lambda(t)}(\omega)d\omega, \quad \text{for } z > 0\end{aligned} \tag{2.10}$$

together with

$$V_\theta(s) = 0 \quad \text{for } z = 0.$$

As in Section 2.2, it will be more convenient to work with the integrated value function $h_\theta(z,t) = E_{x\sim F}[V_\theta(x,z,t)]$. The recursive equation for $h_\theta(z,t)$ is given by

$$\begin{aligned}h_\theta(z,t) =& \frac{E_{x\sim F}[r(x,1)\pi_\theta(1|x,z,t)]}{N} \\ &+ \int e^{-\beta\frac{\omega}{N}} E_{x\sim F}\left[h_\theta\left(z + \frac{G_1(x,t,z)}{N}, t + \frac{\omega}{N}\right)\pi(1|x,z,t) + \dots\right. \\ &\left.\dots + h_\theta\left(z + \frac{G_0(x,t,z)}{N}, t + \frac{\omega}{N}\right)\pi(0|x,z,t)\right] g_{\lambda(t)}(\omega)d\omega, \quad \text{for } z > 0\end{aligned} \tag{2.11}$$

together with

$$h_\theta(z,t) = 0 \quad \text{for } z = 0.$$

The above expression considerably simplifies if one considers the continuous case obtained as the limit of $N \to \infty$. To this end, let us define the quantities

$$\bar{r}_\theta(z,t) := E_{x\sim F}[r(x,1)\pi_\theta(1|x,z,t)]$$

and

$$\bar{G}_\theta(z,t) := E_{x\sim F}\left[G_1(x,z,t)\pi_\theta(1|x,z,t) + G_0(x,z,t)\pi_\theta(0|x,z,t)\right].$$

Now consider subtracting $h_\theta(z,t)\int e^{-\beta\frac{\omega}{N}} g_{\lambda(t)}(\omega)d\omega$ from both sides of equation (2.11), multiplying both sides by N, and taking the limit as $N \to \infty$. This leads (after some re-arrangement of terms) to the following Partial Differential Equation (PDE) for the evolution of $h_\theta(z,t)$:

$$\lambda(t)\bar{G}_\theta(z,t)\partial_z h_\theta(z,t) + \partial_t h_\theta(z,t) - \beta h_\theta(z,t) + \lambda(t)\bar{r}_\theta(z,t) = 0, \quad h_\theta(0,t) = 0\ \forall\ t \tag{2.12}$$

where ∂_z, ∂_t are the partial differential operators, and $h_\theta(0,t) = 0$ is the boundary condition for the PDE. The 'derivation' of (2.12) from (2.11) should be considered heuristic, the rigorous justification is given by Theorem 3. For now, we will take PDE (2.12) to be the law of motion governing the evolution of $h_\theta(z,t)$ and assess its properties directly.

Unfortunately, for nonlinear PDEs of the form (2.12) above, it is unclear whether a classical solution (i.e. a solution $h_\theta(z,t)$ that is continuously differentiable) exist. Consequently, the relevant solution concept that we need to employ here is that of a viscosity solution (Crandall and Lions, 1983), which allows for lack of differentiability while still satisfying (2.12) in a generalized sense. This is a common solution concept for equations of the HJB form; we refer to Crandall, Ishii and Lions (1992) for a user's guide, and Achdou et al. (2017) for a useful discussion. In our context, existence of a unique viscosity solution to (2.12) requires the following conditions:

Assumption 1

(i) *$\bar{G}_\theta(z,t)$ and $\bar{r}_\theta(z,t)$ are Lipschitz continuous in (z,t) uniformly over θ.*

(ii) *$\lambda(t)$ is Lipschitz continuous and uniformly bounded away from 0 for all t.*

(iii) *$\bar{G}_\theta(z,t)$ is uniformly bounded away from 0 for all θ, z, t.*

Assumption 1(i) requires $\bar{G}_\theta(z,t)$ and $\bar{r}_\theta(z,t)$ to be sufficiently smooth. A sufficient condition for this is $\pi_\theta(1|x,t,z)$ is Lipschitz continuous in z,t uniformly over x,θ, but this is still stronger than required. The assumption allows for non-smooth π_θ as long as it can be smoothed by taking expectations over x. Assumption 1(ii) implies that the expected waiting time between arrivals varies smoothly with t and is also bounded from above (recall that the expected waiting times are given by $\lambda(t)^{-1}$). Assumption 1(iii) restricts the policy function class and the functions $G_a(x,z,t)$ so that there is always some expected change to the budget at any given state. This is a mild assumption; as long as there exists a fraction of people that benefit from treatment, it is a dominant strategy to choose a policy that generates a change to the budget.

Lemma 2
Suppose that Assumption 1 holds. Then for each θ there exists a unique viscosity solution $h_\theta(z,t)$ to (2.12).

Proof: see Appendix B.1.

Note that (2.12) define a class of PDEs indexed by θ, the solution to each of which is the integrated value function $h_\theta(z,t)$ from following policy π_θ. The social planner's objective is to choose the optimal policy function parameter, θ^*, that maximizes ex-ante expected utility over the class of the policy functions $\Pi = \{\pi_\theta : \theta \in \Theta\}$:

$$\theta^* = arg \max_{\theta \in \Theta} h_\theta(z_0, t_0), \tag{2.13}$$

where z_0 and t_0 refer to the initial time and budget respectively.

2.3.1 The Sample Version of the Social Planner's Problem

The unknown parameters in the social planner's problem are F, $r(x,a)$ and $\lambda(\cdot)$. As discussed in Section 2.2, the social planner can leverage RCT/observational data to obtain estimates F_n and $\hat{r}(x,a)$ of F and $r(x,a)$. In addition to these, she also now requires an estimate $\hat{\lambda}(t)$ of the waiting times $\lambda(t)$. In some cases, this can be estimated from the RCT data itself.

Alternatively, it is possible to use a second data source and obtain $\hat{\lambda}(t)$ through a Poisson regression, for example. This is feasible in the context of unemployment dynamics.

Given the estimates $F_n, \hat{r}(x,a), \hat{\lambda}(t)$, and the policy $\pi_\theta(.)$, we can plug them in to obtain the quantities

$$\hat{r}_\theta(z,t) \equiv E_{x\sim F_n}[\hat{r}(x,1)\pi_\theta(1|x,z,t)]$$

and

$$\hat{G}_\theta(z,t) = E_{x\sim F_n}\left[G_1(x,z,t)\pi_\theta(1|x,z,t) + G_0(x,z,t)\pi_\theta(0|x,z,t)\right].$$

Based on the above we can construct the sample version of PDE (2.12) as

$$\hat{\lambda}(t)\hat{G}_\theta(z,t)\partial_z\hat{h}_\theta(z,t) + \partial_t\hat{h}_\theta(z,t) - \beta\hat{h}_\theta(z,t) + \hat{\lambda}(t)\hat{r}_\theta(z,t) = 0, \quad \hat{h}_\theta(0,t) = 0 \;\forall\; t. \tag{2.14}$$

A unique solution to PDE (2.14) exists for each θ under analogous conditions to Assumption 1, so we do not repeat them here. The unique solution is $\hat{h}_\theta(z,t)$, the value function under an empirical dynamic environment determined by the quantities $F_n, \hat{r}(x,a), \hat{\lambda}(t)$. As before, one should think of (2.14) as defining a class of PDEs indexed by θ, the solution to each of which is the integrated value function $\hat{h}_\theta(z,t)$ that can be used as an estimate for $h_\theta(z,t)$. Based on these integrated value function estimates, we can now solve a sample version of the social planner's problem as follows:

$$\hat{\theta} = arg\max_{\theta\in\Theta}\hat{h}_\theta(z_0,t_0). \tag{2.15}$$

While the PDE form for $\hat{h}_\theta(z,t)$ is very convenient for our theoretical results, this is not quite practical for computing $\hat{\theta}$. So for estimation we use a discretized version of (2.14), akin to (2.11):

$$\hat{h}_\theta(z,t) = \begin{cases} \frac{\hat{r}_\theta(z,t)}{b_n} + E_{n,\theta}\left[e^{-\beta(t'-t)}\hat{h}_\theta\left(z',t'\right)|z,t\right] & \text{for } z>0 \\ 0 & \text{for } z=0 \end{cases} \tag{2.16}$$

where

$$E_{n,\theta}\left[e^{-\beta(t'-t)}\hat{h}_\theta\left(z',t'\right)|z,t\right] := \int e^{-\beta\frac{\omega}{b_n}} E_{x\sim F_n}\left[\hat{h}_\theta\left(z+\frac{G_1(x,t,z)}{b_n}, t+\frac{\omega}{b_n}\right)\pi(1|x,z,t)\right.$$
$$\left. + \hat{h}_\theta\left(z+\frac{G_0(x,t,z)}{b_n}, t+\frac{\omega}{b_n}\right)\pi(0|x,z,t)\right] g_{\hat{\lambda}(t)}(\omega)d\omega.$$

More generally, for any function f, we let $E_{n,\theta}[f(z',t')|z,t]$ denote the expectation over z',t' conditional on the values of t,z and when following the policy π_θ. Precisely the expectation is joint over three independent probability distributions: (i) The distribution F_n of the covariates, (ii) the probability distribution over the (exponential) waiting time process indexed by $\{\hat{\lambda}(t) : t\in[t_0,\infty)\}$, and (iii) the probability distribution induced on z' due to the randomization of policies using $\pi_\theta(a|s)$.

Note that we have chosen to 'discretize' (2.16) by the factor b_n in analogy with the discussion in Section 2.2.2. One can show that as $b_n\to\infty$, the solution to (2.16) converges to that of (2.14). A formal statement to this effect, along with a bound on the numerical error for a given choice of b_n, is given in Section 2.5.1.

A useful aspect of the estimation of $\hat{\theta}$ is that it is equivalent to solving for the optimal policy function under the sample dynamics described by $F_n, \hat{r}(x,a)$ and $\hat{\lambda}(t)$. Note that in the sample version, both the rewards $\hat{r}(x,a)$, and dynamics (given a policy function $\pi(.)$) are known. This nests the estimation of $\hat{\theta}$ into a standard Reinforcement Learning problem for learning the optimal policy function under a known dynamic environment. In the next section, we apply one such RL algorithm, the Actor-Critic algorithm.

2.4 The Actor-Critic Algorithm

In a standard Reinforcement Learning (RL) framework, an algorithm runs multiple instances, called episodes, of a dynamic environment (or game). At any particular state on any particular episode the algorithm takes an action a according to the current policy function π_θ and observes the reward and the future value of the state. Based on these observed values, it updates the policy parameter to some new value θ' according to a pre-specified criterion. The process then continues with the new updated policy function $\pi_{\theta'}$ until the parameter θ or the cumulative rewards converges.

Estimation of $\hat{\theta}$ in equation (2.16) fits naturally in the above context, since we can simulate a 'sample' dynamic environment as follows: Suppose that the current state is $s \equiv (x, z, t)$, and the policy parameter is θ. The planner chooses an action a according to the policy function $\pi_\theta(a|s)$, which results in a reward of $\hat{r}(x,t)$. The next individual arrives at time $t' = t + \Delta t / b_n$ where

$$\Delta t \sim \text{Exponential}(\hat{\lambda}(t)).$$

New values of the institutional state variables are obtained as $z' = z + \Delta z / b_n$, where

$$\Delta z = G_a(x, t, z).$$

New values of the covariates x' are drawn from the distribution $F_n(.)$, i.e. each individual is drawn with replacement with probability $1/n$ from the sample set of observations. Based on the reward $\hat{r}(x,a)$ and new state (x', z', t'), the policy parameter is updated to a new value θ. This process then repeats indefinitely until θ converges. In between, each time an episode ends, we simply restart a new episode.

In this section, we adapt one of the most widely used RL algorithms - the Actor-Critic algorithm - to our context. We differ from the standard RL approach, however, in employing the integrated value function $\hat{h}_\theta(z,t)$ as the central ingredient of our algorithm instead of the value function $\hat{V}_\theta(s)$ - we explain the rationale for this in Section 2.4.1 below.

Actor-Critic algorithms aim to calculate $\hat{\theta}$ by updating θ at each state of each episode using stochastic gradient descent along the direction $\hat{g}(\theta) \equiv \nabla_\theta \left[\hat{h}_\theta(z_0, t_0)\right]$:

$$\theta \longleftarrow \theta + \alpha_\theta \hat{g}(\theta),$$

where α_θ is the step size parameter or learning rate. Denote by $\hat{Q}_\theta(s,a)$, the action-value function

$$\hat{Q}_\theta(s,a) := \hat{r}_n(x,a) + E_{n,\theta}\left[e^{-\beta(t'-t)}\hat{h}_\theta(z',t')|s,a\right], \tag{2.17}$$

where $\hat{r}_n(x,a) := \hat{r}(x,a)/b_n$. The Policy-Gradient theorem (see e.g. Sutton et al., 2000) provides an expression for $\hat{g}(\theta)$ as

$$\hat{g}(\theta) = E_{n,\theta}\left[e^{-\beta(t-t_0)}\hat{Q}_\theta(s,a)\nabla_\theta \ln \pi(a|s;\theta)\right],$$

where $E_{n,\theta}[.]$ in this context denotes the expectation over the (stationary) distribution of the states s and actions a induced by the policy function π_θ in the (sample) dynamic environment of Section 2.3.1. A well known result (see e.g. Sutton and Barto, 2018) is that

$$E_{n,\theta}\left[e^{-\beta(t-t_0)}\hat{Q}_\theta(s,a)\nabla_\theta \ln \pi(a|s;\theta)\right] = E_{n,\theta}\left[e^{-\beta(t-t_0)}\left(\hat{Q}_\theta(s,a) - b(s)\right)\nabla_\theta \ln \pi(a|s;\theta)\right]$$

for any arbitrary 'baseline' $b(.)$ that is a function of s. Let $\dot{h}_\theta(z,t)$ denote some functional approximation for $\hat{h}_\theta(z,t)$. We exploit the fact that the continuation value of the state-action pair only depends on z,t, and therefore use $\dot{h}_\theta(z,t)$ as the baseline, which gives us

$$\hat{g}(\theta) = E_{n,\theta}\left[e^{-\beta(t-t_0)}\left(\hat{Q}_\theta(s,a) - \dot{h}_\theta(z,t)\right)\nabla_\theta \ln \pi(a|s;\theta)\right].$$

The above is infeasible since we don't know $\hat{Q}_\theta(s,a)$. However we can heuristically approximate $\hat{Q}_\theta(s,a)$ with the one step 'bootstrap' return as suggested by equation (2.17) (here the term 'bootstrap' refers to its usage in the RL literature, see Sutton and Barto, 2018):

$$R^{(1)}(x,a) = \hat{r}_n(x,a) + \mathbb{I}_{\{z'>0\}}e^{-\beta(t'-t)}\dot{h}_\theta(z',t'),$$

This enables us to obtain an approximation for $\hat{g}(\theta)$ as

$$\hat{g}(\theta) \approx \tilde{g}(\theta) = E_{n,\theta}\left[e^{-\beta(t-t_0)}\delta_n(s,s',a)\nabla_\theta \ln \pi(a|s;\theta)\right], \tag{2.18}$$

where $\delta_n(s,s',a)$ is the Temporal-Difference (TD) error defined as

$$\delta_n(s,s',a) := \hat{r}_n(x,a) + \mathbb{I}_{\{z'>0\}}e^{-\beta(t'-t)}\dot{h}_\theta(z',t') - \dot{h}_\theta(z,t).$$

We now describe the functional approximation for $\hat{h}_\theta(z,t)$. Let $\phi_{z,t} = (\phi_{z,t}^{(j)}, j = 1,\ldots,d_\nu)$ denote a vector of basis functions of dimension d_ν over the space of z,t. For the sake of argument, consider approximating $\hat{h}_\theta(z,t)$ by choosing the weights ν to minimize the infeasible mean squared error criterion:

$$arg\min_\nu \hat{S}(\nu|\theta) \equiv arg\min_\nu E_{n,\theta}\left[e^{-\beta(t-t_0)}\left\|\hat{h}_\theta(z,t) - \nu^\intercal\phi_{z,t}\right\|^2\right].$$

Then we can update the value function weights, ν, using gradient descent

$$\nu \longleftarrow \nu + \alpha_\nu \nabla_\nu \hat{S}(\nu|\theta)$$

for some value function learning rate α_ν. Here the gradient is given by

$$\hat{\chi}(\nu|\theta) := \nabla_\nu \hat{S}(\nu|\theta) \propto E_{n,\theta}\left[e^{-\beta(t-t_0)}\left(\hat{h}_\theta(z,t) - \nu^\intercal \phi_{z,t}\right)\phi_{z,t}\right].$$

The above procedure is infeasible since $\hat{h}_\theta(z,t)$ is unknown. However, as before, we can heuristically approximate $\hat{h}_\theta(z,t)$ using the one step bootstrap return $R^{(1)}$ and obtain

$$\hat{\chi}(\nu|\theta) \approx E_{n,\theta}\left[e^{-\beta(t-t_0)}\delta_n(s,s',a)\phi_{z,t}\right]. \tag{2.19}$$

The heuristic for the bootstrap approximation above is based on equation (2.16), which implies that an unbiased estimator of $\hat{h}_\theta(z,t)$ is given by sum of the current reward $\hat{r}_n(x,a)$, and the discounted future value of $\hat{h}(z',t')$.

Using equations (2.18) and (2.19), we can now construct stochastic gradient updates for θ, ν as

$$\theta \longleftarrow \theta + \alpha_\theta e^{-\beta(t-t_0)}\delta_n(s,s',a)\nabla_\theta \ln \pi(a|s;\theta) \tag{2.20}$$

$$\nu \longleftarrow \nu + \alpha_\nu e^{-\beta(t-t_0)}\delta_n(s,s',a)\phi_{z,t}, \tag{2.21}$$

by replacing the expectations in (2.18), (2.19) with their corresponding unbiased estimates obtained from the values of state variables as they come up in each episode. Importantly, the updates (2.20) and (2.21) can be applied simultaneously on the same set of current state values, as long as $\alpha_\nu \gg \alpha_\theta$. This is an example of two-timescale stochastic gradient decent: the parameter with the lower value of the learning rate is said to be updated at the slower time scale. When the timescale for ν is much faster than that for θ, one can imagine that the value of ν has effectively converged to the value function estimate for current policy parameter θ. Thus we can proceed with updating θ as if its corresponding (approximate) value function were already known.

The pseudo-code for this procedure is presented in Algorithm 1.

2.4.1 Basis Dimensions and Integrated Value Functions

The functional approximation for $\hat{h}_\theta(z,t)$ involves choosing a vector of bases $\phi_{z,t}$ of dimension d_ν. From a statistical point of view, the optimal choice of d_ν is in fact infinity. There is no bias-variance tradeoff since we would like to compute $\hat{h}_\theta(z,t)$ exactly. We can simply take as high a value of d_ν as computationally feasible. This useful property is a consequence of employing $\hat{h}_\theta(z,t)$ rather than $\hat{V}_\theta(s)$ in the Actor-Critic algorithm. Since $\hat{r}(x,a)$ could be a function of Y (as with doubly robust estimators, for example), we would need some regularization if we try to obtain a functional approximation for $\hat{V}_\theta(s)$, to ensure we don't overfit to the outcome data. This is not an issue for $\hat{h}_\theta(z,t)$, however, as it only involves the expected value of $\hat{r}(x,a)$ given z,t. Thus by using $\hat{h}_\theta(z,t)$ we are able to avoid an additional regularization term.

Algorithm 1: Actor-Critic

Initialise policy parameter weights $\theta \leftarrow 0$

Initialise value function weights $\nu \leftarrow 0$

Repeat forever:

Reset budget: $z \leftarrow z_0$

Reset time: $t \leftarrow t_0$

$I \leftarrow 1$

While $z > 0$**:**

$x \sim F_n$ (Draw new covariate at random from data)

$a \sim \pi(a|s;\theta)$ (Draw action, note: $s = (x,z,t)$)

$R \leftarrow \hat{r}(x,a)$ (with $R = 0$ if $a = 0$)

$\Delta t \sim \text{Exponential}(\hat{\lambda}(t))$ (Draw time increment)

$t' \leftarrow t + \Delta t / b_n$

$z' \leftarrow z + G_a(x,z,t)/b_n$

$\delta \leftarrow R + \mathbb{I}\{z' > 0\} e^{-\beta(t'-t)} \nu^\intercal \phi_{z',t'} - \nu^\intercal \phi_{z,t}$ (Temporal-Difference error)

$\theta \leftarrow \theta + \alpha_\theta I \delta \nabla_\theta \ln \pi(a|s;\theta)$ (Update policy parameter)

$\nu \leftarrow \nu + \alpha_\nu I \delta \phi_{z,t}$ (Update value parameter)

$z \leftarrow z'$

$t \leftarrow t'$

$I \leftarrow e^{-\beta(t'-t)} I$

2.4.2 Convergence of the Actor-Critic Algorithm

Our proposed algorithm differs from the standard versions of the Actor-Critic algorithm in only using the integrated value function. Consequently, its convergence follows by essentially the same arguments as that employed in the literature for actor-critic methods, see e.g. Bhatnagar et al. (2009). In this section, we restate their main results, specialized to our context. Since all of the convergence proofs in the literature are obtained for discrete Markov states, we need to impose the technical device of discretizing time and making it bounded, so that the states are now discrete (the other terms z and x are already discrete, the latter since we use empirical data). This greatly simplifies the convergence analysis, but does not appear to be needed in practice.

Let $\mathcal{S}$ denote the set of all possible values of (z,t), after discretization. Also, denote by Φ, the $|\mathcal{S}| \times d_\nu$ matrix whose ith column is $(\phi_{z,t}^{(i)}, (z,t) \in \mathcal{S})^\intercal$, where $\phi_{z,t}^{(i)}$ is the ith element of $\phi_{z,t}$.

Assumption 2

(i) *$\pi_\theta(a|s)$ is continuously differentiable in θ for all s, a.*

(ii) *The basis functions $\{\phi_{z,t}^{(i)} : i : 1, \ldots, d_\nu\}$ are linearly independent, i.e. Φ has full rank. Also, for any vector ν, $\Phi\nu \neq e$, where e is the $\mathcal{S}$-dimensional vector with all entries equal to one.*

(iii) *The learning rates satisfy $\sum_k \alpha_\nu^{(k)} \to \infty$, $\sum_k \alpha_\nu^{(k)2} < \infty$, $\sum_k \alpha_\theta^{(k)} \to \infty$, $\sum_k \alpha_\theta^{(k)2} < \infty$ and $\alpha_\theta^{(k)}/\alpha_\nu^{(k)} \to 0$ where $\alpha_\theta^{(k)}, \alpha_\nu^{(k)}$ denote the learning rates after k steps/updates of the algorithm.*

(iv) *The update for θ is bounded i.e.*

$$\theta \longleftarrow \Gamma\left(\theta + \alpha_\theta \delta_n(s, s', a)\nabla_\theta \ln \pi(a|s;\theta)\right)$$

where $\Gamma : \mathbb{R}^{dim(\theta)} \to \mathbb{R}^{dim(\theta)}$ is a projection operator such that $\Gamma(x) = x$ for $x \in C$ and $\Gamma(x) \in C$ for $x \notin C$, where C is any compact hyper-rectangle in $\mathbb{R}^{dim(\theta)}$.

Differentiability of π_θ with respect to θ is a minimal requirement for all Actor-Critic methods. Assumption 2(ii) is also mild and rules out multicollinearity in the basis functions for the value approximation. Assumption 2(iii) places conditions on learning rates that are standard in the literature of stochastic gradient descent with two timescales. Assumption 2(iv) is a technical condition imposing boundedness of the updates for θ. This is an often used technique in the analysis of stochastic gradient descent algorithms. Typically this is not needed in practice, though it may sometimes be useful to bound the updates when there are outliers in the data.

Define $\mathcal{Z}$ as the set of local maxima of $J(\theta) \equiv \hat{h}_\theta(z_0, t_0)$, and $\mathcal{Z}^\epsilon$ an ϵ-expansion of that set. Also, $\theta^{(k)}$ denotes the k-th update of θ. We then have the following theorem on the convergence of our Actor-Critic algorithm.

Theorem 1
Suppose that Assumption 2 holds and additionally that $\nabla_\theta \pi_\theta(s)$ is uniformly Hölder continuous in s. Then, for each $\epsilon > 0$, there exists M such that if $d_\nu \geq M$, then $\theta^{(k)} \to \mathcal{Z}^\epsilon$ with probability 1 as $k \to \infty$.

The above theorem is for the most part a direct consequence of the results of Bhatnagar et al. (2009). We provide further discussion and a justification of the result in Appendix B.1.

2.4.3 Parallel Updates

While Theorem 1 assures convergence of our algorithm, in practice the updates could be volatile and may take a long time to converge. Much of the reason for this is the correlation between the updates as one cycles through each episode - indeed, note that the state pairs (s, s') are highly correlated. Hence the stochastic gradients become correlated and one needs many episodes to move in the direction of the true (i.e. the expected) gradient.

Algorithm 2: Parallel Actor-Critic

Initialise policy parameter weights $\theta \leftarrow 0$
Initialise value function weights $\nu \leftarrow 0$
Batch size B

For $p = 1, 2, ...$ processes, launched in parallel, each using and updating the same global parameters θ and ν:

Repeat forever:

Reset budget: $z \leftarrow z_0$
Reset time: $t \leftarrow t_0$
$I \leftarrow 1$

While $z > 0$**:**

$\theta_p \leftarrow \theta$ (Create local copy of θ for process p)

$\nu_p \leftarrow \nu$ (Create local copy of ν for process p)

batch_policy_upates $\leftarrow 0$

batch_value_upates $\leftarrow 0$

For $b = 1, 2, ..., B$**:**

$x \sim F_n$ (Draw new covariate at random from data)

$a \sim \pi(a|s; \theta_p)$ (Draw action, note: $s = (x, z, t)$)

$R \leftarrow \hat{r}(x, a)$ (with $R = 0$ if $a = 0$)

$\Delta t \sim \text{Exponential}(\hat{\lambda}(t))$ (Draw time increment)

$t' \leftarrow t + \Delta t / b_n$

$z' \leftarrow z + G_a(x, z, t)/b_n$

$\delta \leftarrow R + \mathbb{I}\{z' > 0\} e^{-\beta(t'-t)} \nu_p^\intercal \phi_{z',t'} - \nu_p^\intercal \phi_{z,t}$ (TD error)

batch_policy_upates $\leftarrow$ batch_policy_upates $+ \alpha_\theta I \delta \nabla_\theta \ln \pi(a|s; \theta_p)$

batch_value_upates $\leftarrow$ batch_value_upates $+ \alpha_\nu I \delta \phi_{z,t}$

$z \leftarrow z'$

$t \leftarrow t'$

$I \leftarrow e^{-\beta(t'-t)} I$

If $z \leq 0$, break **For**

Globally update: $\nu \leftarrow \nu +$ batch_value_upates$/B$
Globally update: $\theta \leftarrow \theta +$ batch_policy_upates$/B$

This is a common problem for all Actor-Critic algorithms, but recently Mnih et al. (2015) have proposed to solve this through the use of asynchronous parallel updates. The key idea is to run multiple versions of the dynamic environment on parallel threads or processes, each of which independently and asynchronously updates the shared global parameters θ and v. Since at any given point in time, the parallel threads are at a different point in the dynamic environment (they are started with slight offsets), successive updates are decorrelated. And as an additional benefit, the algorithm is faster by dint of being run in parallel.

Algorithm 2 provides the pseudo-code for parallel updating. It also amends the previous version of the algorithm by adding batch updates. In batch updating, the researcher chooses a batch size B such that the parameter updates occur only after averaging over B observations. This usually results in a smoother update trajectory because extreme values of the updates are averaged out.

2.5 Statistical and Numerical Properties

In this section, we analyze the statistical and numerical properties of the estimated welfare maximizing policy functions. The main result of this section is a probabilistic bound on the regret defined as the maximal difference between the integrated value functions $h_{\hat{\theta}}(z_0, t_0)$ and $h_{\theta^*}(z_0, t_0)$. We derive this using our bound on the maximal difference in the value functions

$$\sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta} |\hat{h}_\theta(z,t) - h_\theta(z,t)| \tag{2.22}$$

since

$$h_{\theta^*}(z_0, t_0) - h_{\hat{\theta}}(z_0, t_0) \le \sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta} 2|\hat{h}_\theta(z,t) - h_\theta(z,t)|.$$

The bound for (2.22) will depend on the sample size n and the complexity of the policy function class, indexed by the Vapnik-Cervonenkis (VC) dimension of some specific collections of functions.

We maintain Assumption 1 that is required for the existence of the value functions. In addition, we impose the following:

Assumption 3

(i) *(Bounded Rewards) There exists $M < \infty$ such that $|Y(0)|, |Y(1)| \le M$.*

(ii) *(Complexity of the Policy Function Space) The collections of functions*

$$\mathcal{I} = \{\pi_\theta(1|\cdot, z, t) : z \in [0, z_0], t \in [t_0, \infty), \theta \in \Theta\}$$

over the covariates x, and which are indexed by z, t and θ, is a VC-subgraph class with finite VC index v_1. Furthermore, for each $a = 0, 1$, the collection of *functions*

$$\mathcal{G}_a = \{\pi_\theta(a|\cdot, z, t) G_a(\cdot, z, t) : z \in [0, z_0], t \in [t_0, \infty), \theta \in \Theta\}$$

over the covariates x is also a VC-subgraph class with finite VC index v_2. We shall let $v = \max\{v_1, v_2\}$.

Assumption 3(i) ensures that the rewards are bounded. This is a common assumption in the treatment effect literature (see e.g. Kitagawa and Tetenov, 2018) and imposed mainly for ease of deriving the theoretical results. Assumption 3(ii) has already been discussed in some detail in Section 2.2. We only add here that in many of the examples we consider $G_a(x, z, t)$ is independent of x, as in equation (2.9) for example. In this case $v_1 = v_2$.

The next set of assumptions relate to the properties of the observational or RCT dataset from which we estimate $\hat{r}(x, a)$.

Assumption 4

(i) *(IID Draws from F) The observed data is an iid draw of size n from the distribution F.*

(ii) *(Selection on Observables) $(Y(1), Y(0)) \perp W|X$.*

(iii) *(Strict Overlap) There exists $\kappa > 0$ such that $p(x) \in [\kappa, 1-\kappa]$ for all x.*

Assumption 4(i) assumes that the observed data is representative of the entire population. If the observed population only differs from F in terms of the distribution of some observed covariates, we can reweigh the rewards, and our theoretical results continue to apply. Assumption 4(ii) assumes that the observed data is taken from an observational study that satisfies unconfoundedness. In Section 2.6.1, we consider extensions to non-compliance. Assumption 4(iii) ensures that the propensity scores are strictly bounded away from 0 and 1. Both Assumptions 4(ii) and 4(iii) are directly satisfied in the case of RCT data.

Under Assumptions 3 and 4, one can propose many different estimates of the rewards $\hat{r}(x, 1)$ that are consistent for $r(x, 1)$. In this paper we recommend doubly robust estimates. As described in Section 2.2, an example of a doubly robust estimate of the reward is

$$\hat{r}(x, 1) = \hat{\mu}(x, 1) - \hat{\mu}(x, 0) + (2W - 1)\frac{Y - \hat{\mu}(x, W)}{W\hat{p}(x) + (1 - W)(1 - \hat{p}(x))}, \tag{2.23}$$

where $\hat{\mu}(x, w)$ and $\hat{p}(x)$ are non-parametric estimates of $\mu(x, w)$ and $p(x)$. To simplify matters, we shall assume that these non-parametric estimates are obtained through cross-fitting (Chernozhukov et al., 2018a). This is done as follows: We divide the data randomly divided into K folds of equal size, and for each fold j, we run a machine learning estimator of our choice on the other $K-1$ folds to estimate $\hat{\mu}^{(-j)}(x, w)$ and $\hat{p}^{(-j)}(x)$. Then for any observation

x_j in some fold j, we set $\hat{\mu}(x_j, w) = \hat{\mu}^{(-j)}(x_j, w)$ and $\hat{p}(x_j) = \hat{p}^{(-j)}(x_j)$. We employ cross-fitting estimators as they require minimal assumptions. Additionally, they have excellent bias properties as demonstrated by Chernozhukov et al., (2018a), and Athey and Wager (2018). We impose the following high level conditions for the machine learning methods used in our cross-fitted estimates:

Assumption 5

(i) *There exists an $a > 0$ such that for $w = 0, 1$*

$$\sup_x |\hat{\mu}(x, w) - \mu(x, w)| = O_p(n^{-a}), \quad \sup_x |\hat{p}(x) - p(x)| = O_p(n^{-a}).$$

(ii) *(L^2 convergence) There exists some $\xi > 1/2$ such that*

$$E\left[|\hat{\mu}(x, w) - \mu(x, w)|^2\right] \lesssim n^{-\xi}, \qquad E\left[|\hat{p}(x) - p(x)|^2\right] \lesssim n^{-\xi}.$$

Assumption 5 is taken from Athey and Wager (2018). The requirements imposed are weak and satisfied by almost all non-parametric estimators including series regression or lasso. Using Assumptions 1-5, one can show that the quantities $\hat{r}_\theta(z, t), \hat{G}_\theta(z, t)$ are uniformly close to $\bar{r}_\theta(z, t), \bar{G}_\theta(z, t)$. In particular, there exists a universal constant C_0 such that

$$E\left[\sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta} \|\hat{r}_\theta(z, t) - \bar{r}_\theta(z, t)\|\right] \le C_0\sqrt{\frac{v_1}{n}} \text{ and} \tag{2.24}$$
$$E\left[\sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta} \left\|\hat{G}_\theta(z, t) - \bar{G}_\theta(z, t)\right\|\right] \le C_0\sqrt{\frac{v_2}{n}}.$$

The above inequalities are based on the work of Kitagawa and Tetenov (2018), and Athey and Wager (2018).

Our final assumption is on the estimation of $\lambda(t)$.

Assumption 6
$\sup_t \left|\hat{\lambda}(t) - \lambda(t)\right| \lesssim \sqrt{v/n}$ *with probability approaching one under the probability distribution $\Omega_\lambda(\mathbb{T})$ over the waiting time process indexed by $\{\lambda(t) : t \in [t_0, \infty)\}$.*

Assumption 6 is made solely in order to obtain a $\sqrt{v/n}$ bound on the regret.[83] Clearly, this is satisfied if one has a parametric model for $\lambda(t)$. Alternatively, in some contexts, e.g. unemployment dynamics, one might estimate $\lambda(t)$ from a second data source that contains information on arrival rates, with the size of this second dataset being much larger that n. In general, however, estimation of $\lambda(t)$ is an exercise in forecasting, and Assumption 6 a statement on the accuracy of the forecasts. Clearly, the further we go into the future, the less confident we would be in our forecast, and the less likely it is that Assumption 6 holds.

[83]On the other hand, if $\sup_t \left|\hat{\lambda}(t) - \lambda(t)\right| \lesssim b(n)$ with $b(n)$ slower than $\sqrt{v/n}$, we can show by the same techniques as the proof of Theorem 1 that the regret will be of the order $b(n)$ as well.

From (2.24) and Assumption 6, we find that the parameters characterizing the PDEs (2.12) and (2.14) are uniformly close to each other. This indicates that the solutions to these PDEs should also be uniformly close. Before we present a formal statement to this effect, we present here a heuristic derivation assuming a classical solution exists and satisfies

$$\sup_{z,t,\theta} |\partial_t h_\theta(z,t)| < \infty. \tag{2.25}$$

Denote $\hat{\delta}_\theta(z,t) = \hat{h}_\theta(z,t) - h_\theta(z,t)$. Then from (2.12) and (2.14), we have

$$\partial_z \hat{\delta}_\theta(z,t) + \frac{1}{\lambda(t)\bar{G}_\theta(z,t)} \partial_t \hat{\delta}_\theta(z,t) - \frac{\beta \hat{\delta}_\theta(z,t)}{\lambda(t)\bar{G}_\theta(z,t)} = \hat{A}_\theta(z,t), \quad \hat{\delta}_\theta(0,t) = 0 \ \forall \ t \tag{2.26}$$

where[84]

$$\sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta} \left\| \hat{A}_\theta(z,t) \right\| \leq C\sqrt{v/n}$$

for some $C < \infty$ under Assumptions 1-6 and (2.25). The derivation of the last inequality in particular makes use of (2.24). Now, equations of the kind (2.26) above can be converted into ones involving only ordinary differentials using the technique of characteristic curves. Intuitively, characteristic curves enable us to 'propagate' the solution from the boundary condition $\hat{\delta}_\theta(0,t) = 0$, by converting an equation that involves partials differentials to one involving ordinary differentials along each characteristic curve. For (2.26), the relevant characteristic curves are given by $t = \Gamma_c(z)$, where each curve - indexed by a scalar $c \in \mathbb{R}$ which determines the initial condition - is defined as the solution to the ODE[85]

$$\frac{d\Gamma_c(z)}{dz} = \frac{1}{\lambda(\Gamma_c(z))\bar{G}_\theta(z,\Gamma_c(z))}, \quad \Gamma_c(0) = c. \tag{2.27}$$

Now for each c, denote $\hat{u}_\theta(z;c) := \hat{\delta}_\theta(z,\Gamma_c(z))$. Then by differentiating $\hat{u}_\theta(z;c)$ with respect to z, and using (2.26),(2.27), we obtain the following equation for the behavior along each characteristic curve,

$$\partial_z \hat{u}_\theta(z;c) - \frac{\beta}{\lambda(\Gamma_c(z))\bar{G}_\theta\left(z,\Gamma_c(z)\right)} \hat{u}_\theta(z;c) = \hat{A}_\theta\left(z,\Gamma_c(z)\right), \quad \hat{u}_\theta(0;c) = 0. \tag{2.28}$$

Since (2.28) is now in the form of an ODE, we can apply Grönwall's inequality as outlined in Section 2.2.1 to show that $|\hat{u}_\theta(z;c)| \leq C\sqrt{v/n}$ uniformly over all the possible values of (θ,z,c), where $C < \infty$ is some constant. But since every vector $(z,t) \equiv (z,\Gamma_c(z))$ for some c, this implies that $\left|\hat{\delta}_\theta(z,t)\right| \leq C\sqrt{v/n}$ uniformly over all θ, z, t.

As noted previously, the derivation is only heuristic; the formal proof makes use of the properties of viscosity solutions. The following is our main theorem of this section:

[84] All inequalities in this discussion should be understood as holding with probability approaching one under the joint distribution of $F \times \Omega_\lambda(\mathbb{T})$.

[85] Under Assumption 1, the Picard-Lindelöf theorem guarantees existence of a unique solution for each c.

Theorem 2

Suppose that Assumptions 1-6 hold. Then with probability approaching one under joint the probability distribution $F \times \Omega_\lambda(\mathbb{T})$,

$$\sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta} \left| e^{-\beta(t-t_0)} \left\{ \hat{h}_\theta(z,t) - h_\theta(z,t) \right\} \right| \leq C\sqrt{\frac{v}{n}}.$$

Consequently,

$$h_{\theta^*}(z_0,t_0) - h_{\hat{\theta}}(z_0,t_0) \leq 2C\sqrt{\frac{v}{n}}.$$

Additionally, the above statements hold uniformly over all $F \times \Omega_\lambda(\mathbb{T})$ if similarly uniform versions of Assumptions 5 and 6 hold.

We prove Theorem 2 in Appendix B.1 by verifying the conditions in Souganidis (1985, Proposition 1.4).

Though we do not formally show this, it seems likely that the $\sqrt{n}$ rate for the regret $h_{\theta^*}(z_0,t_0) - h_{\hat{\theta}}(z_0,t_0)$ cannot be improved upon, especially since Kitagawa and Tetenov (2018) show that this rate is optimal in the static case. At the same time, we do not claim that the VC dimension v in the rate is necessarily tight.[86]

2.5.1 Approximation and Numerical Convergence

In Section 2.3.1, we pointed out that for computation, it is preferable to use an approximate version of PDE (2.14), given by (2.16). Indeed our algorithm in Section 2.4 was based on this. Implementing this algorithm requires choosing a 'approximation' factor b_n. Here we characterize the numerical error resulting from any particular choice of b_n. This is the PDE counterpart of the analysis in Section 2.2.2.

In order to bound the numerical error, we make the additional assumption that $\hat{G}_\theta(z,t)$ is strictly negative. This is because the machinery for viscosity solutions typically requires monotonicity with respect to $h_\theta(z,t)$. We were able to avoid this previously by getting rid of the $h_\theta(z,t)$ term through a change of variables in the proofs of Lemma 2 and Theorem 2. But this is not viable when working with the approximating equation (2.16). Note that in our running example with job training, changes to budget are always negative, so this assumption is obviously satisfied here.

For each $\theta \in \Theta$, denote by $\tilde{h}_\theta(z,t)$ the solution to (2.16), and by $\hat{h}_\theta(z,t)$ the solution to (2.14).

[86] Ideally, one would like to restrict the policy function class in Assumption 3(ii) to the lower dimensional subspace of the values of (z,t) that are 'typically' encountered in following the optimal policies π_{θ^*} and $\pi_{\hat{\theta}}$ (i.e. optimal for the true and sample dynamic environments respectively).

Theorem 3

Suppose that Assumptions 1-6 hold. Assume further that $\hat{G}_\theta(\cdot)$ is strictly negative for all θ, and $|\hat{\lambda}(\cdot)\hat{G}_\theta(.)| \leq M < \infty$. Then, with probability approaching one under the joint distribution of $F \times \Omega_\lambda(\mathbb{T})$, there exists $K < \infty$ independent of θ, z, t such that

$$\sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta} \left|\hat{h}_\theta(z,t) - \tilde{h}_\theta(z,t)\right| \leq K\sqrt{\frac{1}{b_n}}.$$

We prove Theorem 3 in Appendix B.1. The upper bound, which is of the order $b_n^{-1/2}$, appears to be sharp under our assumptions. We refer to Krylov (2005) for some results in this direction. Note that this is of a smaller order than the rate of b_n^{-1} we obtained in Section 2.2.2 for ODEs. One can understand this difference as the price for dealing with solutions $\hat{h}_\theta(z,t)$ that are not differentiable everywhere, but are only valid in a viscosity sense.

Let $\tilde{\theta}$ denote the numerical approximation to $\hat{\theta}$, obtained as the solution to

$$\tilde{\theta} = arg \max_{\theta\in\Theta} \tilde{h}_\theta(z_0,t_0).$$

By a direct application of Theorems 2 and 3,

$$h_{\theta^*}(z_0,t_0) - h_{\tilde{\theta}}(z_0,t_0) \leq 2C\sqrt{\frac{v}{n}} + 2K\sqrt{\frac{1}{b_n}}.$$

Hence, as a rule of thumb, we recommend setting b_n to be some multiple of, or exactly equal to n. One could then try out a few different values, $b_n, 2b_n$ etc to make sure the solution does not change too much.

2.6 Extensions

2.6.1 Non-Compliance

As it is the case for our example in Section 2.7, a common issue in practice is that there is substantial non-compliance. Here we show how our methods can be modified to account for this. We assume that the treatment behaves similarly to a monotone instrumental variable in that we can partition individuals into three categories: compliers, always-takers, and never-takers.

We assume that the social planner cannot change the compliance behavior of any individual. Then the only category of people for whom a social planner can affect a welfare change are the compliers. As for the always-takers and never-takers, the planner has no control over their choices, so its equivalent to assume that the planner would always treat the former and never treat the latter. Formally, the change in reward (conditional on the covariates) for the social planner from treating an individual i, as compared to not treating is

$$r(x_i, 1) = \begin{cases} \text{LATE}(x_i) & \text{if } i \text{ is a complier} \\ 0 & \text{otherwise,} \end{cases} \tag{2.29}$$

where LATE(x) denotes the local average treatment effect for an individual with covariate x. As before, we normalize $r(x, 0)$ to 0 as we only consider expected welfare. Note that always takers and never-takers are associated with 0 rewards. The evolution of the budget is also different for each group. In particular,

$$N(z' - z) = \begin{cases} G_a(x, t, z) & \text{if } i \text{ is a complier} \\ G_1(x, t, z) & \text{if } i \text{ is an always-taker} \\ G_0(x, t, z) & \text{if } i \text{ is a never-taker.} \end{cases} \tag{2.30}$$

The planner does not know the true compliance behavior of any individual, but she can form expectations over them given the observed covariates. Let $q_c(x), q_a(x)$ and $q_n(x)$ denote the probabilities that an individual is respectively a complier, always-taker or never-taker conditional on x. Given these quantities, the analysis under non-compliance proceeds analogously to Section 2.3, after taking relevant expectations over the rewards in (2.29), and over the evolution of z in (2.30). In particular, let $h_\theta(z, t)$ denote the integrated value function in the current setting. Then we have the following PDE for the evolution of $h_\theta(z, t)$:

$$\lambda(t)\bar{G}_\theta(z, t)\partial_z h_\theta(z, t) + \partial_t h_\theta(z, t) - \beta h_\theta(z, t) + \lambda(t)\bar{r}_\theta(z, t) = 0, \quad h_\theta(0, t) = 0 \; \forall \; t,$$

where

$$\bar{r}_\theta(z, t) := E_{x\sim F}\left[q_c(x)\pi_\theta(1|x, z, t)r(x, 1)\right],$$

and (in view of equation 2.30),

$$\begin{aligned} \bar{G}_\theta(z, t) := E_{x\sim F}\big[&q_c(x)\left\{\pi_\theta(1|z, t)G_1(x, t, z) + \pi_\theta(0|z, t)G_0(x, t, z)\right\} \\ &+ q_a(x)G_1(x, t, z) + q_n(x)G_0(x, t, z)\big]. \end{aligned}$$

In order to estimate the optimal policy rule, we need estimates of $q_c(x), q_a(x), q_n(x)$, along with LATE(x). To obtain these, we assume that the planner has access to an observational study involving Z as the intended treatment status or instrumental variable, and W as the observed treatment. As before, Y is the observed outcome variable. Observe that $q_a(x) = E[W|X = x, Z = 0]$ and $q_n(x) = E[1 - W|X = x, Z = 1]$. Hence we can estimate $\hat{q}_a(x)$ by running a Logit regression of W on X for the sub-group of the data with $Z = 0$. Estimation of $\hat{q}_n(x)$ can be done in an analogous manner. Using both these estimates, we can also obtain $\hat{q}_c(x) = 1 - \hat{q}_a(x) - \hat{q}_n(x)$. To estimate LATE($x$), we recommend the doubly robust version of Belloni et al. (2017). In the case where there do no exist any always-takers, the expression for this simplifies and is given by

$$\hat{\text{LATE}}(x) = \theta_y(1) - \theta_y(0)$$

where

$$\theta_y(1) := \frac{\widetilde{\mu}(x, W, 1) + \frac{Z}{\hat{p}(x)}(WY - \widetilde{\mu}(x, W, 1))}{\hat{q}_c(x) + \frac{Z}{\hat{p}(x)}(W - \hat{q}_c(x))}, \qquad \text{and}$$

$$\theta_y(0) := \frac{\widetilde{\mu}(x, 1-W, 1) + \frac{Z}{\hat{p}(x)}\left[(1-W)Y - \widetilde{\mu}(x, 1-W, 1)\right] - \left[\hat{\mu}(x, 0) + \frac{1-Z}{1-\hat{p}(x)}(Y - \hat{\mu}(x, 0))\right]}{\frac{Z}{\hat{p}(x)}\left(\hat{q}_c(x) - W\right) - \hat{q}_c(x)}.$$

In these equations, $\widetilde{\mu}(x, k, Z)$ is an estimator for $E[kY|Z, x]$, which can be obtained through series regression or other non-parametric methods. Furthermore, $\hat{p}(x)$ is an estimator for $p(x) = P(Z = 1|X = x)$ - the IV propensity score.

Given the estimates $\hat{q}_c(x), \hat{q}_a(x), \hat{q}_n(x)$ and $\hat{\text{LATE}}(x)$, it is straightforward to modify the algorithm in Section 2.4 to allow for non-compliance. The main difference from Algorithm 2 is that at each update we would randomly draw the compliance nature of the individual from a multinomial distribution with probabilities $(\hat{q}_c(x), \hat{q}_a(x), \hat{q}_n(x))$. Conditional on this draw, the rewards are given by sample counterpart of (2.29), and the updates to budget by (2.30). The pseudo-code for the resulting algorithm is provided in Appendix B.2.

Probabilistic bounds on the regret for the estimated policy rule can also be obtained by the same techniques as in Section 2.5. If $q_c(x), q_a(x), q_n(x)$ were known exactly, we can show that the rates for the regret remain unchanged at $\sqrt{v/n}$. The key step is to obtain concentration bounds analogous to (2.24), following which we can proceed with the discussion in Section 2.5. A similar analysis when using the estimated quantities $\hat{q}_c(x), \hat{q}_a(x), \hat{q}_n(x)$ is however more involved; we leave the details for future research.

2.6.2 Arrival Rates Varying by Covariates

In many dynamic settings, like our example with JTPA in the next section, different individuals not only respond differently to treatment, but also have (potentially) different dynamics regarding their arrival rates. This is equivalent to saying that we would like to let the distribution F_t of the covariates change with time (and in general be different from the limit of the empirical distribution F_n). Precisely, let $\lambda_x(t)$ denote the covariate specific arrival process. Then we can decompose F_t as a time-varying compound distribution[87]

$$F_t(y) = \int_{x \le y} w_t(x) dF(x) \quad \text{where } w_t(x) := \frac{\lambda_x(t)}{\int \lambda_\omega(t) dF(\omega)}.$$

With the above in mind, the PDE for the evolution of $h_\theta(z, t)$ is the same as (2.12), but

[87]The functions $F(.)$ and $\lambda_{(.)}(t)$ are separately identifiable since we associate F with the limit of F_n.

now

$$\bar{r}_\theta(z,t) := E_{x\sim F_t}\left[\pi_\theta(1|x,z,t)r(x,1)\right],$$
$$\bar{G}_\theta(z,t) := E_{x\sim F_t}\left[G_1(x,z,t)\pi_\theta(1|x,z,t) + G_0(x,z,t)\pi_\theta(0|x,z,t)\right]$$

and $\lambda(t)$ is replaced by $\bar{\lambda}(t)$ where

$$\bar{\lambda}(t) := E_{x\sim F_t}\left[\lambda_x(t)\right].$$

Suppose that $w_t(x)$ were known piece-wise constant over some clusters $j = 1,\dots,J$ on the covariate space $\mathcal{X}$. Then we could write $F_t = \sum_j w_t(j)F_j$, where F_j denotes the distribution of the covariates corresponding to cluster j. If in addition these cluster indices were known, we can estimate the cluster specific arrival rates $\hat{\lambda}_j(t)$, and weights $\hat{w}_t(j) := \hat{q}_j\hat{\lambda}_j(t)/\sum_i \hat{q}_i\hat{\lambda}_i(t)$. Here, $\hat{q}_j$ denotes the empirical proportion of observations within cluster j. Using these, we can replace F_t with the empirical counterpart $F_{n,t} = \sum_j \hat{w}_t(j)F_{n,j}$ and construct the empirical PDE (2.14) using the sample quantities

$$\hat{r}_\theta(z,t) = E_{x\sim F_{n,t}}\left[\pi_\theta(1|x,z,t)\hat{r}(x,1)\right],$$
$$\hat{G}_\theta(z,t) = E_{x\sim F_{n,t}}\left[G_1(x,z,t)\pi_\theta(1|x,z,t) + G_0(x,z,t)\pi_\theta(0|x,z,t)\right], \text{ and}$$
$$\hat{\lambda}(t) = \sum_j \hat{w}_t(j)\hat{\lambda}_j(t).$$

The above discussion is only suggestive since we do not expect $w_t(x)$ to really be piece-wise constant. However, as long as $w_t(x)$ is Lipschitz continuous, we can approximate it with a piece-wise continuous function $\hat{w}_t(x)$ by partitioning the space $\mathcal{X}$ of the covariates into a finite set of clusters $j = 1,\dots,J$. To do this, we can employ iterative partitioning using the median or mean (k-median/means clustering). After randomly choosing one observation per cluster as starting point, each observation is assigned to the cluster which is closest in terms of the median/mean, and then the cluster median/mean is recomputed (see Anderberg, 1973). The value of J is allowed to increase with the sample size.

In terms of the theoretical bounds on the regret rates, we will now have an additional term due to the approximation error from replacing $w_t(x)$ with the cluster estimate $\hat{w}_t(x)$. Let us denote this rate by $\mathcal{R}(n,J)$, i.e.

$$\sup_{x\in\mathcal{X}}\left|F_t(x) - \int_{y\le x}\hat{w}_t(y)dF(y)\right| \lesssim \mathcal{R}(n,J).$$

One can derive the rates $\mathcal{R}(n,J)$ under various conditions on the smoothness of $w_t(x)$, the number of data-points used to estimate $\hat{w}_t(\cdot)$ (allowing that it could be obtained from a different dataset), and the choice of J. To illustrate, suppose that $w_t(x)$ is continuously differentiable and we use the same dataset as the RCT to estimate $\hat{w}_t(x)$, then under some regularity conditions one can show $\mathcal{R}(n,J) \lesssim J^{-d_x} + \sqrt{J/n}$, where d_x denotes the dimension of x (Bonhomme et al., 2017). Different rates are possible under other assumptions; we shall not document these here but simply use $\mathcal{R}(n,J)$ to state our results. In particular, the concentration inequalities (2.24) will now include an additional $\mathcal{R}(n,J)$ term:[88]

$$E\left[\sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta}\|\hat{r}_\theta(z,t)-\bar{r}_\theta(z,t)\|\right]\le C_0\sqrt{\frac{v_1}{n}}+\mathcal{R}(n,J),$$

with related expressions for $\hat{G}_\theta(z,t)-\bar{G}_\theta(z,t)$ and $\hat{\lambda}(t)-\bar{\lambda}(t)$. Subsequently, proceeding with the remainder of the analysis in Section (2.5) enables us to show that

$$h_{\theta^*}(z_0,t_0)-h_{\hat{\theta}}(z_0,t_0)\le 2C\left(\sqrt{\frac{v}{n}}+\mathcal{R}(n,J)\right),$$

with probability approaching one.

It is straightforward to extend our Actor-Critic algorithm to allow for clusters: before each update we sample the cluster index by drawing the value of j from a multinomial distribution with probabilities $(\hat{w}_t(1),\dots,\hat{w}_t(J))$. The pseudo-code for the resulting algorithm is provided in Appendix B.2.

2.6.3 Deterministic Policy Rules

So far our discussion has focused on policy rules that are in general stochastic. This gives more flexibility to the social planner, but there are situations in which randomization is not appealing for legal or ethical reasons. Here we investigate how one might adapt our proposed algorithm when restricted to deterministic policy rules. Our theoretical results require no modification as they already encompass deterministic policies.

One approach for handling deterministic policies is to approximate them by a randomized policy that is arbitrarily close. For instance, suppose that the social planner is restricted to using linear eligibility scores (Kitagawa and Tetenov, 2018) as policy rules, for example:

$$\Pi=\left\{\pi_\theta:\pi_\theta(1|s)=\mathbb{I}(s'\theta>0)\right\}.$$

One issue with such a functional class is that it is not differentiable. However we can employ analytical approximations to the step function to make these functions arbitrarily smooth. For example, instead of Π we could employ the class

$$\tilde{\Pi}_k=\left\{\tilde{\pi}_\theta:\tilde{\pi}_\theta(1|s)=\frac{1}{2}+\frac{1}{\pi}\arctan(ks'\theta)\right\}$$

where $k\in\mathbb{R}^+$ is arbitrarily large. Then as $k\to\infty$, $\tilde{\Pi}_k\to\Pi$.

[88] In deriving this expression, we make use of the fact that the concentration bounds in (2.24) hold uniformly over all probability distributions (and therefore hold uniformly over all $\tilde{F}_t(x)=\int\hat{w}_t(x)dF(x)$). In particular, we can decompose $\hat{r}_\theta(z,t)-\bar{r}_\theta(z,t)$ as the difference between $\hat{r}_\theta(z,t)-\tilde{r}_\theta(z,t)$ and $\tilde{r}_\theta(z,t)-\bar{r}_\theta(z,t)$, where $\tilde{r}_\theta(z,t):=E_{x\sim\tilde{F}_t}[\pi_\theta(1|x,z,t)r(x,1)]$ and $\tilde{F}_t:=\int\hat{w}_t(y)dF(y)$. The first term is the of order $\sqrt{v/n}$ due to the uniform concentration bounds, while the second term $\mathcal{R}(n,J)$ then arises from the difference between $F_t-\tilde{F}_t$ as discussed above.

The main difficulty with applying our algorithm on $\tilde{\Pi}_k$ with a large k is that it does not permit sufficient exploration. Hence when faced with policy rules that are close to deterministic, we recommend using an off-policy actor-critic algorithm (Degris, White and Sutton, 2012). In an off-policy setting, the algorithm chooses actions according to a behavioral policy (e.g. $b(1|s) = 1/2$), but uses the resulting outcomes to update the target policy π_θ. To account for the fact that the sequence of states under the target policy is different from that under the behavioral policy, the updates are adjusted by the importance weights $\rho(a|s) = \pi_\theta(a|s)/b(a|s)$. With these modifications, it is straightforward to extend our actor-critic algorithm to the off-policy context. The pseudo-code for this provided in Appendix B.2. The theoretical properties of the algorithm can be derived in an analogous way to Section 2.4.2; we do not present them here as they follow in a straightforward manner from the results in Degris, White and Sutton (2012).

This approach requires a choice of k. Instead of keeping this constant, we can start from a moderate initial value for k and increase it slowly in the course of the updates. In particular, we recommend the following two-step procedure: First we solve for the optimal policy function at some initial value of k using the on-policy algorithm from Section 2.4. We then use this as the behavioral policy, and use our off-policy actor-critic algorithm to update the policy function as the values of k are increased.

2.7 Empirical Application: JTPA

To illustrate our approach, we use the popular dataset on randomized training provided under the JTPA, akin to e.g. Kitagawa and Tetenov (2018), or Abadie, Angrist, and Imbens (2002). During 18 months, applicants who contacted job centers after becoming unemployed were randomized to either obtain support or not. Local centers could choose to supply one of the following forms of support: training, job-search assistance, or other support. Again akin to Kitagawa and Tetenov (2018), we consolidate all forms of support. Baseline information about the 20601 applicants is available as well as their subsequent earnings for 30 months. We follow the sample selection procedure of Kitagawa and Tetenov (2018) and delete entries with missing earnings or education variables as well as those that are not in the analysis of the adult sample of Abadie, Angrist, and Imbens (2002). This results in 9223 observations.

In this setting, a policy maker is faced with a sequence of individuals who just became unemployed. For each arriving individual, she has to decide whether to offer job training to them or not. The decision is made based on current time, remaining budget, and individual characteristics. For the latter, we follow Kitagawa and Tetenov (2018) and use education, previous earnings, and age. Job training is free to the individual, however, costly to the policy maker who has only limited funds.

The frequency with which people with given characteristics apply is not constant throughout the year. As we use RCT data which contains information regarding when participants arrived, we can estimate Poisson processes that are changing over the course of the year. We first partition the data into clusters using k-median clustering on education, previous earnings, and age.[89] Prior to the clustering, we standardize the variables. The resulting clusters are described in Appendix B.4.

For each cluster, we estimate the arrival probabilities. While we assume that they are constant across years, we allow for variation within a year. In particular, we specify the following functional form for the cluster-specific Poisson parameter:

$$\lambda_c(t) = exp\left(\beta_{0,c} + \beta_{1,c} sin(2\pi t) + \beta_{2,c} cos(2\pi t)\right).$$

Regarding time, t is normalized so that $t = 1$ corresponds to a year. For each cluster, we obtain the estimates $\hat{\beta}_c$ (and hence $\hat{\lambda}_c(t)$) using maximum likelihood estimation. Figure 2.1 shows the estimated dynamic behavior of each cluster. People from cluster 1, for example, display a less pronounced seasonal pattern regarding their arrival rates than people from cluster 2.

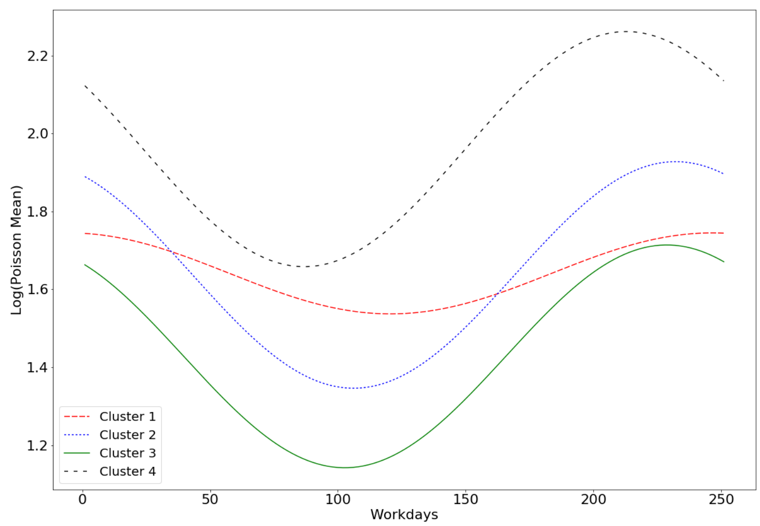


Figure 2.1: Clusters-Specific Arrival Rates over Time

We obtain the reward estimates $\hat{r}(x, 1)$ in two ways: (i) $\hat{r}(x, 1) = \hat{\mu}(x, 1) - \hat{\mu}(x, 0)$, and (ii) from a doubly robust procedure as in (2.23). In both cases we use simple OLS to estimate the conditional means. For this reason we shall call case (i) the case of standard OLS rewards. The relevant covariates are education, previous earnings, and age. Estimating the propensity score is not necessary in this context as it was set by the RCT to be $\frac{2}{3}$. Note that the different reward estimates give rise to different heterogeneity patterns, which crucially affect the resulting policy function. Indeed, while the doubly robust procedure consistently estimates the true heterogeneity structure, the standard OLS does not. Consequently, we expect differing parameters in the policy functions and treatment decisions.

In terms of the other parameters, we set the budget such that 1600 people can be treated,

[89] Given the limited amount of data, the number of clusters we can reliably estimate is limited too. We chose to use four clusters. With more data, more clusters and hence a more detailed picture of differential arrival of applications becomes possible.

which is about a quarter of the expected number of people arriving in a year (given our Poisson rates). Subsequently, we normalize z in such way that $z_0 = 0.25$. We also use a discount factor of $\beta = -\log(0.9)$, which implies an annualized discount rate of 0.9 (since $t = 1$ corresponds to a year). The episode terminates when all budget is used up.

The primary outcome variable is the policy function. We chose the policy function class to be of the logistic form $\pi_\theta \sim \text{Logit}(\theta_1^\intercal \mathbf{x} + \theta_2^\intercal \mathbf{x} \cdot z + \theta_3^\intercal \mathbf{x} \cdot \cos(2\pi t))$, where $\mathbf{x} = (1, \text{age}, \text{education}, \text{previous earnings})$. We use $\cos(2\pi t)$ to ensure that the arrival rates are periodic, and to prevent discontinuities at the end of the year. Note that this allows for episodes potentially lasting longer than a year, but constrains the years themselves to be identical.

To run our Actor-Critic algorithm we need to set the learning rates. There exist some rules of thumb for these, see e.g. Sutton and Barto (2018). In practice, however, we tune the rates manually starting from the rules of thumb to optimize the performance of the algorithm. Choosing a rate that is too high makes the algorithm unstable, while setting the rate too low makes convergence slow. Based on pilot runs we found that by setting $\alpha_\theta = 0.3$ and $\alpha_\nu = 0.8$ we could achieve good performance.

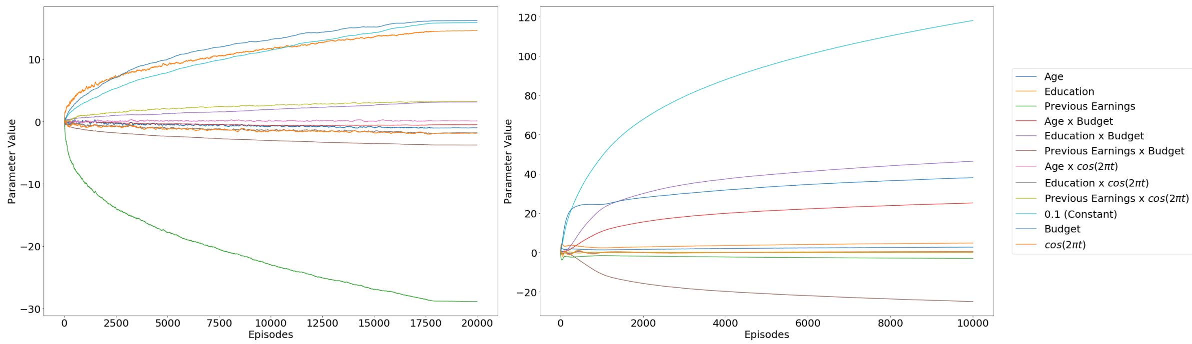
With these rates, employing the parallel actor-critic algorithm with clusters (see Appendix B.2 and section 2.6.2) provides promising results. Figure 2.2 shows that the expected welfare converges as learning occurs through the episodes. In both cases we normalize welfare so that choosing a random policy provides a welfare of 1. The welfare is thus approximately three times higher than that under random treatment in the initial episode. The parameters in the policy function generally converge as well, as shown in Figure 2.3.

A: Doubly Robust Reward Estimates

B: Standard OLS Reward Estimates

Figure 2.2: Converging Episodic Welfare

In sum, we have shown that rewards are substantially higher than under random treat-



A: Doubly Robust Reward Estimates B: Standard OLS Reward Estimates

Figure 2.3: Convergence of Policy Function Parameters

ment. Moreover, both rewards and policy function parameters converge. This illustrates the functionality of our algorithm for given reward estimates.

2.8 Conclusion

In this paper we have shown how to estimate optimal dynamic treatment assignment rules using RCT data under constraints on the policy space. We have proposed an Actor-Critic algorithm to efficiently solve for these rules. Our framework can be naturally extended to incorporate non-compliance using instrumental variables. Separately, our results also point the way to using Reinforcement Learning to solve PDEs characterizing the evolution of value functions. We do so by approximating the PDEs with a dynamic program. We were also able to characterize the numerical error involved in this approximation.

Perhaps the main limitation of the current setting is that the policy is assumed not to affect the environment. We believe this is a reasonable assumption in many contexts, especially in settings like unemployment, arrivals to emergency rooms, childbirth (e.g. for the provision of daycare) etc., where either the time of arrival is not in complete control of the individual, or where it is determined by factors exogenous to the provision of treatment. Moreover, in cases like microcredit or development aid, the budgetary cycle might be unknown to the individuals (and potentially not constant) - prohibiting strategic arrival. At the same time, the assumption is clearly circumspect in other cases. For instance, a treatment rule that is more favorable at specific times may also encourage people to arrive at different times. If the response function of the environment to the policy is known, it could be integrated in the approach outlined in this paper. We leave that to future research.

Chapter 3

Not-For-Profit Firms as Means of Market Entry

3.1 Introduction

In recent years, large companies have created Not-For-Profit (NFP) firms with the stated objective to provide affordable quality goods to the poor.[90] These companies belong to varying industries and only some of them are active in the country where they have created the NFP firm (see Grameen, 2014a). This behavior appears puzzling at first, as creating a NFP firm for the poor seems not to be profit maximizing; partly due to the NFP status itself, and partly as For-Profit (FP) firms regularly ignore markets in which consumers are poor (Yunus, 2008).

I provide a novel explanation for this behavior, treating NFP firms as a signaling device. The NFP firms are created in "markets for the poor". These markets are by themselves not attractive for companies because the profits that can be obtained there do not match the costs of signaling high quality and developing a new product. However, NFP firms have no incentives to lie about the produced quality - by definition they cannot keep any profits from doing so (see Glaeser & Shleifer, 2001). Therefore, when they enter a new market, it is rational to *instantly* believe their quality-claim - hence NFP firms face no signaling costs. After producing sufficiently many high-quality goods with a NFP firm, a company earns a global reputation for high quality. This reputation allows the company to profitably create an additional firm in another (related) market for high-quality products, but with a FP status.

Consequently, the concern for reputation combined with companies being active in several markets not only affects which markets are served, but also the optimal organizational form. Contrary to large parts of the related literature,[91] no altruistic motivation of any agent is required. Nevertheless, profit-oriented companies create NFP firms in "markets for the poor", despite the fact that this is costly (the expected cost of developing a new product). Altruistic motivation is complementary though, and I use motivated workers in an extension of the main mechanism to achieve patterns of the creation of NFP firms that potentially better explain the stylized facts. These motivated workers are part of the NFP firms though and the FP companies creating the NFP firms are still free from any altruistic agents.

The model I present here is dynamic. As the markets grow over time (for example due to population growth), establishing a NFP firm in a "market for the poor" becomes optimal. The profits from operating in the "market for the rich" with a high-quality reputation grow,

[90] Examples include Danone, Uniqlo, Schneider Electric, Veolia, or Felissimo (Grameen, 2014a).

[91] This literature includes Glaeser & Shleifer, 2001; Rose-Ackerman, 1996 & 1997; Ghatak & Müller, 2011; and Besley & Ghatak, 2007. The way altruism drives these models varies. For Rose-Ackerman (1996 & 1997), altruism of firm owners and donations are key. In Besley & Ghatak (2007), altruistic consumers allow for an equilibrium in which prices are high enough to pay for corporate social social responsibility - which can arguably include the creation of NFP firms. In Ghatak & Müller (2011), the abundant supply of labor from altruistic workers is key.

but the expected costs of creating a NFP firm are constant.

Moreover, the model is a hybrid between a moral hazard and an adverse selection model. It distinguishes two types of firms. While one type always produces high quality, producing low quality is cheaper for the other type. Crucially though, the latter type is able to produce high quality (at a high cost) and mimic the former type. This aspect is related to the model by Laffont and Tirole (1986), where firms have different efficiency parameters but also choose a level of effort. The model presented in this paper is simpler in the sense that only the latter type makes a choice, but it extends the idea to an environment where the output's quality is unobserved at the time of purchase (and to more than one market).

The concept that firms need to establish a costly reputation when quality is only observed after the purchase builds on the work by Shapiro (1983). The current paper is also related to the literature regarding the advantages of production with a NFP status, including Francois (2000, 2003), Hansmann (1980), Ghatak and Müller (2011), as well as Glaeser and Shleifer (2001). The mechanism that FP firms, but not NFP firms, require a costly signal to establish a high-quality reputation is also present in parts of this literature. I employ this mechanism in a dynamic environment with more than one market (i.e. more than one variant of the good), extending the work by Hansmann (1980) and Glaeser and Shleifer (2001). Finally, this paper contributes to the literature concerning mechanisms behind the creation of NFP firms (see footnote 91).

Case Studies

Evidence for the mechanism suggested in this paper can be found in case studies of NFP firms that are created by large companies - usually in cooperation with the Grameen Creative Lab, which fosters social businesses in general.

Danone: The French multinational food company Danone founded the NFP firm Grameen Danone in 2006 in Bangladesh. It has designed a high quality yoghurt for children (it includes additional nutrients for the children's optimal development), which is sold at 6 Eurocents (Grameen, 2014a). While the product itself is new and specifically designed for the poor, fresh dairy products are the key component of Danone's sales (58% in 2011; Danone, 2012). More specifically, more expensive yoghurts that are nutritionally valuable for children are among Danone's major brands.[92]

Apart from the initial investment, no profits are allowed to flow back to Danone. Before 2006, Danone itself was not operating in Bangladesh (Yunus, 2007).

BBC business reporter James Melik (2009) believes the underlying reason for this NFP firm is Danone's goal to expand in South Asia. His German colleague at the business-newspaper Handelsblatt had similar suspicions when Adidas evaluated the option to create a NFP firm which would sell a pair of shoes for 1 Euro (Hauschild, 2014). Melik's (2009) suspicion is supported by the fact that the share of the Asia/Pacific market of Danone's sales tripled from 5% in 1996 to 15% in 2011, documenting Danone's ambition in the greater region (Danone, 2012).

Uniqlo: The Japanese multinational clothing company Uniqlo has founded Uniqlo Social

[92]E.g. Fruchtzwerge in Germany, Danimals in the US, Danino in Canada, or Danoninho in Brazil.

Business Bangladesh, and Grameen Uniqlo in 2010 and 2011 respectively (Grameen Uniqlo, 2014b). These are also NFP firms that fully reinvest all potential profits. Consumers of Grameen Uniqlo are not as poor as those from Grameen Danone – but the products are still sold cheaply: prices start at 2.5 US$ (Sadique, 2013). Their primary business policy is "(...) providing affordable (...) exceptional apparel." (Grameen Uniqlo, 2014a). Uniqlo itself is not present with shops in Bangladesh (Uniqlo, 2014). However, Yukihiro Nitta, the CEO of Grameen Uniqlo, talks openly about the potential to become market leader in the fast growing economy of Bangladesh, which has been interpreted as a sign that Uniqlo targets the increasingly wealthy urban population there (see Sadique, 2013).

Schneider Electric: The globally operating French electricity corporation Schneider Electric created a NFP firm called Grameen Schneider Electric. The stated goal of the NFP firm is to provide electricity and associated services to the poor (Schneider Electric, 2012). The difference to the previous examples is that Schneider Electric itself has already been present in Bangladesh prior to establishing the NFP firm (Schneider Electric, 2014).

Further companies that have created NFP firms include Veolia and Felissimo (Grameen, 2014b). In sum, the model presented in this paper is motivated by the following stylized facts:

1. For-profit companies create NFP firms that provide cheap products specifically aimed at the poor - even though the focus on the poor is not the core-business of these companies.

2. In two case-studies, these companies had not been present in the country at all (including the market of their core-business). Journalists suspect that the motivation for the NFP-creation is to allow for an entry in the "core-market" of the respective country.

3. In at least one case, the for-profit company creating the NFP firm was present in the core-market of the country in which the NFP firm was created.

4. The notion that for-profit companies create NFP firms is, to the best of my knowledge, a relatively recent phenomenon.

The remainder of this paper is organized as follows. Section 2 illustrates the basic mechanism and showcases how reputation is obtained and what equilibrium is reached if no information travels across markets. The full model, including the interaction between markets and the idea of global reputation, is outlined in section 3. Section 4 provides the extension to include motivated workers in the model. Finally, section 5 discusses welfare and policy implications of the mechanisms presented in this paper and section 6 concludes.

3.2 Basic Framework

To illustrate the basic mechanism, this section outlines the basic environment of two types of firms and one variant of product that can be produced in two qualities. Firms face the situation that consumers cannot observe quality at the time of purchase. This can be overcome by establishing a reputation for high quality, which is costly (see Shapiro, 1983). If given the incentive to do so, "discounter" firms can produce high quality to appear as a "quality expert" firm and then produce low quality to "milk" this reputation. The full mechanism, including the key aspect of having two variants of products, is outlined in section 3.3.

3.2.1 Quality & Demand

The quality of the product $a \epsilon \{h, l\}$ is either high or low. The product has a credence-good character: its quality is only observed one period after purchase, which defines the length of a period. The quantity of the high-quality product that is demanded at time t is $Q^D_{ht}(p_{ht}) = D_t(k - p_{ht})$. The demanded quantity of the low-quality product is $Q^D_{lt}(p_{lt}) = \hat{D}_t(z - p_{lt})$. The terms D_t and $\hat{D}_t$ are growing over time (e.g. due to population growth). Specifically, they are both assumed to grow at the constant rate g. Consequently, the prices p_{at} potentially vary over time as well.

3.2.2 Firms & Production

The product can be produced by two types of firms. The cost for a firm of type f to produce quantity q of quality a is $C_f(q_a) = c_{fa}q_a$; where $f \epsilon \{d, e\}$: firms are either "discounters" or "quality experts". Discounters have an advantage in producing low quality cheaply: $c_{dl} < c_{el}$. Quality experts are able to produce high-quality cheaper: $c_{eh} < c_{dh}$. I abstract from the case where quality experts are tempted to produce low quality. For that reason, I make the simplifying assumption that for quality experts, producing high and low quality is equally expensive: $c_{el} = c_{eh}$.

This setup contains elements of both adverse selection an moral hazard. There are two different types of firms as in the standard adverse selection case. Moreover, discounters can produce high quality, but are potentially tempted to produce low quality since this is cheaper, as in a moral hazard setting.

Finally, firms can choose to operate with a not-for-profit (NFP) status. This status is permanent and, contrary to quality-claims, always credible (other than quality, it can be enforced through the legal system). By definition, NFP firms cannot make any profits, but this also implies that they have no incentives to lie about the produced quality (Glaeser & Shleifer, 2001). In the single-variant case presented in this section, choosing a NFP status will never be optimal due to the inability to make any profits. It is nevertheless introduced here because it is a fundamental part for the case with two variants outlined in section 3.3.

3.2.3 Equilibrium

Regarding the low-quality product, no firm has an incentive to lie about quality. No form of signaling or reputational investment is required and the market is perfectly competitive. Consequently, the equilibrium price $p^*_{lt} = c_{dl}$, which is constant. Regarding the demand parameter z, I assume $z > c_{dl}$, so that there is positive demand for the low-quality product at any point in time.

The equilibrium for the high-quality products is more subtle. Since quality is unobservable at the time of purchase, consumers have a belief regarding each firm whether it produces high or low quality (reputation of that firm). There is no equilibrium where the reputation to produce high quality is free to obtain for FP firms. Discounters would claim to sell high quality but profitably sell low quality and hence believing this claim would be irrational.[93]

I assume that all N_{ht} firms that have the reputation to produce high quality at time t are in Cournot competition. In terms of possible costly signals to obtain this reputation, I restrict my focus to the production of the goods at any quantity or price, with or without a NFP status.[94] The cost of signaling for a firm of type f at time t is denoted s_{ft}. In equilibrium, the entry of an additional quality expert is not profitable:

$$\pi_{et}^{cournot}\left(p_{ht}, N_{ht}, r\right) - s_{et} \leq 0, \tag{3.1}$$

where $\pi_t^{cournot}(.)$ represents the discounted sum of profits at the interest rate r. The standard Cournot equilibrium with N_{ht} quality experts is given by each firm supplying $q_{ht}^* = \frac{D_t}{N_{ht}+1}\left(k - c_{eh}\right)$. Each of these firms makes a discounted stream of profits of:

$$\pi_{et}^{cournot} = \frac{1}{1 - \frac{1+\hat{g}}{1+r}} \frac{D_t}{\left(N_{ht}+1\right)^2}\left(k - c_{eh}\right)^2.$$

I make the technical assumption that $r > g$. The rate $\hat{g} \leq g$ denotes the growth rate of the per-period profits of an individual firm, $\pi_t^{cournot,period} = \frac{D_t}{(N_{ht}+1)^2}\left(k - c_{eh}\right)^2$.[95] If N_{ht} is constant, then $\hat{g} = g$.

The costs of signaling to each firm-type, s_{ft}, are such that discounters do not mimic quality experts. Specifically, the cost of the considered signal is to produce $\frac{h_t}{1+r}$ high-quality goods in the period prior to market entry and sell them with a loss at price p_{lt}:[96]

$$s_{ft} = h_t\left(c_{fh} - c_{dl}\right). \tag{3.2}$$

Due to condition (3.1) and $s_{et} < s_{dt}$,[97] no discounter enters to permanently produce high

[93] The same is true for any probabilistic belief that attributes a non-zero probability for high quality to a FP firm that does not invest in signaling.

[94] I can also allow for lump sum signals, but these are always dominated by production-based signals: since $c_{eh} < c_{dh}$, producing the high-quality good results in a lower signaling cost for quality experts relative to discounters, which is more efficient.

[95] The rate $\hat{g}$ is not indexed by t because it is assumed to be time-independent. This means I assume a constant rate of change of $\pi_{et}^{cournot,period}$. See Appendix C.2 for a discussion.

[96] I assume that the demand in the low-quality market is large, so that selling at least $\frac{h_t}{1+r}$ units in one period is always possible and that firms can restrict the quantity in order not to sell more, i.e. $\frac{Q_{lt}^D(c_{dl})}{N_{lt}} > \frac{h_t}{1+r}$, or if the price is reduced to $p_{lt} - \varepsilon$, $Q_{lt}^D\left(c_{dl}\right) > \frac{h_t}{1+r}$. The division by $(1+r)$ is for notational convenience as s_{ft} refers to entry at time t by using this signal, but the action of signaling starts in the period prior to market entry. This assumption (and notation) is for convenience only. If the assumption fails, products are sold over multiple periods and discounted accordingly. In such a case, all that changes is the definition of h_t. For example, if two periods are necessary, $\frac{h_{1t}}{1+r}$ units are sold in the period before obtaining the high-quality reputation, and $\frac{h_{2t}}{(1+r)^2}$ units in the period before that, $h_t = h_{1t} + h_{2t}$.

[97] In general, $s_{et} \leq s_{dt}$. If the signal was equally costly to both firms (e.g. "burning" money), then $s_{et} = s_{dt}$,

quality. However, discounters are relatively more efficient at producing low quality. Therefore, they potentially choose to obtain a high-quality reputation to "milk" this reputation by producing low-quality products labeled as high quality. This is feasible for one period. Thereafter, consumers observe the quality and discover the fraud. Another equilibrium condition is therefore:

$$\pi_{dt}^{milk}\left(p_{ht}, N_{ht}\right) - s_{dt} \leq 0. \tag{3.3}$$

The "milking profit" is given by:

$$\pi_{dt}^{milk}\left(p_{ht}, N_{ht}\right) = \frac{Q_{ht}^{D}\left(p_{ht}\right)}{N_{ht}}\left(p_{ht} - c_{dl}\right) = \left(k - c_{dl} - \frac{N_{ht}}{N_{ht}+1}\left(k - c_{eh}\right)\right)\frac{D_t}{N_{ht}+1}\left(k - c_{eh}\right).$$

Finally, once a quality expert has paid the sunk cost to established a high-quality reputation, there is no reason for that firm to exit the market, even if profits become lower over time. This results in condition (3.4):

$$N_{ht} \leq N_{h(t+1)}. \tag{3.4}$$

In equilibrium, if condition (3.4) is slack, condition (3.1) must hold with equality as otherwise more firms would enter. Similarly, if condition (3.4) holds with equality, condition (3.1) is potentially slack. Moreover, I impose that condition (3.3) holds with equality; i.e. I select the most profitable equilibrium for the firms. This selection is supported by backward induction. Using (3.1)-(3.3) (all binding), the conditions determining N_{ht}^{*} and s_{dt}^{*} are given by:

$$\frac{1}{1-\frac{1+\hat{g}}{1+r}}\frac{D_t}{\left(N_{ht}+1\right)^2}\left(k - c_{eh}\right)^2 = s_{dt}\frac{c_{eh}-c_{dl}}{c_{dh}-c_{dl}}, \tag{3.5}$$

$$\left(k - c_{dl} - \frac{N_{ht}}{N_{ht}+1}\left(k - c_{eh}\right)\right)\frac{D_t}{N_{ht}+1}\left(k - c_{eh}\right) = s_{dt}. \tag{3.6}$$

Lemma 3

The absolute cost of signaling is increasing with the market size (for both type of firms), at a non-decreasing rate.

Proof: Plugging the expression of s_{dt} provided in condition (3.6) into condition (3.5), it becomes clear that N_{ht}^{*} is constant in time, condition (3.4) is satisfied, and $\hat{g} = g$. Due to this, and assuming $k > c_{eh}$ (as otherwise there would be no demand for the high-quality product), s_{dt}^{*} is linearly increasing in D_t, and hence increasing at a constant rate over time, and due to (3.2), so does s_{et}^{*}.□

Intuitively, as the demand is increasing over time, so is the milking profit for discounters. For condition (3.3) to hold, s_{dt} needs to increase over time as well. This implies that a higher

but in case of production with a loss and $c_{eh} < c_{dh}$, the inequality is strict.

quantity of high-quality products needs to be sold to obtain a reputation for high quality, which means that s_{et} is also increasing.

Lemma 4
For the high-quality product to be supplied by a for-profit firm ($N^*_{ht} \geq 1$), the demand parameter k (highest willingness to pay for the high-quality good) and the production costs of discounters (c_{dl}, c_{dh}) need to be sufficiently large relative to c_{eh}, the cost of the quality experts to produce the high-quality good.

Proof: Again from the conditions (3.5) and (3.6), N^*_{ht} is given by:

$$N^*_{ht} = \left(\frac{\frac{c_{dh}-c_{dl}}{c_{eh}-c_{dl}}}{1 - \frac{1+g}{1+r}} - 1 \right) \frac{k - c_{eh}}{c_{eh} - c_{dl}} - 1.$$

Consequently, $N^*_{ht} \geq 1$ reduces to:

$$k - c_{eh} \geq \frac{2\left(c_{eh} - c_{dl}\right)^2}{\frac{c_{dh}-c_{dl}}{1-\frac{1+g}{1+r}} - \left(c_{eh} - c_{dl}\right)}. \tag{3.7}$$

While $k > c_{eh}$ and $c_{dh} > c_{eh} > c_{dl}$, and hence both sides of the above inequality are positive, the inequality does not hold if c_{eh} is sufficiently close to k and/or c_{eh} is close to c_{dh} but substantially different from c_{dl}.[98] □

Lemma 4 implies that if the willingness (or ability) to pay of the consumers for the high-quality product is low (k is low relative to c_{eh}), then the high-quality market does not exist.

Finally, in this simplified setting, there are no incentives to create a NFP firm. Note though that a NFP firm has no incentives to lie about quality and can consequently operate even if condition 3.7 fails - as long as $k > c_{eh}$.

3.3 Full Model

In the full model, there are two variants of the good, each of which can be produced in high or low quality. Quality experts have an advantage in producing high quality regarding both variants of the good. Consequently, one variant of the good can be used to signal being a quality expert regarding both variants. Under some conditions, the optimal way to signal is to operate a NFP firm that produces one of the two variants.

[98] Intuitively, π^{milk}_{dt} is increasing in $(k - c_{eh})(k - c_{dl})$. Regarding $\pi^{cournot}_{dt}$, both production costs and "price defining costs" are c_{eh} and hence $\pi^{cournot}_{dt}$ is increasing in $(k - c_{eh})^2$. It is always the case that $(k - c_{eh})(k - c_{dl}) > (k - c_{eh})^2$ and other terms in the respective equations for the equilibrium profits can lead to $\pi^{*milk}_{dt} < \pi^{*cournot}_{et}$ (e.g. because the quality experts are in the market for more than one period). However, if $(k - c_{eh})(k - c_{dl}) - (k - c_{eh})^2$ is too large, then the high-quality market collapses as π^{milk}_{dt} is too high. If k is large, then $(k - c_{eh})(k - c_{dl})$ is similar to $(k - c_{eh})^2$.

3.3.1 Quality & Demand

The key difference to section 3.2 is that there are two variants of the good. The idea is that one of these variants is cheaper and aimed at “the poor”, while the other type is aimed at “the rich”. For the remainder of this paper, I will use the following illustrative example. The good is a shoe, the variant (v) of good for those with a sufficiently large willingness to pay (WTP) for it (“the rich”) is a work boot and the other type (for “the poor”) is a sandal; i.e. $v \epsilon \{b, s\}$. Both boots and sandals can be produced in high and low quality. The (relevant) demand functions are summarized in Table 3.1.

$Q^D_{vqt}(p_{vqt})$	High WTP for Boots	Low WTP for Boots
High-Quality Sandal	0 at $p_{bht} < k$	$d_t(j - p_{sht})$
Low-Quality Sandal	0 at $p_{blt} < z$	$\hat{d}_t(y - p_{slt})$
High-Quality Boot	$D_t(k - p_{bht})$	0 at $p_{bht} \geq c_{beh}$
Low-Quality Boot	$\hat{D}_t(z - p_{blt})$	0 at $p_{blt} \geq c_{bdl}$

Table 3.1: Demand Functions by Consumer Type

Regarding the demand parameters (j, k, y, z) in Table 3.1, I assume $z > c_{bdl}$ and $y > c_{sdl}$, so that these markets are served at perfectly competitive prices, as well as

$$k \geq \frac{2\left(c_{beh} - c_{bdl}\right)^2}{\frac{c_{bdh} - c_{bdl}}{1 - \frac{1+g}{1+r}} - \left(c_{beh} - c_{bdl}\right)} + c_{beh}, \ and \ \frac{2\left(c_{seh} - c_{sdl}\right)^2}{\frac{c_{sdh} - c_{sdl}}{1 - \frac{1+g}{1+r}} - \left(c_{seh} - c_{sdl}\right)} + c_{seh} > j > c_{seh}. \tag{3.8}$$

The former implies that the market for high-quality boots is guaranteed to be served (see Lemma 4 and section 3.3.3), while the latter means that the market for high-quality sandals would not be served in isolation.

Intuitively, “the poor” are highly price sensitive. Their willingness to pay for a high-quality sandal is barely higher than its production cost. Moreover, for sandals, producing low quality is cheap, while producing a high-quality durable sandal is rather expensive, so c_{sdl} is low relative to c_{seh}.

Finally, the demand scale parameters d_t, D_t, and $\hat{D}_t$ are assumed to grow at the constant rate g.

3.3.2 Firms & Production

The cost for a firm of type f to produce quantity q of variant v and quality a is $C_f(q_{va}) = c_{vfa} q_{va}$. Regarding both boots and sandals, discounters have an advantage in producing low quality and quality experts have an advantage in producing high quality: $c_{bdl} < c_{bel}$, $c_{sdl} < c_{sel}$, $c_{beh} < c_{bdh}$, $c_{seh} < c_{sdh}$. I again abstract from the case where quality experts are tempted to produce low quality and assume $c_{bel} = c_{beh}$, $c_{sel} = c_{seh}$.

I focus on the situation where high-quality sandals are not yet a well established product - namely they do not yet exist. Consequently, there is a risk of failure associated with attempting to produce high-quality sandals. While I assume that the cost function of production (conditional on success) remains at $C_f(q_{sh}) = c_{sfh}q_{sh}$, there is a risk that the actual output is useless (e.g. sandals that fall apart, or such that do not meet the taste of the consumers). Specifically, I assume that there is an additional expected ex-ante cost of producing high-quality sandals. This cost is constant over time, equal for both types of firms, and denoted M.

Finally, firms are allowed to operate with different organizational forms in the boots and the sandals market. Namely, a company can operate with a NFP status in the sandals market but with a FP status in the boots market.

3.3.3 Equilibrium

Regarding both low-quality boots and sandals, markets are perfectly competitive with the constant prices $p^*_{blt} = c_{bdl}$ and $p^*_{slt} = c_{sdl}$. A key implication of having two variants of the product is that reputation carries through markets. If a firm has credibly signaled to be a quality expert regarding the production of sandals, this firm is a quality expert in general - also regarding the production of boots. Due to the assumed condition (3.8), the market for high-quality sandals is not served in isolation (even if M was zero). Consequently, this market can only be served for one of the following reasons:

Establishing Reputation Establish a high-quality reputation in the sandals market to profitably enter the market for high-quality boots.

Benefiting from Reputation Firms that have already established a high-quality reputation in the boots market enter the sandals market without the need for any further signaling.

3.3.3.1 Establishing Reputation

A firm can obtain a high-quality reputation via the boots market - namely by selling high-quality boots with a loss as outlined in section 3.2. Alternatively, a firm can obtain a high-quality reputation via the sandals market. This can be achieved either with a FP or a NFP status. As a NFP firm is not allowed to make profits, $p^{NFP}_{sht} = c_{seh}$.[99] The ex-ante expected cost M cannot be recouped as ex post, when successful, this would imply making a profit.

Lemma 5
If the optimal signal is producing high quality in the sandals market, the NFP status is superior to the FP status.

[99] Discounters could theoretically charge $p^{NFP,d}_{sht} = c_{sdh}$ but that is never optimal as it immediately reveals their type.

Proof: If a quality expert produces high-quality sandals with a FP status, the expected cost of that are M plus the signaling costs in the sandals market. The expected benefits are the profits from permanently operating in both the boots and sandals market with a high-quality reputation. If it is done with a NFP status, the expected cost is M and the expected benefit is only the profits from the boots market. As a direct consequence of Lemma 4 and due to the assumed condition (3.8), the cost of signaling in the sandals market is higher than the profit from operating in that market with a high reputation.□

In equilibrium, a sandal-producing NFP firm instantly obtains a high-reputation in the sandals market, and after producing $\frac{\tilde{h}_t}{1+r}$ high-quality sandals within one period, this reputation carries to the boots market.[100] It is necessary to wait for one period, as otherwise low-type firms would have an incentive to immediately milk the high reputation in the boots market. With this one-period wait for the reputation to carry through markets, the argument of Glaeser and Shleifer (2001) comes fully into play and it is rational to believe that NFP firms produce high quality if they declare to do so (see section 3.3.4 for more details). The role of $\tilde{h}_t$ is akin to that of h_t previously, and the equilibrium conditions (replacing (3.1) - (3.4) from section 2) that allow both signaling via producing high-quality boots at a loss and via a sandal-producing NFP firm are:

$$\pi_{bt}^{cournot} \leq min\left\{\tilde{s}_{et}^{NFP}, s_{bet}\right\}, \tag{3.9}$$

$$\tilde{s}_{ft}^{NFP} = \tilde{h}_t\left(c_{sfh} - c_{seh}\right) + M, \tag{3.10}$$

$$s_{bft} = h_t\left(c_{bfh} - c_{bdl}\right). \tag{3.11}$$

$$\pi_{bdt}^{milk} \leq \tilde{s}_{dt}^{NFP}, \tag{3.12}$$

$$\pi_{bdt}^{milk} \leq s_{bdt}. \tag{3.13}$$

$$N_{bht} \leq N_{bh(t+1)}. \tag{3.14}$$

Lemma 6

The absolute cost of NFP-signaling via the sandals market is constant in time for quality experts, even as the markets grow in size.

Proof: The proof follows directly from equation (3.10).□

[100] Analogue to footnote 96, I assume that one period provides enough demand at $p_{sht} = c_{seh}$ to sell the necessary $\frac{\tilde{h}_t}{1+r}$ units (and also to sell $\frac{h_t}{1+r}$ boots at c_{bdl}) throughout this paper. In case of $\tilde{h}_t$, the issue becomes more subtle as it is linked to when the expected cost M occurs. As shown in Appendix C.5, it is sufficient to assume that "market for the poor" is large relative to the "market for the rich".

Moreover, note that for discounters, this cost $\left(\tilde{s}_{dt}^{NFP}\right)$ varies with time and is adjusted accordingly (via the time-varying $\tilde{h}_t$) such that no mimicking occurs.

Lemma 7
If the optimal signal is creating a NFP firm in the sandals market (particularly implying $\tilde{s}_{et}^{NFP} < s_{bet}$), the period-profit made by each high-reputation firm is constant.

Proof: Combining conditions (3.9) (binding) and (3.10):

$$\frac{1}{1-\frac{1+\hat{g}}{1+r}} \frac{D_t}{\left(N_{bht}+1\right)^2} \left(k - c_{beh}\right)^2 = M.$$

The number of firms with a high-quality reputation in the boots market, N_{bht}, is growing in equilibrium to ensure that $\frac{D_t}{(N_{bht}+1)^2}$ remains constant.[101] Consequently, condition (3.14) is slack, as required for condition (3.9) to be binding, $\pi_{bet}^{cournot,period} = \frac{D_t}{(N_{bht}+1)^2}\left(k - c_{beh}\right)^2$ is constant, and hence $\hat{g} = 0$.□

Regarding the conditions (3.9) - (3.14), note that again either (3.9) or (3.14) is binding. Moreover, condition (3.13) is always binding and condition (3.12) is binding unless this implies a negative $\tilde{h}_t$ (i.e. always binding for sufficiently high t). The latter two being binding is again due to the selection of the equilibrium that is supported by backward induction. Finally note that M is assumed to be constant, which the fundamental reason for Lemma 7 to hold (see section 3.4 for a mechanism where M drops to zero after the first producer of high-quality sandals is operational).

3.3.3.2 Benefiting from Reputation

Firms that have obtained a high-quality reputation in the boots market could enter the sandals market directly with a high reputation and make a positive period-profit until a NFP firm is created and the market price p_{sht} becomes $p_{sht}^{NFP} = c_{seh}$. The expected cost of doing so is M. I assume that M is sufficiently large relative to the demand for high-quality sandals - in particular relative to the share of the "poor" (people that would buy high-quality sandals in principle) that buy at a price of c_{seh}. As shown in Theorem 4 below, this assumption implies that firms with a high-quality reputation (in the boots-market) do not benefit from it in the sandals market. A sufficient assumption is the following:

$$\frac{M}{Q_{sh1}^{D}\left(c_{seh}\right)} > \frac{\left(j - c_{seh}\right)}{4\left(1-\frac{1+g}{1+r}\right)}, \tag{3.15}$$

Intuitively, this is another condition stating that "the market for the poor" by itself is unattractive for firms. Condition (3.15) is a stronger assumption than necessary. The necessary and sufficient condition, which is considerably more difficult to interpret, is provided in the proof of Theorem 4 below.

[101] The equilibrium is again assumed to have a constant rate of change of $\pi_{bet}^{cournot,period}$ if $\tilde{s}_{et}^{NFP} < s_{bet}$. See Appendix C.2 for a discussion.

3.3.3.3 Equilibrium

The previous lemmas and assumptions lead to Theorem 4:

Theorem 4

1) *If the initial demand in the boots market, namely D_0, is sufficiently low, the initially optimal form of market entry is signaling within the boots market.*

2) *As this market grows over time (as D_t grows with rate g), there exists a point in time $t = \tau$ when the optimal form of market entry becomes "NFP-signaling".*

3) *For $t > \tau$, the profit each firm makes is constant, which implies that the number of firms in the market for high-quality boots is increasing. Consequently, NFP firms are not just a "credible threat" but in fact created.*

4) *Under the assumed condition (3.15) (and even under a weaker version), no firm that already has a high-quality reputation will invest M in order to operate in the sandals market.*

Proof: See Appendix C.1.[102]

Theorem 4 states that if the boots market is sufficiently small initially, quality experts enter initially by producing high-quality boots with a loss. The number of firms in that market then remains constant until at some point (denoted τ), a NFP firm is created in the market for high-quality sandals. Due to condition (3.15), this NFP firm, created by a quality expert, is the first to ever offer high-quality sandals. After confirming the high-quality reputation, the quality expert then enters the market for high-quality boots with an FP status.

3.3.4 Role of the NFP Status

In the mechanism above, the NFP status is critical as it avoids signaling costs. Since quality claims are not enforceable, the NFP status is what guarantees that only quality experts offer the high-quality sandals and consequently benefit from *instant* high-reputation in the sandals market. Specifically, I assume:

- quality claims are not enforceable (i.e. sandal producers cannot be sued for producing low-quality sandals, especially once they have stopped operations),
- the NFP status is enforceable (i.e. NFP sandal producers cannot exit the market (or remain in it) with a profit; the owners would be sued). This implies the "non-distribution constraint" of NFP firms to hold perfectly (see Hansmann, 1980).

[102]Intuitively, s_{bet} is initially increasing since it is tied (via h_t) to s_{bdt}, which must be increasing to avoid mimicking from becoming optimal as the markets grow. Due to the increasing s_{bet}, entry is not attractive for quality experts after the initial period. Once "NFP signaling" becomes optimal, the relevant signaling cost is now $\tilde{s}_{bet}^{NFP}$ for quality experts, which is no longer tied to s_{bdt} or $\tilde{s}_{bdt}^{NFP}$. While the latter are still increasing to avoid mimicking, $\tilde{s}_{bet}^{NFP}$ is constant. Now quality experts create NFP firms (and enter subsequently) until entry is not profitable as a larger N_{bht} implies lower prices and profits.

Suppose instead a firm would commit to selling sandals at a discounted price such as c_{seh}. While this leads to the equivalent consumer surplus if these sandals are indeed of high quality, consumers cannot rationally believe the high-quality claim. Discounters can mimic this approach, but produce low quality sandals and make a profit of:

$$\pi_{sdt}^{milk} = \frac{Q_{sht}^{D}\left(c_{seh}\right)}{N_{sht}}\left(c_{seh} - c_{sdl}\right) > 0.$$

After one period, consumers learn about the true type of the firm and the (high-quality) demand for that firm drops to 0. Note that this results in an overall profit of π_{sdt}^{milk}; since no high-quality sandals are sold, no development-costs (M) need to be invested.

The NFP status, however, ensures incentive-compatibility. Given that discounters cannot make a profit in the first period and they can only carry the reputation to the boots market after one period, they have no incentive to establish a NFP firm. Quality experts, however, have an incentive to establish a NFP firm once $t \geq \tau$. Consequently, it is rational to believe that a newly established NFP firm is a quality expert.

The above is true for any discounted price larger than c_{sdl}. Alternatively, quality experts could operate a FP firm and commit to a discounted price $p \leq c_{sdl}$ to ensure incentive compatibility. However, this results in a signaling cost that is strictly larger than M and is consequently not optimal.[103]

In sum, no quantity of high-quality sandals will be demanded if they are sold by a FP firm at a price $p > c_{sdl}$ unless that firm has previously established a high reputation. Similarly, promising to sell any quantity at $p > c_{sdl}$ is not a sufficient signal by itself. Selling at a price $p \leq c_{sdl}$ (as a FP firm) results in a high reputation once enough units have been sold, but this strategy is dominated by creating an NFP firm. Furthermore, selling at a price $p > c_{sdl}$ and promising to sell at $p < c_{sdl}$ later is not a credible way of obtaining instant high reputation. Therefore, any form of signaling by producing sandals with a FP firm is either infeasible or dominated by using a NFP firm.

3.4 Extension: Motivated Workers

A key characteristic of the mechanism outlined above is that NFP firms are created even though this creation is costly (the development cost M) and no pro-social motivation of any sort is included in the model. The fundamental reason is that NFP firms allow for instant high-quality reputation in the sandals market. However, the dynamic equilibrium described in section 3.3.3.3 implies that NFP entry - and hence the creation of new and additional NFP firms - occurs regularly once $t > \tau$. Moreover, as mentioned in the introduction, it can be observed that firms that are already active with a FP firm in a market "for the rich" create NFP firms. So far, this cannot be explained with the mechanism presented above. Finally,

[103] The development costs M have to be spent in any case. Therefore, even if one sandal sold at a price $p \leq c_{sdl} < c_{seh}$ was sufficient to signal being a quality expert, this is more expensive than establishing an NFP firm. Moreover, given $\pi_{sdt}^{milk} > 0$, the number of sandals needed to be sold at this price is generally larger than one.

I have assumed above that the cost M is constant in both time and the number of firms in the high-quality sandals market. Given that the cost M is in part motivated by the fact that high-quality sandals are a novel product, this assumption is potentially too strong in the long run.

The (expected) cost of NFP creation being constant not only in time but also in the number of NFP firms is also the underlying reason for the former two issues. This can be addressed by extending the setting to include motivated workers. The idea of workers caring for the impact of their work is common in the literature (in the context of NFP firms, see for example Ghatak & Müller, 2011).

I model the production/labor market side in the following way. First, potential workers have the simple utility function $u = w + \alpha i - \bar{u}$, where w denotes wage, $\bar{u}$ is the outside option (to work elsewhere), $0 < \alpha < 1$, and i denotes impact of their work on "the poor". More specifically, i is defined as the increase in consumer surplus of consumers in the sandals market that can be attributed to a firm entering (and remaining in) the market per worker of that firm.

Second, regarding the cost of production, I assume $c_{sfa}(p) = A_{sfa} w_{sap}$; where A_{sfa} denotes the productivity of a firm of type f to produce sandals of quality a and w_{sap} the wage paid to workers to produce sandals of the (believed[104]) quality a that are sold at price p. Since workers only care about "the poor", who do not demand any boots, the cost of producing boots is not affected.

Regarding the market for low-quality sandals, which is perfectly competitive, this results in a constant cost of production and a constant price p^*_{slt} (see Appendix C.4). This is analogue to the sections 3.2 and 3.3. The market for low-quality boots is not affected by the introduction of motivated workers in the sandals market.

Regarding high-quality production, there are important differences to section 3.3. These are driven by the fact that the first (NFP) firm to operate in the high-quality sandals market has a significant impact on consumer surplus, while the effect of subsequent firms is generally smaller, and under the current assumption of constant returns to scale, even zero. This is the major aspect that workers internalize: they receive utility from being part of a new firm that sells a product to the poor that makes their lives better, while they receive no utility from being part of a firm that merely copies the product and sells it at the same price to the same people that are also able to buy the product from the former firm. The consequences of this are outlined in Theorem 5:

Theorem 5

1) *With motivated workers, the expected cost of establishing a firm that produces high-quality sandals (M) can drop to zero after the first such firm has successfully been established, without causing an unbounded number of firm entries.*

[104] Similar to consumers only trusting NFP firms instantly, the argument could be extended that FP (but not NFP) firms have an incentive to cheat workers regarding the produced quality. This is not necessary in the current setting as the consumer beliefs are sufficient to guarantee that a NFP status is optimal. Note though, that is another potential advantage of the NFP status.

2) Specifically, if M drops to zero and producing high-quality boots is a superior signal to producing high-quality sandals, only one NFP firm is created at $t = \tau$. This leads to a unique increase in the number of firms in the high-quality boots market (N_{bh}) by at most one.

3) If the "inside firms" of the boots market can overcome the associated coordination problem, it is an "inside firm" that creates this NFP firm. This is guaranteed to be the case for monopolists.

Proof: See Appendix C.3.

Intuitively, the expected cost M is not allowed to drop to zero in section 3.3 because there, the cost of "NFP signaling" always equals M for quality experts. If M dropped to zero, there would be an explosion of NFP firms. With motivated workers, however, the first NFP firm can operate cheaper, and due to its NFP status, the price (p_{sht}) equals those cheaper production costs. All following firms, even NFP firms established by quality experts, face a loss of producing at this price, and hence the cost M is no longer required to prevent unboundedly many of them entering the market.

After the first NFP firm has been established, the signaling options are therefore to either produce high-quality boots or high-quality sandals at a loss. If the loss of doing the former is lower for quality experts relative to discounters than doing the latter, producing high-quality boots becomes again the best possible signal - as it was for $t < \tau$ (see Theorem 4.1). Consequently, except for the creation of one NFP firm at $t = \tau$, the situation is as described in section 3.2.

If the creator of the NFP firm is not yet a high-reputation producer in the boots market, she will enter the high-quality boots market. However, expecting this, the high-reputation boots producers have an incentive to create the NFP firm themselves to avoid such a market entry. For a monopolist, it is always profitable and also obvious that it is the monopolist himself that needs to create the NFP firm. For $N_{bh} > 1$, the "inside firms" need to coordinate regarding whom to create the NFP firm and how to split the costs. It is no longer profitable for a single firm to spend M in order to prevent an outsider from entering.

3.5 Welfare & Policy Implications

In the mechanisms presented in the sections 3.3 and 3.4, NFP firms create high-quality sandals and are universally beneficial to "the poor", i.e. people with a low willingness to pay for boots. They obtain access to a high-quality product (sandals) at production costs - or in case of section 3.4 even cheaper.

The mechanisms are also beneficial to the "rich" consumers, provided time has progressed sufficiently far (i.e. for a sufficiently large t). In the case outlined in section 3.3, the number of firms in the high-quality boots market (N_{bht}) would be constant absent of the possibility of "NFP entry" (see section 3.2). With "NFP entry", however, ($N_{bht} + 1$) is growing constantly at the rate $\sqrt{1+g}-1$, hence eventually, N^*_{bht} is guaranteed to be larger than equilibrium number of firms in isolation obtained in section 2 $\left(N^{*isolation}_{bht}\right)$. A larger number of firms implies lower prices and higher quantities, and hence higher consumer welfare. However, for low values of t, the anticipated entry of additional boots producers lead to $N^*_{bht} \leq N^{*isolation}_{bht}$ (at least for $t \leq \tau$; see proof of Theorem 4.2). This implies higher prices for high-quality boots compared to a case where no signaling through another market is possible. The latter effect is mitigated if τ is large. In that case, the initial number of firms (N_{bh0}) is close to $N^{*isolation}_{bht}$ and the condition that N_{bht} is non-decreasing guarantees it to remain at that level (for $t \leq \tau$). However, a large τ also implies that it takes longer for entry via NFP creation to occur and the benefits of "large t" to materialize.

The mechanism outlined in section 3.4 is qualitatively similar regarding welfare. The key difference is that N_{bh} increases by at most one after $t=\tau$. Consequently, $N^*_{bht} = N^{*isolation}_{bht}$ either initially, or eventually, or both. So initially, the consumer welfare of "the rich" is weakly lower and eventually, it is weakly higher.

Regarding potential policy interventions, I first consider markets that satisfy the assumptions made in this paper and will therefore experience the creation of NFP firms eventually. For these markets, the relevant task is not to create the initial conditions (as these were set in the past), but to affect the environment for $t>0$, and typically for $0<t<\tau$. This implies that condition (3.14) is binding: the number of firms in the high-quality market cannot be decreasing. Imposing $t<\tau$ sets the focus on those markets that have not yet experienced NFP entry, which is arguably more interesting for policy. As the number of firms in the high-quality boots market is constant at N^*_{bh0} for $0<t<\tau$, the period in which a NFP firm is created is given by:

$$\tau = \frac{ln\left(\frac{M(c_{bdh}-c_{bdl})\left(N^*_{bh0}+1\right)}{D_0(c_{beh}-c_{bdl})(k-c_{beh})\left(k-c_{bdl}-\frac{N^*_{bh0}}{N^*_{bh0}+1}(k-c_{beh})\right)}\right)}{ln\left(1+g\right)}.$$

A lower value of τ is beneficial for both "rich" and "poor" consumers (but harms the profits of the "inside firms"). If, for example, a policy can lower the cost M (e.g. by supporting research and development), this leads to an earlier creation of the first NFP firm. The optimal number of "inside firms" (N_{bht}) is, if anything, lower for an unexpected decrease in M (see proof of Theorem 4.2). However, due to condition (3.14), N_{bht} remains constant.

Focusing on the extension with motivated workers, workers that are willing to work for a lower wage for the initial NFP firm imply a larger set of markets that can be served by a NFP firm.[105] Therefore, policy interventions that increase the motivation of workers, or offer other benefits for working at NFP firms, expand the set of markets that can benefit from the mechanism outlined in this paper.

[105] Essentially, the c_{seh} in condition (3.15) becomes lower.

3.6 Conclusion

When quality is only observed after purchase and companies can operate several firms with different organizational forms, NFP firms are created in "markets for the poor" as signaling devices under the assumptions outlined above. A key part of these assumptions concern the demand functions. The parameters are assumed such that (1) FP firms ignore "the market for the poor" (at least in isolation) but (2) NFP firms are able to operate in this market (see condition (3.8)). The former is in line with the observations of Yunus (2008). While assuming (2) on top of (1) is restrictive to some degree, it directs the focus to those environments that can benefit from "NFP signaling". Moreover, introducing motivated agents (e.g. motivated workers) partially alleviates the restrictions on the demand parameters.

For environments that meet these assumptions, NFP firms are eventually created. No pro-social motivation is required, but employing motivated workers in the model (for example) regulates the number of created NFP firms.

In terms of the four stylized empirical facts listed in section 3.1, the first point is taken into the model as possible action of the companies. Under the assumptions mentioned above, the outcome is in line with the second point and companies create NFP firms as their optimal action to enter the FP market - both with and without motivated workers. With motivated workers, also the third point is accounted for: it can be optimal for FP companies to create a NFP firm in order to avoid entry of an additional competitor. Finally, the fourth point is also part of the model's outcome: only as markets grow to be sufficiently large, market entry via NFP creation is optimal.

Once NFP firms are created, the mechanism outlined in this paper is beneficial to both "the rich" and "the poor". The former benefit from lower prices, while the latter benefit from being offered a tailored quality product at production cost. Initially though, "the rich" are potentially harmed by this mechanism: the expectation of future firm entry can lead to fewer firms in the early periods, which implies higher prices. Consequently, given the conditions for NFP entry are (and remain to be) met, any policy that leads to earlier NFP creation is beneficial for all consumers (but harmful to firm profits).

It is left to future research to allow for interactions of the demands for the different goods. For example, the model could be extended such that the demand for high-quality boots decreases once high-quality sandals become available. Note though that such forms of cannibalization between the markets would not affect the results qualitatively as none of the driving forces are affected: no changes occur prior to NFP signaling being optimal, markets and signaling costs still grow over time, and once NFP entry is optimal, the cost of signaling remains constant for quality experts. Consequently, the most important associated future research arguably concerns empirically analyzing these markets and the testable predictions provided in the current paper.

Appendix

A Supplementary Material and Proofs to Chapter 1

A.1 Synthetic Control Method

The synthetic control method has been introduced in the seminal papers of Abadie and coauthors (Abadie & Gardeazabal, 2003; Abadie et al., 2010, 2015). It is a difference-in-difference estimate where instead of one comparison, several comparisons are made, based on the pre-event period. Under rather strong assumptions, a variant of the synthetic control method can be applied to analyze the 'Brexit effect'. This is done here for the example of London. The result suggests a short term effect of the Brexit vote of one to two months.

Method

The synthetic control method requires panel data with a 'donor pool' of other entities that can be used to model the entity in question in the post-treatment period. In my case, the entity in question is racial or religious hate crime. For the donor pool, I require time series of variables that are (1) affected similarly by factors that are not the Brexit vote, but (2) unaffected by the Brexit vote. Other types of crime are my major candidates - although several issues require attention. Provided sufficiently similar types are chosen, the requirement (1) is arguably only a weak assumption. However, requirement (2), the unaffectedness by the Brexit vote, is a rather strong assumption. Analyzing the time series of the relevant crime types for 'Brexit effects' can test this assumption to some extent. Another problem is that for the London data, only aggregate information on crime is available to us. Hence a crime can be part of the aggregate of racial or religious hate crime, but also part of the aggregate of harassment crime. In principle, this violates the first requirement mechanically. However, racial or religious hate crimes are rather rare.[106] Thus while there will be bias, I expect that bias to be small. Moreover, if anything, that bias will negate any effects of the Brexit vote specifically to racial or religious hate crime, so the results below would be too conservative.

Another challenge when using a panel of different crime types is the fact that other than for example states or countries, crime types do not have entity-level properties (such as GDP, or inflation rate; see Abadie, Diamond & Hainmueller, 2015). The synthetic control method relies on such properties as predictors of the main effect of interest. Plainly using month of the year or a time trend is also infeasible because it is identical for all entities. However, a key property of crime is its behavior across time, especially its seasonality. I generate averages for each month of the year, as well as averages for each year, and use these as predictor variables.[107] That way, each entity in the panel has generally unique predictor variables.

[106]Example: For Manchester, where the data is on the incident-level, at most 21.5% of any given crime-type (public order offense) used in the donor pool is flagged as racial hate crime. This is an outlier, with the second most being violence without injury at 3.2%. Given that public order offenses (which are part of state based crime) were only given weight 0.001, strong concerns about that outlier are arguably unwarranted.

[107]The average of the given year or calendar month was calculated for each observation based on all other

Result

The synthetic control result is fully consistent with the visual inspection in section 1.3. I use again July 2016 as the period affected by Brexit (results are robust to using June 2016). The variables of the donor pool and their respective weights are depicted in Table A.1 below. The result of using the synthetic control method with this panel data is illustrated in Figure A.1. It becomes clear that at the time of the event, the synthetic racial or religious hate crime completely lacks the spike of the true racial or religious hate crime. This is possibly the case for June already and August thereafter, but seems to be over by September 2016, providing evidence for the effect being temporary. There is another spike in 2017 which coincides with the month after the Manchester terror attacks, which is in line with the findings of Ivandic, Kirchmaier and Machin (2018), and hardly related to the Brexit vote itself.

Crime	Control Weight
Assault with Injury	0.004
Common Assault	0.004
Harassment	0.11
Other Violence	0.016
Criminal Damage: Dwelling	0.009
Criminal Damage: Car	0.007
Criminal Damage: Other Building	0.013
Criminal Damage: Other	0.01
Rape	0.802
Other Sexual	0.016
Theft from Person	0.005
Theft from Shop	0.005
State Based Offense	0.001

Table A.1: Synthetic Control Weight (London)

observations, although using all observations lead to almost exactly the same results.

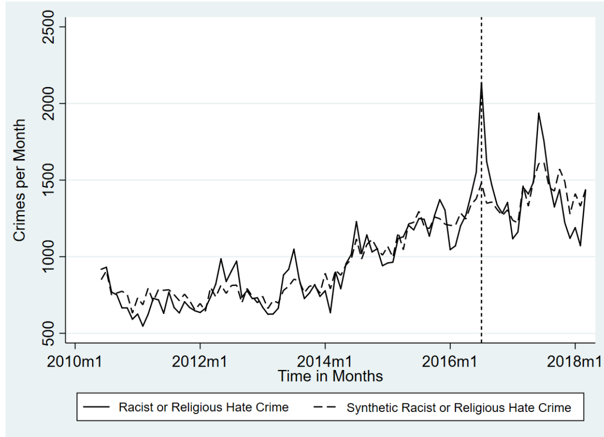
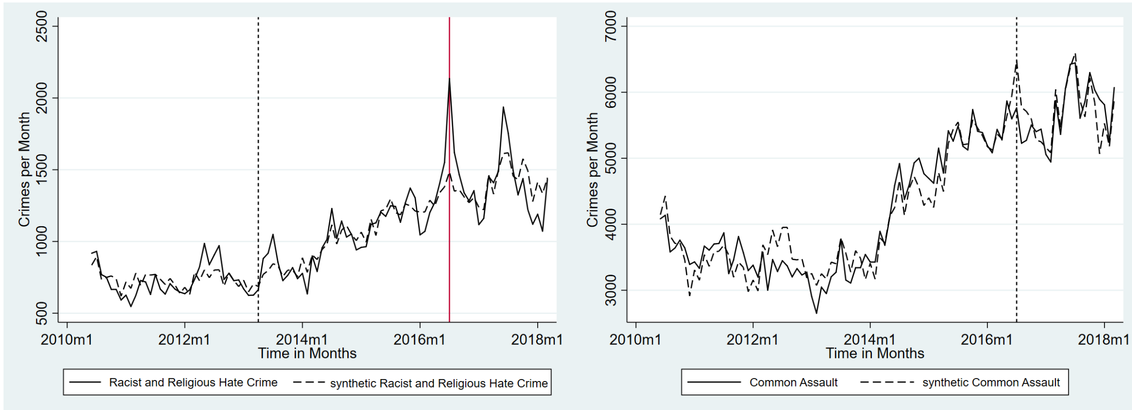


Figure A.1: Synthetic Control (London)

What is rather surprising is that rape is such an important part of the synthetic construct. The weights are generally quite different from Figure 1.5 and footnote 26, which served (other than data availability) as guide which crime types to include. However, as shown in Figure A.1, the match of the pre-period is reasonable, at least compared to the massive discrepancy around the Brexit vote. I also emphasize that it was a key argument, that only few crimes of each type are flagged as hate crimes, so Figure 1.5 is indeed a weak guideline. Moreover, both in-time and 'in-crime' placebo exercises with a random other month before Brexit and a random other crime-type are in line with the analysis. The in-time placebo produces virtually the same result as Figure A.1, and the in-crime placebo shows if anything that only the synthetic crime spikes at Brexit, which makes sense since that includes partially racial or religious hate crime (see Figure A.2). Finally, in a simple regression of the monthly counts of racial or religious hate crimes on the counts of all other crime types of the donor pool, it is indeed rape that has the highest conditional correlation.

The absence of a spike for the synthetic crime suggests that the fact that some crime was classified both as hate as well as other crime is indeed a minor issue. Moreover, it suggests that crime that is not racially or religiously aggravated was not affected in a noticeable magnitude. Additionally, I use each individual time series in the donor pool as independent variable in a regression equivalent to the one used for the synthetic control method (month and year averages), but also add a either a dummy for July 2016 or post (and including) July 2016 (i.e. Chow test for the intercept). Out of these 26 regressions, only one contains an estimator for the dummy variable that is significant on the 10%-level (July 2016 dummy in the harassment regression).



Note: The dashed vertical line divides the pre- from the post-period. In the right-hand figure, it coincides with the Brexit vote as this is the shock of interest. The left-hand figure depicts an 'in-time placebo', hence the pre-post divide is at a random other point in time to show that not the mere pre-post divide is causing the difference but the shock of interest. July 2016, i.e. the month after the Brexit vote, is indicated with a solid red vertical line in that figure.

Figure A.2: London: In-Time and In-Crime Synthetic Control Placebos

A.2 Difference-in-Difference Estimation

For London and Manchester, the data sources contain racial or religious hate crime but also total crime. This can be used for a difference-in-difference estimation. As the magnitudes are dramatically different, only the relative effect is analyzed. The assumption that the Brexit vote did not affect the overall crime level is required for this analysis to be valid (along with parallel trends).

Column 1 of Table A.2 shows the simplest possible difference-in-difference approach, regressing log(monthly crimes) on a dummy for July 2016 (denoted 'Brexit'), the crime-type dummy (hate or total crime), and the term of interest: their interaction (standard errors in parentheses). As shown in Figure A.3 though, this result is likely misleading. Hate crime seems to have a slightly different trend and more pronounced seasonal effects than total crime.

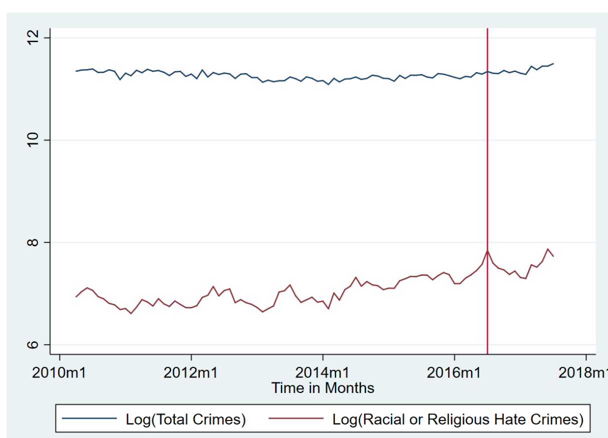


Figure A.3: Trends

Column 2 shows the result where a type specific time trend (time and time squared) and type specific seasonalities (month-of-the-year dummies) are added as controls. The magnitude of the effect is now at 0.27 (0.25 if the insignificant total Brexit effect is subtracted) for hate crime, highly similar to the 0.21 obtained in section 1.3.2.

	(1)	(2)	(3)
	log(Crimes)	log(Crimes)	log(Crimes)
Hate Crime * Brexit	0.70**	0.27***	0.38***
	(0.30)	(0.00)	[0.19, 0.57]
Brexit	0.07	-0.02	-0.13
	(0.21)	(0.08)	[-0.40, 0.15]
Type Fixed Effects	Yes	Yes	Yes
Borough Specific Trends/Season.	No	Yes	Yes
Std. Error Adjusted	No	No	Wild C. Bootstr.
Observations	176	176	2785

Table A.2: Overall Effect of Brexit: Difference-in-Difference Estimation

The standard errors are not adjusted in any way in Table A.2 and consequently invalid (see e.g. Cameron et al., 2008). Obtaining valid standard errors is difficult in this setting. Clustering on the crime type results in only two clusters, prohibitively few for any clustering or adjusted clustering.

Running this difference-in-difference estimation (of column 2) separately for each month in the data reveals that July 2016 results in the highest of all 88 estimates (and the second highest absolute value). In that respect, the estimate is significant from a permutation-test point of view.

The results of repeating this exercise with not only the short term but also potential long run effects are displayed in the first two columns of Table A.3. Running that difference-in-difference estimation separately for each month reveals that while the transitory effect (Hate Crime * Brexit) is the second largest, the long term effect (Hate Crime * Post-Brexit) is in 21 out of 88 cases lower. So again, there does not seem to be a significant long-run effect of the Brexit vote (the 0.40 in column 1 is also not a tail result if that estimation is run for every month respectively). Regarding the short run effect, 0.18 is again close to the 0.21 obtained in section 1.3.2.

Finally, in column 3 of both tables, slightly different data is used. The total crime is split into sub categories. While GMP and the London Metropolitan police do not categorize identical, 14 categories could be matched (losing 49% of the total crimes as measured in July 2016, but a trend comparison still highly resembles Figure A.3). That results in panel data of 15 crime types (including racial and religious hate crime) over 88 months. The advantage is that now clustering the standard error by type results in 15 clusters. Following Cameron et al. (2008), I use wild cluster bootstrapping (95% bootstrapped CI in brackets).

	(1)	(2)	(3)
	log(Crimes)	log(Crimes)	log(Crimes)
Hate Crime * Brexit	0.76***	0.18	0.23***
	(0.24)	(0.11)	[0.08, 0.38]
Hate Crime * Post-Brexit	0.40***	-0.18***	-0.30***
	(0.07)	(0.05)	[-0.52, -0.08]
Brexit	0.08	0.00	-0.12
	(0.17)	(0.08)	[-0.40, 0.16]
Post-Brexit	0.11	0.04	0.35
	(0.05)	(0.04)	[-0.30, 0.99]
Controlling for Type and Brexit	Yes	Yes	Yes
Borough Specific Trends/Season.	No	Yes	Yes
Std. Error Adjusted	No	No	Wild C. Bootstr.
Observations	176	176	2785

Table A.3: Overall Effect of Brexit: Difference-in-Difference Estimation with Long Run Effects

A.3 Proof of Lemma 1

Take the regression $y = X\beta + Z\gamma + \nu$. By the Frisch-Waugh-Lovell theorem, the "classic" OLS estimator is $\hat{\beta}_c = (X'M_Z X)^{-1}X'M_Z y$. The OLS estimator obtained from the regression $M_Z y = X\beta + \varepsilon$ is $\hat{\beta}_p = (X'X)^{-1}X'M_Z y$. Further, $X'X - X'M_Z X = X'Z(Z'Z)^{-1}Z'X$, which is positive semi-definite.[108] In the single variable case, X is a vector (hence now denoted x) and β is a scalar, and consequently $|\hat{\beta}_c| \geq |\hat{\beta}_p|$. Unfortunately, the same cannot be stated for the multi variable case.

Moreover, $E\left(\hat{\beta}_p|X, Z\right) = \beta - (X'X)^{-1}X'P_Z X\beta$. For the single variable case, this reduces to $E\left(\hat{\beta}_p|x, Z\right) = \left(1 - \frac{\hat{x}'\hat{x}}{x'x}\right)\beta$, where $\hat{x}$ refers to the fitted values of a regression of x on Z. Since $\hat{x}'\hat{x} \geq x'x$ and both sides are positive, there is attenuation bias in $\hat{\beta}_p$. Moreover, since $\hat{x}'\hat{x} = ESS + n\bar{x}$ and $x'x = TSS + n\bar{x}$, the bias is directly related to $R^2_{x,Z}$. Again, little can be said about the multi variable case.

q.e.d.

A.4 Robustness: Alternative Measures & Forms of Cross Validation

In this appendix-section, I examine the robustness of the results with respect to alternative measures and forms of cross validation. The alternative measures concern both the crime-increase in July 2016 and the heterogeneity that is captured with the candidate variables.

Table A.4 shows the lasso models for different measures for the crime-increase in July 2016 (the increase being the difference between predicted and observed hate crime). The first two

[108] Another argument for the single variable case is that if x is regressed on Z (with error ε), then $x'x = \hat{x}'\hat{x} + \hat{\varepsilon}'\hat{\varepsilon} = \hat{x}'\hat{x} + x'M_Z x$, hence $x'x \geq x'M_Z x$.

columns report again the measure used in the main section. Columns 3 and 4 as well as 5 and 6 show that not using July 2016 (not using June, July, and August 2016 respectively) to obtain the prediction leads to an identical choice of variables. If only the pre-period is used to predict the hate crime in July 2016, then a similar model is obtained for the absolute effect (the difference being the lack of the share of individuals with no qualifications in the model) and an intercept-only model for the relative effect (columns 7 and 8). Using a lagged dependent variable in addition to the prediction-specification in the main text leads again to the same model as in the main text (columns 9 and 10). When the lagged dependent variable is used instead of time squared, a slightly different model is obtained (columns 11 and 12). Instead of using the difference between predicted and observed hate crime as a measure, it is also possible to run a partially penalized regression of equation 1.1 and only penalize the interaction terms. The result of this is depicted in columns 13 and 14. The model of the absolute effect still contains the share of recent immigrants, but now also the share of people born in countries that were part of the EU in 2000 rather than the share of people stating no religion and the share of people with no formal qualification. For the relative effect, an intercept-only model results.

Table A.5 depicts the second measure of heterogeneity that is captured with the candidate variables suggested by Chernozhukov et al. (2018b) - the first being displayed in Table 1.4 in the main text. Here, the predicted effect (that is a function of candidate variables only) is not used for defining bins, but rather used in the simple regression of observed hate crimes on predicted effect. Similar to Table 1.4, conventional significance is only obtained for the absolute effect but significance based on permutation inference is obtained for both the absolute and the relative effect.

The Tables A.6 and A.7 provide different forms of cross validation used in the approach behind Table 1.6 in the main text. Columns 1 to 3 of Table A.6 (columns 3 to 5 of Table A.7) are equivalent to panel A (B) of Table 1.6 except for using 3-fold instead of 10-fold cross validation. The columns 6 and 7 (8 and 9 of Table A.7) differ from the lasso and associated CPSL result in panel A (B) only be not using the 'rule of parsimony' discussed in footnote 45. Similarly, the columns 4 and 5 (6 and 7 of Table A.7) differ in both using 3 folds and disregarding the 'rule of parsimony'. Finally, the columns 1 and 2 of Table A.7 differ from the result in panel B of Table 1.6 by imposing the binding condition of having at least 2 variables included in the splitting based approach. This condition is not binding regarding the absolute increase in hate crimes.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	HC	log(HC)	HC	log(HC)	HC	log(HC)	HC	log(HC)	HC	log(HC)	HC	log(HC)	HC	log(HC)
Rec. Immigr.	16.3		19.0		19.7		15.7		16.5		17.3	0.00	0.05	
No Rel. Stated	11.8		13.8		13.5		11.5		10.1		6.2			
No Quali.	-5.8	-0.02	-6.7	-0.02	-6.9	-0.03			-4.0	-0.01		-0.00		
Born Old EU													0.07	
Measure	As used in this paper and described in section 1.4.2		Not using July 2016 in prediction		Not using June, July, August 2016 in prediction		Only using pre-Brexit period		Using lags in addition to setup described in section 1.4.2		Using lags instead of time squared		Using partly penalized full regression (alike regression 1.1)	

Note: Results of using a simple lasso with standardized data. HC refers to racial or religious hate crime. Dependent variable demeaned, detrended and deseasonalized at the borough level. A lack of any entries means that a constant is chosen by the lasso as the best prediction model. Despite the fact that no direct wealth or income proxy is chosen by the full regression method, section 1.6.3 provides evidence for their importance in this setting. Columns 1 to 12 are cross sectional in July 2016. Columns 13 and 14 use a panel of 88 months and do not use the variables as levels but interacted with a dummy for July 2016. The correlations are unsurprisingly lower.

Table A.4: Lasso Models of Different Measures of Abnormal Hate Crime in July 2016

	(1) Hate Crimes per Population Million	(2) log(Hate Crimes per Population Million)
Predicted Effect	1.28***	0.58
Permutation Significance	**	**
Permutation: 90% Benchmark	0.91 [71%]	0.48 [83%]
Permutation: 95% Benchmark	0.80 [62%]	0.25 [43%]
Candidate Variables	69	69
Observations	42	42
Placebos (Permut. Test)	87	86

Note: Measures of heterogeneity that is captured by the candidate variables. Permutation inference uses other months than July 2016 (the month after Brexit) as placebos. The benchmarking refers to subtracting the 90th/95th percentile of the placebo values. Percentages in brackets indicate how much of the heterogeneity is attributed to the Brexit vote if July 2016 had spatial noise equal to the 90th/95th percentile (using 88 months of data). Method used for the predictions: Lasso with 69 candidate variables (vote and census data, plus a dummy for Manchester). The lasso's inherent attenuation bias is a potential reason for estimates larger than one. The one month where one of the boroughs experienced 0 hate crimes was not used as a placebo for the relative case as the logarithm is not defined.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Alternative Measure for the Captured Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Recent Immigrants	1572***	1237***	902	1139***	1087	842***	1091
No Relig. Stated	806**	641**	337	631***	518	842	545
No Qualific.		-235	-79	-224*	-171	-297	-181
Industry Code D				-2015	-340	-545	-593
No Religion						-166	-1
Sikh						166	17
Method	Splitting	CPSL	Lasso	CPSL	Lasso	CPSL	Lasso
Folds	3	3	3	3	3	10	10
Parsimony Rule	Never	Yes	Yes	No	No	No	No
Frequency	2.4%	NA	NA	NA	NA	NA	NA
Candidate Var.	68	69	69	69	69	69	69
Observations	42	42	42	42	42	42	42

** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (for CPSL and Splitting)*

Table A.6: Results using Different Forms of Cross Validation (Absolute Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Rec. Immigr.						2.02	1.09		
No Quali.	-1.97	-2.26*	-2.51***	-2.55*	-0.34	-1.76**	-1.31	-2.18	-1.67
Ind. Code D		-25.11				-22.17	-8.53	-11.17	-13.65
Ind. Code E	-3.22								
Male								3.36	2.40
No Rel. St.								1.65	0.47
No Religion								-0.88	-0.04
Method	Splitting	Splitting	Splitting	CPSL	Lasso	CPSL	Lasso	CPSL	Lasso
Folds	3	10	3	3	3	3	3	10	10
Pars. Rule	Never	Never	Never	Yes	Yes	No	No	No	No
Frequency	1.5%	1.3%	2.5%	NA	NA	NA	NA	NA	NA
Min 2 Var	Yes	Yes	No	Never	Never	Never	Never	Never	Never
Cand. Var.	68	68	68	69	69	69	69	69	69
Obs.	42	42	42	42	42	42	42	42	42

** p < 0.1, ** p < 0.05, *** p < 0.01 (for CPSL and Splitting)*

Table A.7: Results using Different Forms of Cross Validation (Relative Effects)

A.5 Racially or Religiously Aggravated vs Non-Aggravated Offenses

Figure A.4 shows the types of crime that are relatively frequently racial or religious hate crimes and are hence especially tracked by the Home Office (2017). These crimes are: assault with injury, assault with no injury, harassment, public distress, and criminal damage. The figure depicts data for the 38 forces in England and Wales and both time series are standardized to 100 in April 2014. While an increase in racially or religiously aggravated offenses is observed, non-aggravated offenses appear stable in July 2016.

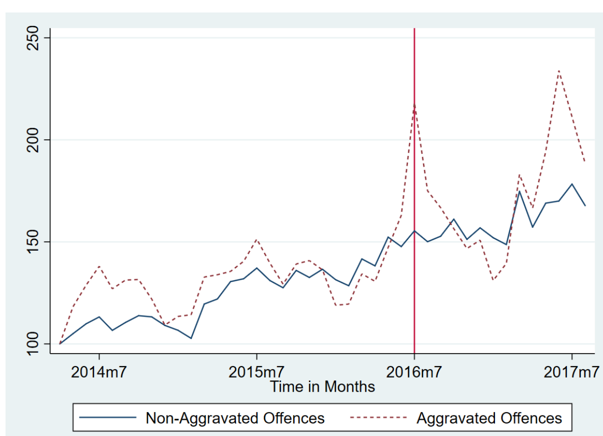


Figure A.4: Racially or Religiously Aggravated vs Non-Aggravated Offenses

A.6 Further Simulations, Visualizations & Regressions

Further Simulations

True DGP	Bias in % of CI	Param. in CI	Mean Freq.	Best	Significant
Noise Only	2.2%	94%	5.2%	NA	6%
X_1	-0.9%	100%	36.1%	78%	99%
$2X_1$	2.2%	99%	92.9%	100%	100%
$2X_2+2X_3$	1.4%	100%	48.0%	100%	100%
$2X_1+X_4+X_5$	2.9%	100%	36.1%	52%	99%

Note: Each simulated 100 times. X_1: share of immigrants arrived within 2 years; X_2: share of remain votes in the Brexit referendum; X_3: share of people in a same-sex civil partnership; X_4: share of people with no qualifications; X_5: share of people not stating a religion. Possible predictors: the 68 candidate variables from the census and vote data.

Table A.8: Simulation Results Using a Single Variable: Bias Small, Coverage High

Table A.8 simulates the search for a single variable that explains best the treatment heterogeneity. When the true model contains more than one variable, it is not necessarily the case that any of those variables is best. In the fifth row of Table A.8, for example, the correlation structure among all 69 candidate variables leads to the case that despite the fact that only X_1, X_4, and X_5 are part of the true DGP, X_7 is almost the best single variable to explain the heterogeneity in an approximation model (it is highly correlated with both X_1 and X_4, the difference to the truly best variable, X_1, are marginal). This is desired in this setting, as it directs the focus to the right boroughs using the most parsimonious model possible.[109]

The second column of Table A.8 indicates the average bias of the coefficients in terms of the length of their 95% confidence interval. The next column indicates how often it is the case that the "true"[110] parameter of the chosen approximating model is contained in the 95% confidence interval. Column four indicates the mean frequency with which the "winning variable" was chosen in each iteration of the simulation. Since the frequency itself is measured across splits and not across simulation-iterations, it is also obtained in section 1.6 and can therefore serve as a measure of similarity between the simulations and the real regressions. The last two columns show to what percentage (across the number of simulation iterations) the best variable was chosen, and to what percentage the chosen variable was significant at the 95% level. The simulation results show that the confidence intervals are, if anything, too large and the bias negligible. While it is not always the case that the best variable is found, it needs to be stressed that it is often the case in my setting of correlated candidate variables that the second or third best variables are almost equivalently informative predictors.

[109] Finding the variables best describing the heterogeneity is the question at hand, and it is not trivial given the issue of multiple hypothesis testing. Finding variables that causally interact with the shock caused by the Brexit vote would be even more desirable in order to pin down the exact mechanisms at play, but as of the best of my knowledge, it is (currently) virtually impossible to do so quantitatively in this setting.

[110] True in the sense of its predictive property (i.e. including omitted variable bias).

Further Time Series Visualizations

Figure A.5 shows the time-series of racial or religious hate crimes for London and the other 37 forces excluding London. While the left-hand side shows the effect to be highly pronounced in London, the right-hand side shows that a clear spike is also visible for the other 37 forces. The phenomenon is consequently not just London-specific, but it was clearly more pronounced there.

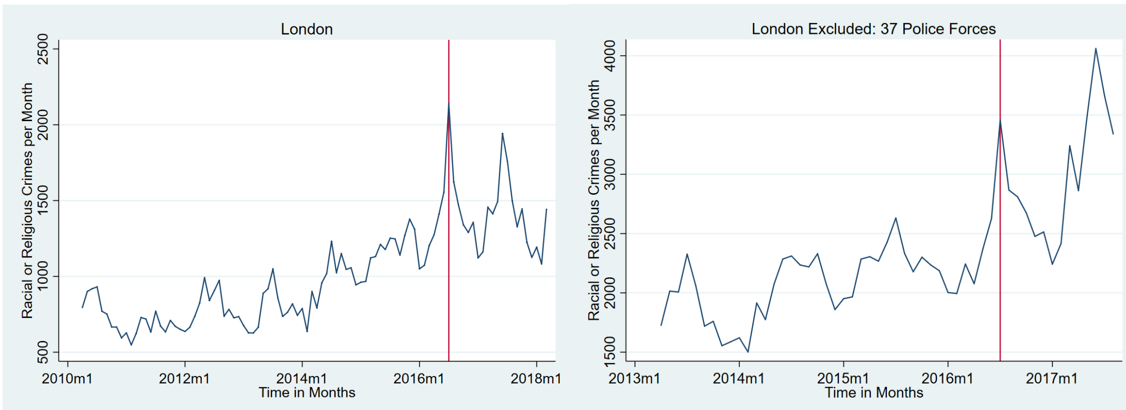


Figure A.5: Racial or Religious Hate Crime over Time: London vs the other 37 Forces

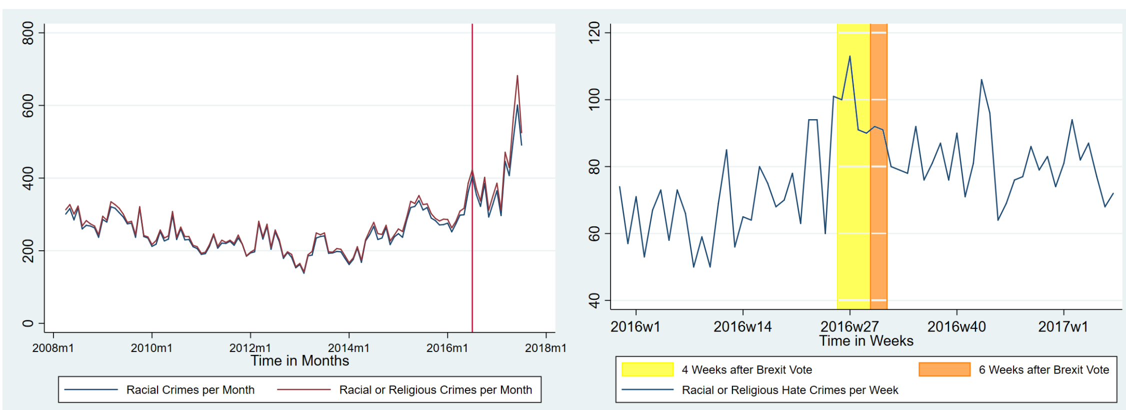


Figure A.6: Racial Hate Crimes over Time: Manchester

Figure A.6 depicts details that are only available for the data from Manchester. The left-hand side shows that 'racial or religious hate crime' and 'racial hate crime' numbers are almost identical (with the possible exception of the spike in summer 2017). This means that very few 'racial or religious hate crime' are recorded to be driven by religious but not racial hate. The right-hand side (as well as Figure A.8) shows weekly instead of monthly data - and Figure A.7 even shows daily data. With higher frequencies, a specific spike is more difficult to see (especially as the spike in Manchester is considerably less pronounced than in London). However, in the weeks after the Brexit vote, there are more (and somewhat higher) peaks than in the rest of the timeseries and fewer 'valleys'.

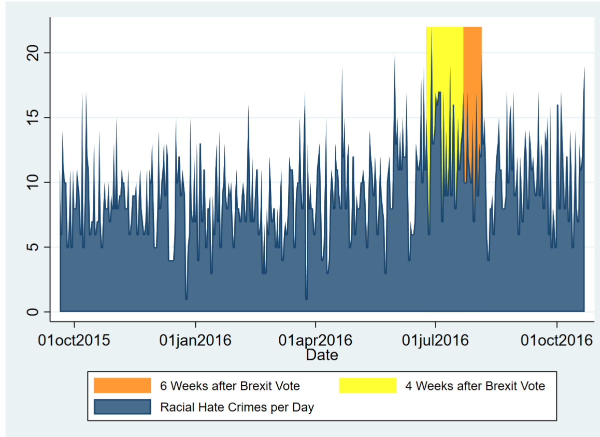


Figure A.7: Racial Hate Crimes over Time: Daily Data for Manchester

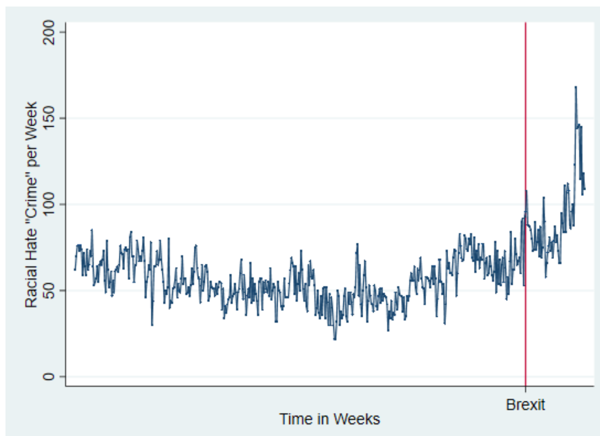


Figure A.8: Racial Hate Crimes over Time: Manchester Complete Weekly Data

Further Regression Tables

Table A.9 depicts the result from regression 1.1 in section 1.4.1, where the remain-voteshare is used as only dimension of heterogeneity. It shows that the increase in hate crime was more pronounced in boroughs with a high remain vote. While this is interesting (and confirmed in Albornoz et al., 2018), it suffers from the problem that it can be viewed as an arbitrary choice from the set of candidate variables.

	(1)
Brexit	-110.03***
Remain X Brexit	2.85***
Time Linear	-27.00***
Time Squared	0.02***
Brexit Def.	July 2016
Area	Borough
Time	Month
Time X Area FE	Yes
Area FE X Time Linear	Yes
Area FE X Time Squared	Yes
Cluster-Level	Area
Clusters	42
Observations	3696

** p < 0.1, ** p < 0.05, *** p < 0.01*

Table A.9: Regressing Racial or Religious Crime on the Sole Heterogeneity Dimension Remain Vote

Table A.10 shows the results from the regressions 1.2 and 1.3. In the first row, it depicts the average effect of July 2016 - i.e. the month after the Brexit vote.

	(1)	(2)
July 2016	51.92***	0.2057***
Time Linear	-26.4***	-0.2164***
Time Squared	0.02***	0.0002***
Month X Borough FE	Yes	Yes
Borough FE X Time Linear	Yes	Yes
Borough FE X Time Squared	Yes	Yes
Cluster-Level	Borough	Borough
Clusters	42	42
Observations	3696	3696

** p < 0.1, ** p < 0.05, *** p < 0.01*

Table A.10: Regression Equations 1.2 and 1.3

Table A.11 shows the effects of the most important single variable (in column 1), and of the variables chosen in the resulting model of heterogeneity on hate crimes prior to the Brexit vote (in column 2). Not surprisingly, the share of recent immigrants is also positively correlated with hate crimes in the months before the Brexit vote (opportunity channel). Moreover, the share of people with no formal qualifications is positively correlated with hate crimes before the Brexit vote. This is in contrast with it being negatively associated with the increase of hate crimes in the month after the Brexit vote (see Table 1.6, panel A).

	(1)	(2)
Recent Immigrants	1782.51***	2003.24***
No Religion Stated		417.51**
No Qualifications		315.80***
Time Linear	-27.13***	-27.13***
Time Squared	0.02***	0.02***
Month FE	Yes	Yes
Cluster-Level	Borough	Borough
Clusters	42	42
Observations	3108	3108

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Regression on the Pre-Brexit Period Using Heterogeneity Variables instead of Borough Fixed Effects

Residual Distributions

Figure A.9 shows the distribution of the July 2016-residuals of the regressions 1.2 and 1.3.

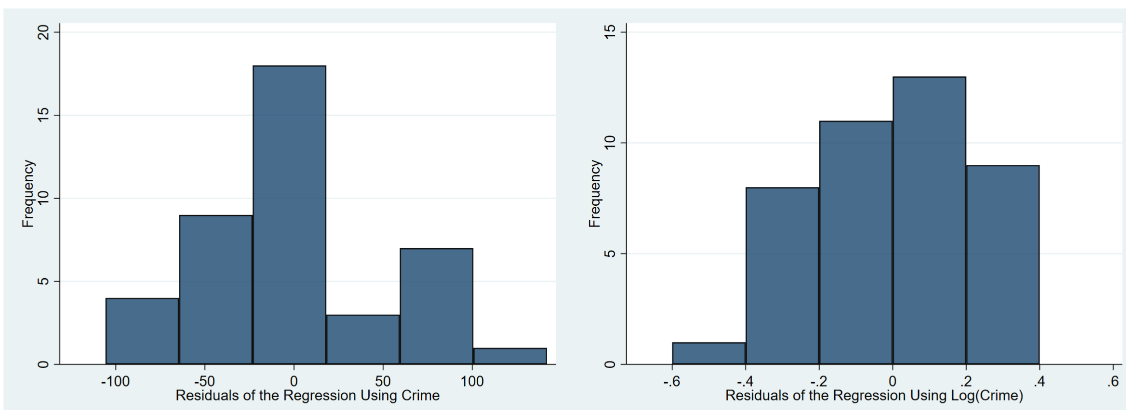


Figure A.9: Sample Distribution of July 2016-Residuals of Regression 1.2 and 1.3

A.7 Seed Dependency Examples: Repeated Splitting vs Single Split

In this appendix-section, the importance of the seed-dependency issue is visualized for the current application. The Tables A.12 and A.14 illustrate the problem in the task of finding the single most important candidate variable to predict the heterogeneity of the increase in hate crime. The first four columns show the result of the splitting based method outlined in section 1.5.2. While this approach uses 1000 seeds, it could still be seed dependent if 1000 is not enough. However, the fact that the results are highly similar when the seed numbers 1-1000, 1001-2000, 2001-3000, and 3001-4000 are used shows that the splitting based approach has virtually overcome the seed-dependency problem. Conversely, for a single split, the seed dependency problem is severe in the current application. As shown in the last four columns, very different results (different variables in particular) occur for each seed if just a single split is used.

The Tables A.13 and A.15 show the same for the full model rather than a single variable. The only difference in the splitting based approach is the inclusion or exclusion of one insignificant variable (share of people with no qualifications in the case regarding the absolute increase in hate crime). Conversely, vastly different models occur based on a single seed - stressing again the importance of seed-dependency in the current application.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rec. Immigr.	2321***	2273***	2310***	2261***	2492***		2790***	
No Rel. St.								2401***
Social Gr. C						-749***		
Method	Splitting	Splitting	Splitting	Splitting	1 Split	1 Split	1 Split	1 Split
Seed	1-1000	1001-2000	2001-3000	3001-4000	1	2	3	4
Freq	46.2%	46.5%	46.0%	46.8%	NA	NA	NA	NA
Cand. Var.	68	68	68	68	68	68	68	68
Obs.	42	42	42	42	42	42	42	42

** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Table A.12: Seed Dependency Issue: Single Variable, Levels

	(1)	(2)	(3)	(4)	(5)	(6)
Recent Immigrants	1150*	1581***	1469***	1962**		2312**
No Religion Stated	715**	791**	872**	863	933*	
Sikh					460	
No Qualifications	-274					-368
Fully Deprived					-121	
Self Employed					609	
Mixed Ethn						412
Age 0 - 15				348		
Age 16 - 29				-310		
Age 30 - 65					-74	
Industry F				-43		
Industry G					256	
Industry J					161	
Industry L					-1356	
Industry O					-387	
Centr. Heating						3892*
Method	Splitting	Splitting	Splitting	1 Split	1 Split	1 Split
Seed	1-1000	1001-2000	2001-3000	1	2	3
Freq	1.6%	1.7%	2.0%	NA	NA	NA
Candidate Var.	68	68	68	68	68	68
Observations	42	42	42	42	42	42

** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Table A.13: Seed Dependency Issue: Full Model, Levels

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No Quali.	-2.47***	-2.38***	-2.41***	-2.40***				
Industry D					-57.63**			
Remain Vote						0.77*	0.69*	
No Rel. St.								10.68**
Method	Splitting	Splitting	Splitting	Splitting	1 Split	1 Split	1 Split	1 Split
Seed	1-1000	1001-2000	2001-3000	3001-4000	1	2	3	4
Freq	26.6%	28.5%	29.7%	29.0%	NA	NA	NA	NA
Cand. Var.	68	68	68	68	68	68	68	68
Obs.	42	42	42	42	42	42	42	42

** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Table A.14: Seed Dependency Issue: Single Variable, Logs

	(1)	(2)	(3)	(6)	(7)	(8)
Recent Immigrants				5.12*		0.90
Jewish						-0.01
No Qualifications	-2.67***	-2.37**	-2.50**			-1.83
Industry E					-84.07	
Industry G					4.18	
Industry J					4.57	
Industry O					-1.70	
Voted Remain					-0.94	
Method	Splitting	Splitting	Splitting	1 Split	1 Split	1 Split
Seed	1-1000	1001-2000	2001-3000	1	2	3
Freq	2.0%	2.0%	2.4%	NA	NA	NA
Candidate Var.	68	68	68	68	68	68
Observations	42	42	42	42	42	42

** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Table A.15: Seed Dependency Issue: Full Model, Logs

A.8 Correlation Structure

Recent Immigrants	Correlation	No Qualifications	Correlation
Born in the UK	-0.88	Social Grade AB	-0.93
Born Rest of the World	0.86	Industry E[111]	0.89
Born in a 2000-EU State	0.84	Disabled	0.87
2 Bedrooms or Fewer	0.81	Industry J[112]	-0.86
Provides Unpaid Care	-0.81	Social Grade DE	0.83
		Social Grade C2	0.83
		Industry O[113]	-0.82

Table A.16: Candidate Variables with an Absolute Correlation to the Model Variables Larger than 0.8 (None for No Religion Stated)

A.9 Further Methods & Results Ignoring Model Selection

Binning

Chernozhukov et al. (2018b) suggest an additional method to find variables that describe the heterogeneity well. As for the measure of captured heterogeneity (see Table 1.4 in the main text), the observations are binned according to the predicted effect. Next, the mean-difference between the top and bottom bin is recorded for each candidate-variable. This is done in each

[111] Water supply, sewerage, waste management and remediation
[112] Information and communication
[113] Public administration and defense

of the 100 splits, and the measures and p-values are aggregated akin to section 1.4.4, with the addition of also conducting an FDR adjustment. Similar to section 1.6.3 but contrary to the sections 1.6.1 and 1.6.2, no choice between correlated (groups of) variables is made. The candidate variables with the largest difference (in percentage points, percent of the mean, and percent of the tercile span) that are also significant at the 10% level are listed in Table A.17. The complete list of significant variables is shown in Table A.18. Interestingly, after conducting the FDR adjustment, none of the candidate variables are significant in characterizing boroughs with the largest relative effect. As before, the tercile span is the difference between the mean of the top and the mean of the bottom tercile of the relevant variable (across all 42 boroughs). Overall, the variables seem to point to wealthy areas with a large fraction of immigrants.

Variable	Difference	Variable	Difference	Variable	Difference
Born in the UK	-28**	Recent Immigrants	126%**	Industry Code E	-86%**
Remain Vote	25**	Industry Code E[114]	-102%**	Prov. Unpaid Care	-84%**
$<$ 3 Bedrooms	20*	Buddhist	93%**	Mixed Ethnicity	84%**
Christian	-17*	Born Rest World	91%*	Industry Code J[115]	83%**
Married	15*	Mixed Ethnicity	80**	Remain Vote	83%**
Percentage Points		*Percent of the Mean*		*Percent of Tercile Span*	

Table A.17: Statistically Significant Candidate Variables (Absolute Effect): Top 5

Comparing these results to those in section 1.6.3, there are clearly similarities. This shows that the simple mechanism there is already rather informative. Also in line with the previous section, unemployment (as any other insignificant variable in the previous section) is not a significant characteristic of boroughs that were strongly affected. On the technical side, it is not surprising that the list of significant variables is lower in this section. The p-values are obtained from only half of the data and the rather conservative correction to double the p-value after taking the median drives them up further.

[114] Water supply, sewerage, waste management and remediation
[115] Information and communication

Variable	Difference in Perc. Pts.	% of Mean	% of Tercile Span
Born in the UK	-27.60**	-39.6%	-81.5%
Remain Vote	25.21**	44.3%	82.7%
2 Bedrooms or fewer	19.69*	38.8%	76.9%
Christian	-16.55*	-31.8%	-72.2%
Married	14.92*	35.2%	74.6%
Socal Grade AB	14.04*	52.1%	69.9%
No Qualifications	-8.93*	-46.1%	-78.6%
Social Grade C2	-8.89**	-53.7%	-81.8%
Industry Code M	7.36*	75.6%	80.0%
30 $>$ Age $>$ 15	6.64*	30.6%	69.9%
Age $>$ 64	-5.44**	-45.1%	-76.4%
Industry Code G	-5.05*	-35.9%	-74.4%
Industry Code J	4.65**	77.1%	83.4%
Disabled	-4.31*	-28.0%	-76.6%
Industry Code F	-3.86*	-56.6%	-73.4%
Recent Immigrants	3.02**	125.7%	82.0%
Providing Unpaid Care	-2.71**	-30.3%	-84.4%
Industry Code R,S,T,U,Other	2.55**	42.8%	79.2%
Mixed Ethnicity	2.36**	80.0%	83.7%
Buddhist	0.78**	93.1%	80.0%
Industry Code E	-0.47**	-101.9%	-86.1%
Born Rest World	0.01*	91.3%	74.6%

Absolute Effect

Table A.18: Complete List of Statistically Different Variables in Most vs Least Affected Terciles

Best Single Variable

Table A.19 is the equivalent of Table 1.6, but restricting each lasso to use exactly one variable. For the lasso, and hence CPSL, the choice of the candidate variable is mechanically the same as in Table 1.8 (Table 1.9 for the relative effect). The magnitude of the CPSL's coefficient is nevertheless interesting. The splitting based approach is not guaranteed to find the same variable (e.g. in case of extreme outliers), but in this case it does. The magnitudes found by CPSL and splitting are remarkably similar.

	(1)	(2)	(3)	(4)	(5)	(6)
	Hate Crimes per Pop. Million			log(Hate Crimes per Pop. Million)		
Recent Immigrants	2308***	2321***	390			
No Qualifications				-2.55**	-2.47***	-0.79
Method	CPSL	Splitting	Lasso	CPSL	Splitting	Lasso
Frequency	NA	46.2%	NA	NA	26.6%	NA
Candidate Var.	69	69	69	69	69	69
Observations	42	42	42	42	42	42

Note: Method-chosen models from 69 candidate variables. Cross sectional analysis across 42 boroughs in July 2016 (the month after Brexit). Dependent variable is detrended, deseasonalized, and demeaned at the borough level using 88 months of data. Mean share of people that have arrived in the UK within 2 years of the 2011 census: 0.024. Mean log(hate crimes per borough population million) in July 2016: 4.567. Mean share of people not stating a religion: 0.080. Mean share of people with no formal qualifications: 0.194. Mean hate crimes per borough population million in July 2016: 219. C.P.S.L assumes i.i.d. errors by construction. In case of splitting heteroskedasticity robust errors are used. Significance not defined for plain lasso (which serves as benchmark only).

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (for CPSL and Splitting)

Table A.19: Best Single Variable for the Absolute/Relative Increase in Hate Crime

Individually Interacted Candidate Variables

For completeness, the Tables A.20 to A.23 provide the full versions of the Tables 1.8 and 1.9 in the main text. In Table A.20 (and A.21), it becomes visible that the share of people with no qualifications is also strongly negatively associated with the absolute (and not just the relative) increase in hate crimes. Similarly - although not picked up by the selected model in Table 1.6 and only significant at the 10% level - the share of recent immigrants is also correlated with the relative (and not just the absolute) increase in hate crime (see Tables A.22 and A.23).

Candidate.Variable	Estimate	FDR Adjusted P	Unadjusted Min. 95	Unadjusted Max. 95
Recent Immigrants	99.10	0.00	69.20	129.10
Social Grade C2	-90.80	0.00	-123.00	-58.50
Mixed Ethnicity	90.00	0.00	55.00	124.90
Industry Code E	-89.60	0.00	-123.80	-55.40
Born UK	-89.20	0.00	-120.50	-58.00
Industry Code J	88.80	0.00	49.00	128.50
Remain Vote	86.70	0.00	48.10	125.40
Industry Code M	85.90	0.00	49.90	121.90
No Qualifications	-85.90	0.00	-117.50	-54.20
Industry Code G	-84.30	0.00	-122.80	-45.80
Provide Unpaid Care	-83.70	0.00	-116.40	-51.00
Industry Code Q	-82.20	0.00	-111.90	-52.50
< 3 Bedrooms	81.90	0.00	50.90	112.90
Industry Code F	-80.80	0.00	-114.30	-47.30
Born Rest World	79.40	0.00	44.50	114.40
Industry Code R,S,T,U,Other	79.10	0.00	37.90	120.40
Single	79.00	0.00	40.80	117.20
Christian	-78.60	0.01	-120.80	-36.40
Fully Deprived	78.00	0.00	39.80	116.20
Buddhist	77.90	0.00	41.40	114.40
Born Recent EU	77.10	0.00	40.90	113.30
Social Grade AB	76.60	0.01	34.90	118.20
Same Sex Marriage	76.40	0.01	34.70	118.20
Aged Over 64	-72.90	0.00	-103.80	-42.00
Aged 16-29	71.70	0.00	36.30	107.10
Disabled	-70.10	0.00	-99.30	-40.90
Industry Code C	-67.70	0.00	-95.50	-39.90
Industry Code D	-66.70	0.00	-94.20	-39.10
Industry Code O	-64.90	0.06	-111.10	-18.60
No Religion Stated	64.60	0.00	51.00	78.20
Self Employed	62.60	0.02	23.00	102.20
Ethnic Other Asian	60.10	0.06	16.80	103.30
White	-59.80	0.01	-92.50	-27.10
Work Part Time	-58.90	0.10	-105.00	-12.80
Divorced	-58.00	0.02	-93.10	-22.80
Aged 0-15	-54.50	0.03	-90.10	-19.00
Industry Code K	53.40	0.00	27.00	79.80
Work from Home	53.30	0.07	13.90	92.80
Industry Code L	50.40	0.00	26.80	73.90
Central Heating	-50.00	0.02	-81.20	-18.90
Industry Code A	-48.40	0.10	-86.40	-10.50
Social Grade DE	-47.90	0.26	-93.50	-2.40
Industry Code I	46.90	0.00	23.90	69.90
Arab	46.40	0.01	20.00	72.90
Muslim	43.30	0.02	16.30	70.40
Rent Social H.	42.00	0.33	-0.50	84.40
1 Pers. Household	41.80	0.26	2.10	81.50
No English	41.50	0.03	14.40	68.60
Industry Code B	39.60	0.18	4.80	74.40
Male	38.60	0.06	10.80	66.40
Black	35.90	0.30	0.60	71.30
Industry Code H	-34.00	1.00	-84.60	16.70
Born 2000-EU	33.50	0.20	3.40	63.50
Bad Health	-32.30	0.68	-72.20	7.60
Industry Code P	-23.00	0.86	-54.20	8.10
Social Grade C1	22.00	1.00	-24.70	68.70
Ethnic South Asian	19.60	1.00	-15.20	54.30
Aged 30-64	16.40	1.00	-26.40	59.30
Lone Parent	-14.10	1.00	-52.00	23.80
Econ. Active	13.70	1.00	-25.50	53.00
Other Religion	10.50	1.00	-23.60	44.50
Jewish	9.90	1.00	-10.70	30.50
Sikh	9.20	1.00	-9.00	27.40
Not Deprived	-8.00	1.00	-43.10	27.00
No Religion	7.00	1.00	-34.20	48.20
Industry Code N	-6.60	1.00	-61.40	48.30
Hindu	6.10	1.00	-22.10	34.40
Unemployed	1.30	1.00	-39.00	41.60

Table A.20: Absolute Effect Using Full Regression

Candidate.Variable	Estimate	FDR.Adjusted.P	Unadjusted.Min.95	Unadjusted.Max.95
Recent Immigrants	85.00	0.00	58.90	111.00
Social Grade C2	-77.80	0.00	-105.00	-50.70
Mixed Ethnicity	77.10	0.00	47.70	106.50
Industry Code E	-76.80	0.00	-106.40	-47.20
Born UK	-76.50	0.00	-102.60	-50.40
Industry Code J	76.10	0.00	42.10	110.00
Remain Vote	74.30	0.00	41.90	106.70
Industry Code M	73.70	0.00	43.00	104.30
No Qualifications	-73.60	0.00	-100.20	-47.00
Industry Code G	-72.30	0.00	-104.50	-40.00
Provide Unpaid Care	-71.70	0.00	-99.10	-44.40
Industry Code Q	-70.40	0.00	-95.10	-45.80
< 3 Bedrooms	70.20	0.00	44.10	96.40
Industry Code F	-69.20	0.00	-97.40	-41.10
Born Rest World	68.10	0.00	39.10	97.10
Industry Code R,S,T,U,Other	67.80	0.00	32.70	102.90
Single	67.70	0.00	35.40	100.10
Christian	-67.40	0.01	-104.00	-30.70
Fully Deprived	66.80	0.00	34.20	99.50
Buddhist	66.80	0.00	36.10	97.50
Born Recent EU	66.10	0.01	28.70	103.50
Social Grade AB	65.60	0.01	30.10	101.20
Same Sex Marriage	65.50	0.01	28.10	102.90
Aged Over 64	-62.50	0.00	-88.30	-36.70
Aged 16-29	61.50	0.00	31.50	91.50
Disabled	-60.10	0.00	-85.20	-35.00
Industry Code C	-58.00	0.00	-82.00	-34.10
Industry Code D	-57.10	0.00	-80.90	-33.30
Industry Code O	-55.60	0.07	-96.70	-14.50
No Religion Stated	55.30	0.00	43.00	67.70
Self Employed	53.60	0.02	19.90	87.40
Ethnic Other Asian	51.50	0.07	13.60	89.30
White	-51.30	0.01	-78.80	-23.70
Work Part Time	-50.50	0.12	-91.40	-9.50
Divorced	-49.70	0.02	-80.20	-19.20
Aged 0-15	-46.70	0.03	-77.00	-16.50
Industry Code K	45.80	0.01	20.50	71.10
Work from Home	45.70	0.08	11.50	79.90
Industry Code L	43.20	0.00	23.30	63.00
Central Heating	-42.90	0.02	-68.80	-17.00
Industry Code A	-41.50	0.11	-74.50	-8.60
Social Grade DE	-41.10	0.25	-79.60	-2.50
Industry Code I	40.20	0.00	20.80	59.60
Arab	39.80	0.26	2.10	77.50
Muslim	37.10	0.02	14.30	60.00
Rent Social H.	36.00	0.33	-0.20	72.20
1 Pers. Household	35.80	0.25	2.30	69.40
No English	35.60	0.03	12.50	58.60
Industry Code B	34.00	0.59	-6.00	74.00
Male	33.10	0.06	9.20	57.00
Black	30.80	0.28	1.10	60.50
Industry Code H	-29.10	1.00	-79.30	21.00
Born 2000-EU	28.70	0.21	3.00	54.40
Bad Health	-27.70	0.67	-61.80	6.40
Industry Code P	-19.70	0.93	-47.20	7.70
Social Grade C1	18.80	1.00	-21.00	58.70
Ethnic South Asian	16.80	1.00	-14.20	47.70
Aged 30-64	14.10	1.00	-24.00	52.20
Lone Parent	-12.10	1.00	-44.40	20.20
Econ. Active	11.80	1.00	-20.90	44.50
Other Religion	9.00	1.00	-80.60	98.50
Jewish	8.50	1.00	-14.40	31.30
Sikh	7.90	1.00	-9.60	25.50
Not Deprived	-6.90	1.00	-37.70	23.90
No Religion	6.00	1.00	-29.20	41.30
Industry Code N	-5.70	1.00	-55.50	44.20
Hindu	5.30	1.00	-29.40	39.90
Unemployed	1.10	1.00	-33.80	36.00

Table A.21: Absolute Effect Using Residuals

Candidate.Variable	Estimate	FDR.Adjusted.P	Unadjusted.Min.95	Unadjusted.Max.95
No Qualifications	-0.34	0.04	-0.51	-0.17
Industry Code E	-0.32	0.15	-0.54	-0.09
Industry Code J	0.31	0.08	0.13	0.49
Recent Immigrants	0.30	0.08	0.12	0.48
Remain Vote	0.30	0.11	0.10	0.49
Born UK	-0.29	0.14	-0.49	-0.09
Social Grade C2	-0.29	0.08	-0.46	-0.12
Social Grade AB	0.29	0.10	0.11	0.46
Mixed Ethnicity	0.28	0.15	0.07	0.49
Disabled	-0.27	0.19	-0.49	-0.06
Industry Code Q	-0.27	0.08	-0.42	-0.11
Industry Code G	-0.26	0.14	-0.45	-0.08
Industry Code M	0.26	0.11	0.09	0.43
Provide Unpaid Care	-0.26	0.15	-0.45	-0.07
Buddhist	0.25	0.15	0.07	0.42
Industry Code D	-0.24	0.08	-0.39	-0.10
Industry Code F	-0.24	0.11	-0.40	-0.08
Industry Code R,S,T,U,Other	0.23	0.17	0.06	0.41
Born Rest World	0.23	0.27	0.03	0.44
Aged Over 64	-0.23	0.19	-0.41	-0.05
Ethnic Other Asian	0.22	0.36	0.01	0.44
Divorced	-0.22	0.26	-0.41	-0.03
Self Employed	0.22	0.26	0.03	0.41
Christian	-0.22	0.49	-0.45	0.01
Industry Code C	-0.22	0.19	-0.39	-0.04
Born Recent EU	0.22	0.15	0.06	0.37
Single	0.20	0.19	0.04	0.36
White	-0.20	0.46	-0.41	0.00
Industry Code O	-0.20	0.32	-0.39	-0.02
Social Grade DE	-0.20	0.50	-0.41	0.01
Same Sex Marriage	0.20	0.17	0.05	0.35
< 3 Bedrooms	0.19	0.17	0.05	0.34
Bad Health	-0.19	0.55	-0.41	0.02
Aged 16-29	0.19	0.28	0.02	0.35
Work from Home	0.18	0.25	0.03	0.34
Work Part Time	-0.18	0.48	-0.36	0.01
Fully Deprived	0.18	0.42	0.00	0.35
Industry Code B	0.17	0.12	0.06	0.28
No Religion Stated	0.17	0.02	0.09	0.24
Aged 0-15	-0.15	0.30	-0.29	-0.02
Industry Code I	0.15	0.32	0.01	0.30
Industry Code L	0.15	0.32	0.01	0.28
Industry Code A	-0.14	1.00	-0.35	0.07
Male	0.14	0.56	-0.01	0.29
Industry Code K	0.13	0.32	0.01	0.25
Born 2000-EU	0.13	0.70	-0.03	0.29
Muslim	0.12	1.00	-0.05	0.29
Arab	0.12	0.64	-0.02	0.26
Econ. Active	0.12	1.00	-0.06	0.30
Lone Parent	-0.12	1.00	-0.28	0.05
Black	0.11	1.00	-0.07	0.30
No English	0.11	1.00	-0.05	0.27
Aged 30-64	0.10	1.00	-0.09	0.30
Industry Code H	-0.09	1.00	-0.27	0.10
Ethnic South Asian	0.08	1.00	-0.08	0.24
Central Heating	-0.08	1.00	-0.24	0.08
Rent Social H.	0.06	1.00	-0.11	0.23
1 Pers. Household	0.06	1.00	-0.10	0.21
Unemployed	-0.06	1.00	-0.25	0.14
Hindu	0.05	1.00	-0.08	0.19
Not Deprived	0.05	1.00	-0.14	0.24
Sikh	0.05	1.00	-0.03	0.12
Other Religion	0.04	1.00	-0.09	0.17
Industry Code P	-0.04	1.00	-0.17	0.09
Social Grade C1	0.03	1.00	-0.19	0.26
Industry Code N	0.03	1.00	-0.18	0.24
No Religion	-0.03	1.00	-0.20	0.15
Jewish	0.02	1.00	-0.06	0.10

Table A.22: Relative Effect Using Full Regression

Candidate.Variable	Estimate	FDR.Adjusted.P	Unadjusted.Min.95	Unadjusted.Max.95
No Qualifications	-0.29	0.03	-0.43	-0.15
Industry Code E	-0.27	0.16	-0.47	-0.07
Industry Code J	0.27	0.08	0.12	0.41
Recent Immigrants	0.25	0.08	0.10	0.41
Remain Vote	0.25	0.10	0.09	0.42
Born UK	-0.25	0.13	-0.42	-0.08
Social Grade C2	-0.25	0.08	-0.39	-0.10
Social Grade AB	0.24	0.08	0.10	0.39
Mixed Ethnicity	0.24	0.16	0.06	0.42
Disabled	-0.23	0.23	-0.42	-0.05
Industry Code Q	-0.23	0.08	-0.36	-0.10
Industry Code G	-0.23	0.13	-0.38	-0.07
Industry Code M	0.22	0.08	0.08	0.36
Provide Unpaid Care	-0.22	0.16	-0.38	-0.06
Buddhist	0.21	0.15	0.06	0.36
Industry Code D	-0.21	0.08	-0.33	-0.09
Industry Code F	-0.21	0.10	-0.34	-0.07
Industry Code R,S,T,U,Other	0.20	0.16	0.05	0.35
Born Rest World	0.20	0.25	0.03	0.37
Aged Over 64	-0.20	0.18	-0.35	-0.04
Ethnic Other Asian	0.19	0.37	0.01	0.37
Divorced	-0.19	0.25	-0.35	-0.03
Self Employed	0.19	0.25	0.03	0.35
Christian	-0.19	0.56	-0.39	0.02
Industry Code C	-0.19	0.23	-0.34	-0.03
Born Recent EU	0.18	0.23	0.03	0.34
Single	0.18	0.18	0.04	0.31
White	-0.17	0.45	-0.35	0.00
Industry Code O	-0.17	0.37	-0.34	-0.01
Social Grade DE	-0.17	0.50	-0.35	0.01
Same Sex Marriage	0.17	0.18	0.04	0.30
< 3 Bedrooms	0.17	0.16	0.04	0.29
Bad Health	-0.17	0.56	-0.35	0.02
Aged 16-29	0.16	0.25	0.02	0.30
Work from Home	0.16	0.23	0.03	0.29
Work Part Time	-0.15	0.53	-0.32	0.01
Fully Deprived	0.15	0.38	0.00	0.30
Industry Code B	0.14	0.15	0.04	0.25
No Religion Stated	0.14	0.03	0.08	0.21
Aged 0-15	-0.13	0.30	-0.25	-0.01
Industry Code I	0.13	0.37	0.01	0.26
Industry Code L	0.13	0.33	0.01	0.24
Industry Code A	-0.12	1.00	-0.32	0.07
Male	0.12	0.66	-0.02	0.25
Industry Code K	0.11	0.36	0.01	0.22
Born 2000-EU	0.11	0.71	-0.02	0.25
Muslim	0.10	1.00	-0.05	0.25
Arab	0.10	1.00	-0.10	0.31
Econ. Active	0.10	1.00	-0.05	0.26
Lone Parent	-0.10	1.00	-0.24	0.04
Black	0.10	1.00	-0.05	0.25
No English	0.10	1.00	-0.04	0.23
Aged 30-64	0.09	1.00	-0.09	0.26
Industry Code H	-0.08	1.00	-0.26	0.11
Ethnic South Asian	0.07	1.00	-0.07	0.21
Central Heating	-0.07	1.00	-0.20	0.07
Rent Social H.	0.05	1.00	-0.09	0.20
1 Pers. Household	0.05	1.00	-0.09	0.18
Unemployed	-0.05	1.00	-0.22	0.12
Hindu	0.05	1.00	-0.13	0.22
Not Deprived	0.04	1.00	-0.13	0.22
Sikh	0.04	1.00	-0.03	0.11
Other Religion	0.03	1.00	-0.30	0.37
Industry Code P	-0.03	1.00	-0.14	0.07
Social Grade C1	0.03	1.00	-0.17	0.22
Industry Code N	0.02	1.00	-0.17	0.22
No Religion	-0.02	1.00	-0.17	0.13
Jewish	0.02	1.00	-0.06	0.09

Table A.23: Relative Effect Using Residuals

B Supplementary Material and Proofs to Chapter 2

B.1 Proofs of Main Results

We recall here the definition of a viscosity solution. Consider a second order differential partial equation of the Dirichlet form

$$F\left(D^2u(y), Du(y), u(y), y\right) = 0, \quad u = 0 \text{ on } \partial\mathcal{Y}, \tag{B.1}$$

where the domain is $\mathcal{Y}$ which is an open set, y is a vector, Du, D^2u denote the first and second order derivatives with respect to y, and $\partial\mathcal{Y}$ is the boundary of $\mathcal{Y}$ on which the initial conditions are specified.[116] We restrict ourselves to functions $F(\cdot)$ that are *proper*, i.e. $F(\cdot)$ is non-decreasing in its third argument. Let $\mathcal{C}^2(\mathcal{Y})$ denote the space of all twice continuously differentiable functions on $\mathcal{Y}$.

Definition B1
A function u is said to be a viscosity solution to (B.1) on the domain $\mathcal{Y}$ if it satisfies the following:

(i) *$u = 0$ on $\partial\mathcal{Y}$, and*

(ii) *for each $\phi \in \mathcal{C}^2(\mathcal{Y})$, if $u - \phi$ has a local maximum at $y_0 \in \mathcal{Y}$, then*

$$F\left(D^2\phi(y_0), D\phi(y_0), u(y_0), y_0\right) \leq 0;$$

(iii) *similarly, for each $\phi \in \mathcal{C}^2(\mathcal{Y})$, if $u - \phi$ has a local minimum at $y_0 \in \mathcal{Y}$, then*

$$F\left(D^2\phi(y_0), D\phi(y_0), u(y_0), y_0\right) \geq 0.$$

We shall also say that a function u satisfies the inequality

$$F\left(D^2u(y), Du(y), u(y), y\right) \leq 0$$

in a viscosity sense if for each $\phi \in \mathcal{C}^2(\mathcal{Y})$, if $u - \phi$ has a local maximum at $y_0 \in \mathcal{Y}$, then

$$F\left(D^2\phi(y_0), D\phi(y_0), u(y_0), y_0\right) \leq 0.$$

Proof of Lemma 2

Recall that our partial differential equation is of the form

$$\lambda(t)\bar{G}_\theta(z,t)\partial_z h_\theta(z,t) + \partial_t h_\theta(z,t) - \beta h_\theta(z,t) + \lambda(t)\bar{r}_\theta(z,t) = 0, \quad h_\theta(0,t) = 0 \; \forall \; t. \tag{B.2}$$

[116] Note that the first and second arguments of $F(\cdot)$ are a matrix and a vector respectively.

We shall first get rid of the inhomogeneous term $-\beta h_\theta(z,t)$ as follows: Define $u_\theta(z,t) := e^{-\beta(t-t_0)}h_\theta(z,t)$. Then (B.2) implies

$$\lambda(t)\bar{G}_\theta(z,t)\partial_z u_\theta(z,t) + \partial_t u_\theta(z,t) + e^{-\beta(t-t_0)}\lambda(t)\bar{r}_\theta(z,t) = 0, \quad u_\theta(0,t) = 0\ \forall\ t.$$

Rewrite the above in the form

$$\partial_z u_\theta + H_\theta(t,z,\partial_t u_\theta) = 0, \quad u_\theta(0,t) = 0\ \forall\ t, \tag{B.3}$$

where $H(z,t,p)$ is the Hamiltonian defined by

$$H_\theta(t,z,p) = \frac{1}{\lambda(t)\bar{G}_\theta(z,t)}p + \frac{e^{-\beta(t-t_0)}\bar{r}_\theta(z,t)}{\bar{G}_\theta(z,t)}.$$

Because $e^{-\beta(t-t_0)}$ is strictly positive and twice continuously differentiable, it is straightforward to show that if a viscosity solution $u_\theta(z,t)$ exists for (B.3), then there also exists a viscosity solution $h_\theta(z,t) = e^{\beta(t-t_0)}u_\theta(z,t)$ to (B.2). Furthermore, the converse is also true. Hence it remains to show existence of a unique viscosity solution to (B.3).

Under Assumption 1, $H(.)$ is uniformly continuous in all its three arguments. Furthermore, for any $t_1, t_2 \in \mathbb{R}$, it is true that

$$|H_\theta(t_1,z,p) - H_\theta(t_2,z,p)| \le M(1+|p|)\,|t_1 - t_2|$$

for some constant M that is independent of θ, z, t_1, t_2, p. Now as long as the Hamiltonian satisfies the above properties, Souganidis (1985, Theorem 1) shows that a unique viscosity solution exists for (B.3).□

Proof of Theorem 2

As in the proof of Lemma 2, we also convert PDE (2.14) into a form that does not include the term $-\beta\hat{h}_\theta(z,t)$. We then have the following PDEs:

$$\partial_z u_\theta + H_\theta(t,z,\partial_t u_\theta) = 0, \quad u_\theta(0,t) = 0\ \forall\ t \tag{B.4}$$
$$\partial_z \hat{u}_\theta + \hat{H}_\theta(t,z,\partial_t \hat{u}_\theta) = 0, \quad \hat{u}_\theta(0,t) = 0\ \forall\ t \tag{B.5}$$

where

$$H_\theta(t,z,p) = \frac{1}{\lambda(t)\bar{G}_\theta(z,t)}p + \frac{e^{-\beta(t-t_0)}\bar{r}_\theta(z,t)}{\bar{G}_\theta(z,t)},$$
$$\hat{H}_\theta(t,z,p) = \frac{1}{\hat{\lambda}(t)\hat{G}_\theta(z,t)}p + \frac{e^{-\beta(t-t_0)}\hat{r}_\theta(z,t)}{\hat{G}_\theta(z,t)}.$$

Note that as in Lemma 2, if $\hat{u}_\theta(z,t)$ is the unique solution of (B.5), then $\hat{h}_\theta(z,t) = e^{\beta(t-t_0)}\hat{u}_\theta(z,t)$ is the unique solution for (2.14), and vice versa. Therefore we first obtain bounds on the sup norm of $\hat{u}_\theta(z,t) - u_\theta(z,t)$.

We start with some properties of the solution $u_\theta(z,t)$. We show that there exist $L_1, L_2 < \infty$ such that

$$\sup_\theta \|u_\theta\| < L_1, \text{ and} \tag{B.6}$$

$$\sup_\theta \|\partial_t u_\theta\| < L_2 \tag{B.7}$$

where $\|u_\theta\|$ denotes the sup-norm of $u_\theta(z,t)$ over its domain, and

$$\|\partial_t u_\theta\| := \sup_{t_1,t_2\in[t_0,\infty),z\in[0,z_0]} \frac{|u_\theta(z,t_1) - u_\theta(z,t_2)|}{|t_1 - t_2|}$$

denotes the Lipschitz constant for $u_\theta(z,t)$ in terms of t (if $u_\theta(z,t)$ is not Lipschitz continuous, we set $\|\partial_t u_\theta\|$ to ∞). To prove the above, we first observe that $H_\theta(t,z,0)$ is uniformly bounded for all θ, z, t. Then (B.6) follows by Crandall and Lions (1987, Theorem VII.1, Remark i)[117]. Furthermore, by Assumptions 1-6, there exists some constant B independent of θ, z, t_1, t_2, p such that

$$H_\theta(t_1, z, \lambda(t_2 - t_1)) - H(t_2, z, \lambda(t_2 - t_1)) \le B\lambda|t_1 - t_2|^2 + B|t_1 - t_2|,$$

for any positive λ. Subsequently, by Crandall and Lions (1987, Theorem VII.1, Remark ii), $u_\theta(z,t)$ is uniformly Lipschitz continuous in t (i.e. uniformly over all θ, z). This proves (B.7).

Now define

$$A = \{z, t, p : z \in [0, z_0], t \in [t_0, \infty), |p| \le L_2\}.$$

By Souganidis (1985, Proposition 1.4),[118] we have that for every $z \in [0, z_0]$ and $\theta \in \Theta$,

$$\sup_t |u_\theta(z,t) - \hat{u}_\theta(z,t)| \le z \sup_{(s,t,p)\in A} \left| H_\theta(t,s,p) - \hat{H}_\theta(t,s,p) \right|. \tag{B.8}$$

Hence it suffices to bound the right hand side of (B.8) uniformly over θ. To do this, observe that by the results of Athey and Wager (2018), there exists $C_0 < \infty$ such that

$$E\left[\sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta} \|\hat{r}_\theta(z,t) - \bar{r}_\theta(z,t)\|\right] \le C_0\sqrt{\frac{v_1}{n}}, \text{ and} \tag{B.9}$$

$$E\left[\sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta} \left\|\hat{G}_\theta(z,t) - \bar{G}_\theta(z,t)\right\|\right] \le C_0\sqrt{\frac{v_2}{n}}.$$

The last inequality can also be derived from Kitagawa and Tetenov (2018, Lemma A.4). Using (B.9) together with Assumptions 1-6, straightforward algebra enables us to show that

$$\sup_{\theta\in\Theta} \sup_{(z,t,p)\in A} \left| H_\theta(t,z,p) - \hat{H}_\theta(t,z,p) \right| \le C_3\sqrt{\frac{v}{n}}$$

[117] A careful examination of the proofs of Crandall and Lions (1987) shows that their results hold uniformly in θ as long as the underlying properties on the Hamiltonian also hold uniformly.

[118] We take the special case of the Proposition in Souganidis (1985) where $\varepsilon \to 0$.

with probability approaching one under $F \times \Omega_\lambda(\mathbb{T})$ (henceforth wpa1), where C_3 is some constant independent of θ, t, z, p. Substituting the above in (B.8), we obtain wpa1

$$\sup_{z\in[0,z_0],t\in[t_0,\infty),\theta\in\Theta} |u_\theta(z,t) - \hat{u}_\theta(z,t)| \le z_0 C_3 \sqrt{\frac{v}{n}}. \tag{B.10}$$

In view of (B.10) and the definitions of $u_\theta(z,t), \hat{u}_\theta(z,t)$, we have wpa1,

$$\left| e^{-\beta(t-t_0)} \left\{ h_\theta(z,t) - \hat{h}_\theta(z,t) \right\} \right| \le z_0 C_3 \sqrt{\frac{v}{n}}.$$

This concludes the proof of the theorem.□

Proof of Theorem 3

The argument leading to the proof of Theorem 3 here was sketched by Souganidis (2009) in an unpublished paper. We formalize the argument here.

All the inequalities in this section should be understood to be holding with probability approaching 1 under the joint distribution of $F \times \Omega_{\lambda(t)}$. In what follows, we drop this qualification for ease of notation and hold this to be implicit. We also employ the following notation: For any function f over (z,t), Df denotes its Jacobean. Additionally, $\|\partial_z f\|, \|\partial_z f\|$ and $\|Df\|$ denote the Lipschitz constants for $f(\cdot,t), f(z,\cdot)$ and $f(\cdot,\cdot)$. We shall represent PDE (2.14) by

$$\hat{F}_\theta(\partial_t f, \partial_z f, f, z, t) = 0, \quad f(0,t) = 0 \;\forall\; t \tag{B.11}$$

with f denoting a function, and where

$$\hat{F}_\theta(l,p,q,z,t) := l + \frac{1}{\hat{\lambda}(t)\hat{G}_\theta(z,t)} p - \frac{\beta}{\hat{\lambda}(t)\hat{G}_\theta(z,t)} q + \frac{\hat{r}_\theta(z,t)}{\hat{G}_\theta(z,t)}.$$

Additionally, denote our approximation scheme (2.16) by

$$S_\theta([f], f, z, t) = 0, \quad f(0,t) = 0 \;\forall\; t \tag{B.12}$$

where for any two functions f_1, f_2,

$$S_\theta\left([f_1], f_2(z,t), z, t, b_n\right) := \frac{b_n}{|\hat{G}_\theta(z,t)|} \left(f_2(z,t) - E_{n,\theta}\left[e^{-\beta(t'-t)/b_n} f_1(z',t') | z, t \right] \right) - \frac{\hat{r}_\theta(z,t)}{|\hat{G}_\theta(z,t)|}. \tag{B.13}$$

Here $[f]$ refers to the fact that it is a functional argument. Note that $\hat{h}_\theta$ and $\tilde{h}_\theta$ are the functional solutions to (B.11) and (B.12) respectively. Also note that $\hat{F}_\theta(l,p,q_1,z,t) \ge \hat{F}_\theta(l,p,q_2,z,t)$ for all $q_1 \ge q_2$ by our assumption that $\bar{G}_\theta(z,t)$ is strictly negative. We shall also make use of the following two properties for $S_\theta(\cdot)$:

(i) $S_\theta(\cdot)$ is monotone in its first argument i.e.

$$S_\theta([f_1], f, z, t, b_n) \ge S_\theta([f_2], f, z, t, b_n) \;\forall\; f_2 \ge f_1.$$

(ii) For all f and $m \in \mathbb{R}$,

$$S_\theta([f+m], f+m, z, t, b_n) \geq S_\theta([f], f, z, t) + \chi m,$$

where $\chi = c\beta \inf_{\theta,z,t} |\hat{\lambda}(t)\hat{G}_\theta(z,t)|^{-1}$ for some $c = (0,1)$. Note that $\chi > 0$ by our assumptions.

The first property is trivial to show. As for the second, observe that

$$S_\theta([f+m], f+m, z, t, b_n) - S_\theta([f], f, z, t) = m\frac{b_n}{|\hat{G}_\theta(z,t)|}\left(1 - E_{n,\theta}\left[e^{-\beta(t'-t)/b_n}|t\right]\right) \geq \chi m.$$

We will also make use of the following properties of the solution $\hat{h}_\theta(z,t)$: There exist K_1, K_2 satisfying

$$\sup_\theta \left\|\hat{h}_\theta\right\| < K_1, \text{ and} \tag{B.14}$$

$$\sup_\theta \left\|D\hat{h}_\theta\right\| < K_2. \tag{B.15}$$

Equation (B.14) follows by similar arguments as in the proof of Theorem 2. For (B.15), note that also by similar arguments as in the proof of Theorem 2, $\sup_\theta \left\|\partial_t \hat{h}_\theta\right\| < L_2 < \infty$. Set

$$A = \{z, t, q, p : z \in [0, z_0], t \in [t_0, \infty), |q| \leq K_1, |p| \leq L_2\}.$$

Souganidis (1985, Proposition 1.5(e)) shows that

$$\sup_t \left\|\hat{h}_\theta(z_1, t) - \hat{h}_\theta(z_2, t)\right\| \leq |z_1 - z_2| \sup_{(s,t,q,p)\in A} |H_\theta(t, s, q, p)|, \qquad \text{where} \tag{B.16}$$

$$H_\theta(t, s, q, p) := \frac{1}{\hat{\lambda}(t)\hat{G}_\theta(z,t)}p - \frac{\beta}{\hat{\lambda}(t)\hat{G}_\theta(z,t)}q + \frac{\hat{r}_\theta(z,t)}{\hat{G}_\theta(z,t)}.$$

In view of (B.16) and Assumptions 1-6, straightforward algebra shows that $\sup_\theta \left\|\partial_z \hat{h}_\theta\right\| \leq L_3 < \infty$. Hence $\sup_\theta \left\|D\hat{h}_\theta\right\| \leq K_2 := L_2 + L_3 < \infty$. This proves (B.15).

We provide here an upper bound for

$$m_\theta := \sup_{z\in[0,z_0], t\in[t_0,\infty)} \left(\hat{h}_\theta(z,t) - \tilde{h}_\theta(z,t)\right). \tag{B.17}$$

A lower bound for $\hat{h}_\theta - \tilde{h}_\theta$ can be obtained in an analogous manner. Clearly, we may assume $m_\theta > 0$, as otherwise we are done. Denote (z^*_θ, t^*_θ) as the point at which the supremum is attained in (B.17). Such a point exists since $\hat{h}_\theta$ and $\tilde{h}_\theta$ are both continuous. We shall consider separately the two cases: (i) $z^*_\theta \leq 2K_2\epsilon$, and (ii) $z^*_\theta > 2K_2\epsilon$.

We start with case (i). In view of (B.15), and the fact $\hat{h}_\theta(0,t) = 0\ \forall t$, we have $|\hat{h}_\theta(z^*_\theta, t^*_\theta)| \leq 4K_2^2\epsilon$. Furthermore, using Assumption 3, which implies $\|\hat{r}_\theta\| \leq 2M < \infty$, we can show using the properties of contraction mappings that $|\tilde{h}_\theta(z,t)| \leq 2Mz$ for all z, t. Hence, $|\tilde{h}_\theta(z^*_\theta, t^*_\theta)| \leq$

$4MK_2\epsilon$. Combining the above gives

$$m_\theta \leq 4(K_2^2 + MK_2)\epsilon. \tag{B.18}$$

This completes the case where $z_\theta^* \leq 2K_2\epsilon$.

We now turn to Case (ii), i.e. $z_\theta^* > 2K_2\epsilon$. To obtain the bound on m_θ in this case, we shall employ the sup-convolution of $\hat{h}_\theta(z,t)$, denoted by $\hat{h}_\theta^\epsilon(z,t)$:

$$\hat{h}_\theta^\epsilon(z,t) := \sup_{r,w}\left\{\hat{h}_\theta(r,w) - \frac{1}{\epsilon}\left(|z-r|^2 + |t-w|^2\right)\right\}.$$

We discuss sup and inf-convolutions and their properties in Appendix B.3. In view of (B.15) and Lemma B2 (below),

$$\left\|\hat{h}_\theta - \hat{h}_\theta^\epsilon\right\| \leq 4K_2^2\epsilon, \tag{B.19}$$

where the norm in the above expression should be understood as the supremum over all $t \in [t_0, \infty)$ and $z > 2K_2\epsilon$. Also, by Lemma B4 (below), there exists $c < \infty$ independent of θ, z, t such that for all $z > 2K_2\epsilon$,

$$\hat{F}_\theta(\partial_t \hat{h}_\theta^\epsilon, \partial_z \hat{h}_\theta^\epsilon, \hat{h}_\theta^\epsilon, z, t) \leq c\epsilon, \tag{B.20}$$

where the above expression is to be interpreted in the viscosity sense. Finally, we also note from Lemma B2 that $\hat{h}_\theta^\epsilon$ is semi-convex with coefficient $1/\epsilon$.

We now compare $S_\theta(\cdot)$ and $\hat{F}_\theta(\cdot)$ at the function $\hat{h}_\theta^\epsilon$. Consider any (z,t) with $z \neq 0$ at which $\hat{h}_\theta^\epsilon$ is differentiable (note that because of semi-convexity, it is differentiable almost everywhere). We can then expand

$$\begin{aligned}
S_\theta([\hat{h}_\theta^\epsilon], \hat{h}_\theta^\epsilon, z, t, b_n) &= \frac{b_n \hat{h}_\theta^\epsilon(z,t)}{|\hat{G}_\theta(z,t)|}\left(1 - E_{n,\theta}\left[e^{-\beta(t'-t)/b_n}|z,t\right]\right) \\
&\quad + \frac{1}{|\hat{G}_\theta(z,t)|}E_{n,\theta}\left[e^{-\beta(t'-t)/b_n}\left\{\hat{h}_\theta^\epsilon(z,t) - \hat{h}_\theta^\epsilon(z',t')\right\}|z,t\right] - \frac{\hat{r}_\theta(z,t)}{|\hat{G}_\theta(z,t)|} \\
&:= A_\theta^{(1)}(z,t) + A_\theta^{(2)}(z,t) + \frac{\hat{r}_\theta(z,t)}{\hat{G}_\theta(z,t)}.
\end{aligned} \tag{B.21}$$

Using the fact $\|h_\theta^\epsilon\| \leq \left\|\hat{h}_\theta\right\| \leq L_1$, straightforward algebra enables us to show using Assumptions 1-6 that

$$A_\theta^{(1)}(z,t) \leq -\frac{\beta}{\hat{\lambda}(t)\hat{G}_\theta(z,t)|}\hat{h}_\theta^\epsilon(z,t) + \frac{C_1}{b_n}, \tag{B.22}$$

for some C_1 independent of θ, z, t. We now consider $A_\theta^{(2)}(z,t)$. Observe that by semi-convexity of $\hat{h}_\theta^\epsilon$ (Lemma B1 below),

$$\hat{h}_\theta^\epsilon(z',t') \geq \hat{h}_\theta^\epsilon(z,t) + \partial_z \hat{h}_\theta^\epsilon(z,t)(z-z') + \partial_t \hat{h}_\theta^\epsilon(z,t)(t-t') - \frac{1}{2\epsilon}\left\{|z-z'|^2 + |t-t'|^2\right\}.$$

Substituting the above into the expression for $A_\theta^{(2)}(z,t)$, and using Assumptions 1-6, straightforward algebra enables us to show

$$A^{(2)}_{\theta}(z,t) \leq \partial_z \hat{h}^{\epsilon}_{\theta} + \frac{1}{\hat{\lambda}(t)\hat{G}_{\theta}(z,t)} \partial_t \hat{h}^{\epsilon}_{\theta} + \frac{C_2}{\epsilon b_n}, \tag{B.23}$$

where again C_2 is independent of θ, z, t. Combining (B.21)-(B.23), and setting $C = \max(C_1, C_2)$, we thus find

$$S_{\theta}([\hat{h}^{\epsilon}_{\theta}], \hat{h}^{\epsilon}_{\theta}, z, t, b_n) \leq \hat{F}_{\theta}(\partial_t \hat{h}^{\epsilon}_{\theta}, \partial_z \hat{h}^{\epsilon}_{\theta}, \hat{h}^{\epsilon}_{\theta}, z, t) + \frac{C}{b_n}\left(1 + \frac{1}{\epsilon}\right). \tag{B.24}$$

In view of (B.24) and (B.20), we obtain

$$S_{\theta}([\hat{h}^{\epsilon}_{\theta}], \hat{h}^{\epsilon}_{\theta}, z, t, b_n) \leq c\epsilon + \frac{C}{b_n}\left(1 + \frac{1}{\epsilon}\right) \quad \text{a.e.} \tag{B.25}$$

where the qualification almost everywhere (a.e.) refers to the points where $D\hat{h}^{\epsilon}_{\theta}$ exists.

Let (here $f^+ := \max(f, 0)$)

$$m^{\epsilon}_{\theta} := \sup_{z \in [0, z_0], t \in [t_0, \infty)} \left(\hat{h}^{\epsilon}_{\theta}(z,t) - \tilde{h}_{\theta}(z,t)\right)^+,$$

and denote $(\breve{z}_{\theta}, \breve{t}_{\theta})$ as the point at which the supremum is attained (or where the right hand side of the above expression is arbitrarily close to m^{ϵ}_{θ} in case $\breve{z}_{\theta} = \infty$). Note that in view of $z^*_{\theta} > 2K_2\epsilon$, equation (B.19) implies $\breve{z}_{\theta} > 0$. Now,

$$\hat{h}^{\epsilon}_{\theta} \leq \tilde{h}_{\theta} + m^{\epsilon}_{\theta}.$$

Then in view of properties (i), (ii) of $S(\cdot)$ derived above,

$$\begin{aligned}
\chi m^{\epsilon}_{\theta} &= S_{\theta}\left([\tilde{h}_{\theta}], \tilde{h}_{\theta}(\breve{z}_{\theta}, \breve{t}_{\theta}), \breve{z}_{\theta}, \breve{t}_{\theta}, b_n\right) + \chi m^{\epsilon}_{\theta} \\
&\leq S_{\theta}\left([\tilde{h}_{\theta} + m^{\epsilon}_{\theta}], \tilde{h}_{\theta}(\breve{z}_{\theta}, \breve{t}_{\theta}) + m^{\epsilon}_{\theta}, \breve{z}_{\theta}, \breve{t}_{\theta}, b_n\right) \\
&\leq S_{\theta}\left([\hat{h}^{\epsilon}_{\theta}], \hat{h}^{\epsilon}_{\theta}(\breve{z}_{\theta}, \breve{t}_{\theta}), \breve{z}_{\theta}, \breve{t}_{\theta}, b_n\right).
\end{aligned} \tag{B.26}$$

Without loss of generality, we may assume $\hat{h}^{\epsilon}_{\theta}$ is differentiable at $(\breve{z}_{\theta}, \breve{t}_{\theta})$ as otherwise we can choose a point arbitrarily close, given that $\hat{h}^{\epsilon}_{\theta}$ is Lipschitz continuous and differentiable a.e. Now we can combine (B.26) and (B.25) to obtain

$$m^{\epsilon}_{\theta} \leq c_1\epsilon + \frac{C_1}{b_n}\left(1 + \frac{1}{\epsilon}\right), \tag{B.27}$$

where $c_1 = \chi^{-1}c$ and $C_1 = \chi^{-1}C$ are independent of θ, z, t. Hence, in view of (B.19) and (B.27),

$$m_{\theta} \leq (4K_2^2 + c_1)\epsilon + \frac{C_1}{b_n}\left(1 + \frac{1}{\epsilon}\right). \tag{B.28}$$

This completes the derivation of the upper bound for m_{θ} under case (ii), i.e. $z^*_{\theta} > 2K_2\epsilon$.

Finally, in view of (B.18) and (B.28), setting $\epsilon = b_n^{-1/2}$ gives the desired rate.□

B.2 Pseudo-Codes and Additional Details for the Algorithm

This Section consists of two parts. In the first part, we give details about the convergence of the Actor-Critic algorithm in Section 2.4.2, and provide a proof of Theorem 1 in the main text. In the second part, we provide psuedo-codes and some additional discussion for various extensions to the basic algorithm that were proposed in Section 2.6.

Convergence of the Actor-Critic Algorithm

Let $\bar{h}_\theta := \bar{\nu}_\theta^\intercal \phi_{z,t}$, where $\bar{\nu}_\theta$ denotes the fixed point of the value function updates (2.21) for any given value of θ. This is the 'Temporal-Difference fixed point', and is known to exist and also to be unique (Tsitsiklis and van Roy, 1997). We will also make use of the quantities

$$\bar{h}_\theta^+(z,t) \equiv E_\theta[\hat{r}_n(x,a) + \mathbb{I}_{\{z'>0\}} e^{-\beta(t'-t)} \bar{h}_\theta(z',t')|z,t]$$

and

$$\mathcal{E}_\theta = E_\theta\left[e^{-\beta(t-t_0)}\left\{\nabla_\theta \bar{h}_\theta^+(z,t) - \nabla_\theta \bar{h}_\theta(z,t)\right\}\right].$$

Define $\mathcal{Z}$ as the set of local maxima of $J(\theta) \equiv \hat{h}_\theta(z_0,t_0)$, and $\mathcal{Z}^\epsilon$ an ϵ-expansion of that set. Also, $\theta^{(k)}$ denotes the k-th update of θ. The following theorem is a straightforward consequence of the results of Bhatnagar et al. (2009).

Theorem B1 (Bhatnagar et al., 2009)
Suppose that Assumption 2 holds. Then, given $\epsilon > 0$, there exists δ such that, if $\sup_k |\mathcal{E}_{\theta^{(k)}}| < \delta$, it holds that $\theta^{(k)} \to \mathcal{Z}^\epsilon$ with probability 1 as $k \to \infty$.

Intuition for the above theorem can be gleaned from the fact that the expected values of updates for the policy parameter are approximately given by

$$E\left[e^{-\beta(t-t_0)}\delta_n(s,s',a)\nabla_\theta \ln \pi(a|s;\theta)\right] \approx \nabla_\theta J(\theta) + \mathcal{E}_\theta.$$

Thus the term $\mathcal{E}_\theta$ acts as bias in the gradient updates. One can show from the properties of the temporal difference fixed point that if $d_\nu = \infty$, then $\bar{h}_\theta(z,t) = \bar{h}_\theta^+(z,t) = \hat{h}_\theta(z,t)$, see e.g. Tsitsiklis and van Roy (1997). Hence, in this case $\mathcal{E}_\theta = 0$. More generally, it is known that

$$\bar{h}_\theta(z,t) = P_\phi[\bar{h}_\theta^+(z,t)],$$

where P_ϕ is the projection operator onto the vector space of functions spanned by $\{\phi^{(j)} : j = 1,\dots,d_\nu\}$. This implies that $\nabla_\theta \bar{h}_\theta^+(z,t) - \nabla_\theta \bar{h}_\theta(z,t) = (I - P_\phi)[\nabla_\theta \bar{h}_\theta](z,t)$[119]. Now, $\nabla_\theta \bar{h}_\theta$ is uniformly (where the uniformity is with respect to θ) Hölder continuous as long as $\nabla_\theta \pi_\theta(s)$ is also uniformly Hölder continuous in s.[120] Hence for a large class of sieve approximations (e.g. Trigonometric series), one can show that $\sup_\theta \left\|(I - P_\phi)[\nabla_\theta \bar{h}_\theta]\right\| \le A(d_\nu)$ where $A(.)$ is

[119]To verify this, note that we can associate $\bar{h}_\theta, \bar{h}_\theta^+$ with vectors and P_ϕ with a matrix since we assumed discrete values for z,t.

[120]This can be shown easily from the definition of the temporal difference fixed point.

some function satisfying $A(x) \to 0$ as $x \to \infty$. This implies $\sup_\theta |\mathcal{E}_\theta| \leq A(d_\nu)$. The exact form of $A(.)$ depends on the smoothness of $\nabla_\theta \bar{h}_\theta$, and therefore that of $\nabla_\theta \pi_\theta(s)$, with greater smoothness leading to faster decay of $A(.)$. In view of the above discussion, we have thus shown the following:

Corollary B1
Suppose that Assumption 2 holds and additionally that $\nabla_\theta \pi_\theta(s)$ is uniformly Hölder continuous in s. Then, for each $\epsilon > 0$, there exists M such that if $d_\nu \geq M$, then $\theta^{(k)} \to \mathcal{Z}^\epsilon$ with probability 1 as $k \to \infty$.

The above was stated as Theorem 1 in the main text.

Extensions and Pseudo-Codes

Algorithm 3 and 4 provide the pseudo-codes for the algorithm with non-compliance and clusters respectively.

Algorithm 3: Parallel Actor-Critic with Non-Compliance (part I/II)

Initialise policy parameter weights $\theta \leftarrow 0$
Initialise value function weights $\nu \leftarrow 0$
Batch size B

For $p = 1, 2, ...$ processes, launched in parallel, each using and updating the same global parameters θ and ν:

Repeat forever:

Reset budget: $z \leftarrow z_0$
Reset time: $t \leftarrow t_0$
$I \leftarrow 1$

While $z > 0$**:**

$\theta_p \leftarrow \theta$ (Create local copy of θ for process p)

$\nu_p \leftarrow \nu$ (Create local copy of ν for process p)

batch_policy_upates $\leftarrow 0$

batch_value_upates $\leftarrow 0$

Algorithm 3: Parallel Actor-Critic with Non-Compliance (part II/II)

For $b = 1, 2, ..., B$**:**

$x \sim F_n$ (Draw new covariate at random from data)

hetero $\sim$ multinomial($\hat{q_c}(x), \hat{q_a}(x), \hat{q_n}(x)$) (Draw compliance heterogeneity)

$a \sim \pi(a|s;\theta_p)$ (Draw action, note: $s = (x, z, t)$)

If hetero $= 1$ (Sample draw is a complier)

$R \leftarrow \widehat{LATE}(x) \cdot \mathbb{I}(a = 1)$ (I.e. $\hat{r}(x, a)$)

$z' \leftarrow z + G_a(x, z, t)/b_n$

Elseif hetero $= 2$ (Sample draw is always-taker)

$R \leftarrow 0$

$z' \leftarrow z + G_1(x, z, t)/b_n$

Elseif hetero $= 3$ (Sample draw is never-taker)

$R \leftarrow 0$

$z' \leftarrow z + G_0(x, z, t)/b_n$

$\Delta t \sim$ Exponential($\hat{\lambda}(t)$) (Draw time increment)

$t' \leftarrow t + \Delta t/b_n$

$\delta \leftarrow R + \mathbb{I}\{z' > 0\}e^{-\beta(t'-t)}\nu_p^{\intercal}\phi_{z',t'} - \nu_p^{\intercal}\phi_{z,t}$ (TD error)

batch_policy_upates $\leftarrow$ batch_policy_upates $+ \alpha_\theta I \delta \nabla_\theta \ln \pi(a|s;\theta_p)$

batch_value_upates $\leftarrow$ batch_value_upates $+ \alpha_\nu I \delta \phi_{z,t}$

$z \leftarrow z'$

$t \leftarrow t'$

$I \leftarrow e^{-\beta(t'-t)} I$

If $z \leq 0$, break **For**

Globally update: $\nu \leftarrow \nu +$ batch_value_upates$/B$

Globally update: $\theta \leftarrow \theta +$ batch_policy_upates$/B$

Algorithm 4: Parallel Actor-Critic with Clusters

Initialise policy parameter weights $\theta \leftarrow 0$

Initialise value function weights $\nu \leftarrow 0$

Batch size B

Clusters $c = 1, 2, \dots, C$

Cluster specific arrival rates $\hat{\lambda}_c(t)$

For $p = 1, 2, \dots$ processes, launched in parallel, each using and updating the same global parameters θ and ν:

Repeat forever:

Reset budget: $z \leftarrow z_0$

Reset time: $t \leftarrow t_0$

$I \leftarrow 1$

While $z > 0$**:**

$\theta_p \leftarrow \theta$ (Create local copy of θ for process p)

$\nu_p \leftarrow \nu$ (Create local copy of ν for process p)

batch_policy_upates $\leftarrow 0$

batch_value_upates $\leftarrow 0$

For $b = 1, 2, ..., B$**:**

$\hat{\lambda}(t) \leftarrow \sum_c \hat{\lambda}_c(t)$ (Calculate arrival rate for next individual)

$\Delta t \sim \text{Exponential}(\hat{\lambda}(t))$ (Sample time increment until next arrival)

$t' \leftarrow t + \Delta t / b_n$

$z' \leftarrow z + G_a(x, z, t)/b_n$

$c \sim \text{multinomial}(p_1, \dots, p_C)$ (where $p_c := \hat{\lambda}_c(t)/\hat{\lambda}(t)$)

$x \sim F_{n,c}$ (Draw new covariate at random from data cluster c)

$a \sim \pi(a|s; \theta_p)$ (Draw action, note: $s = (x, z, t)$)

$R \leftarrow \hat{r}(x, a)$ (with $R = 0$ if $a = 0$)

$\delta \leftarrow R + \mathbb{I}\{z' > 0\} e^{-\beta(t'-t)} \nu_p^\intercal \phi_{z',t'} - \nu_p^\intercal \phi_{z,t}$ (TD error)

batch_policy_upates $\leftarrow$ batch_policy_upates $+ \alpha_\theta I \delta \nabla_\theta \ln \pi(a|s; \theta_p)$

batch_value_upates $\leftarrow$ batch_value_upates $+ \alpha_\nu I \delta \phi_{z,t}$

$z \leftarrow z'$

$t \leftarrow t'$

$I \leftarrow e^{-\beta(t'-t)} I$

If $z \leq 0$, break **For**

Globally update: $\nu \leftarrow \nu +$ batch_value_upates$/B$

Globally update: $\theta \leftarrow \theta +$ batch_policy_upates$/B$

Additionally, Algorithm 5 provides the pseudo-code for an off-policy actor critic algorithm that is useful for deterministic policy rules. For simplicity we provide the last algorithm without parallel updates, though this could be easily extended to here as well.

Algorithm 5: Off Policy Actor-Critic following Degris et al. (2012)

Initialise policy parameter weights $\theta \leftarrow 0$
Initialise value function weights $\nu \leftarrow 0$
Initialise correction weights for value function update $w \leftarrow 0$

Repeat forever:

Reset budget: $z \leftarrow z_0$
Reset time: $t \leftarrow t_0$
$I \leftarrow 1$

While $z > 0$**:**

$x \sim F_n$ (Draw new covariate at random from data)

$a \sim B(a|s)$ (sample action from behavioural policy)

$R \leftarrow \hat{r}(x, a)$ (with $R = 0$ if $a = 0$)

$\Delta t \sim \text{Exponential}(\hat{\lambda}(t))$ (Draw time increment)

$t' \leftarrow t + \Delta t / b_n$

$z' \leftarrow z + G_a(x, z, t)/b_n$

$\delta \leftarrow R + \mathbb{I}\{z' > 0\} e^{-\beta(t'-t)} \nu^\intercal \phi_{z',t'} - \nu^\intercal \phi_{z,t}$ (Temporal-Difference error)

$\rho \leftarrow \frac{\pi(a|s;\theta)}{b(a|s)}$ (Importance sampling)

$\nu \leftarrow \nu + \alpha_\nu I \left(\delta \rho \phi_{z,t} - \rho e^{-\beta(t'-t)} (w^\intercal \phi_{z,t}) \phi_{z,t} \right)$ (Update value parameter)

$w \leftarrow w + \alpha_w I \left(\delta \rho \phi_{z,t} - (w^\intercal \phi_{z,t}) \phi_{z,t} \right)$ (Correction term for off policy)

$\theta \leftarrow \theta + \alpha_\theta I \delta \rho \nabla_\theta \ln \pi(a|s;\theta)$ (Update policy parameter)

$z \leftarrow z'$

$t \leftarrow t'$

$I \leftarrow e^{-\beta(t'-t)} I$

B.3 Semi-Convexity, Sup-Convolution etc

In this Section, we collect various properties of semi-convex/concave functions, and sup/inf-convolutions used in the proof of Theorem 3. Many of these results are well known in the literature of viscosity solutions. However we still provide proofs for the sake of completeness. Also, in some cases we provide simpler proofs at the expense of obtaining results that are not as sharp, but they will suffice for the purpose of proving Theorem 3.

In what follows we take y to be a vector in $\mathbb{R}^n$. Additionally, for some function u, we let $\|Du\|$ denote the Lipschitz constant for u, with the convention that it is ∞ if u is not Lipschitz continuous.

Semi-Convexity and Concavity

Definition B2
A function u on $\mathbb{R}^n$ is said to be semi-convex with the coefficient c if $u(y) + \frac{c}{2}|y|^2$ is a convex function. Similarly, u is said to be semi-concave with the coefficient c if $u(y) - \frac{c}{2}|y|^2$ is concave.

The proof of Theorem 3 makes use of the following property of semi-convex functions. An analogous property also holds for semi-concave functions. We can also extend the scope of the theorem (i.e. also to points where Du does not exist) by considering one-sided derivatives, which can be shown to exist everywhere for semi-convex functions.

Lemma B1
Suppose that u is semi-convex. Then u is twice differentiable almost everywhere. Furthermore, for every point at which Du exists, we have for all $h \in \mathbb{R}^n$,

$$u(y+h) \geq u(y) + h^\intercal Du(y) - c|h|^2.$$

Define $g(y) = u(y) + \frac{c}{2}|y|^2$. Since $g(y)$ is convex, the Alexandrov theorem implies $g(\cdot)$ is twice continuously differentiable almost everywhere. Hence $u(y) = g(y) - \frac{c}{2}|y|^2$ is also twice differentiable almost everywhere.

For the second part of the theorem, observe that by convexity,

$$g(y+h) \geq g(y) + h^\intercal Dg(y).$$

Note that where the derivative exists, $Dg(y) = Du(y) + cy$. Hence,

$$u(y) + \frac{c}{2}|y+h|^2 \geq u(y) + \frac{c}{2}|y|^2 + h^\intercal Du(y) + ch^\intercal y.$$

Rearranging the above expression gives the desired inequality.□

Sup and Inf Convolutions

Let $u(y)$ denote a continuous function on some open set $\mathcal{Y}$. Let $\partial\mathcal{Y}$ denote the boundary of $\mathcal{Y}$, and $\bar{\mathcal{Y}}$ its closure.

Definition B3
The function u^ϵ is said to be the sup-convolution of u if

$$u^\epsilon(y) = \sup_{w \in \bar{\mathcal{Y}}} \left\{ u(w) - \frac{1}{2\epsilon}|w-y|^2 \right\}.$$

Similarly, u_ϵ is said to be the inf-convolution of u if

$$u_\epsilon(y) = \inf_{w \in \bar{\mathcal{Y}}} \left\{ u(w) + \frac{1}{2\epsilon}|w-y|^2 \right\}.$$

We shall also define y^ϵ as the value for which

$$u(y^\epsilon) - \frac{1}{2\epsilon}|y^\epsilon - y|^2 = u^\epsilon(y),$$

if y^ϵ lies in $\mathcal{Y}$ (otherwise it is taken to be undefined). Analogously, y_ϵ is the value for which

$$u(y_\epsilon) + \frac{1}{2\epsilon}|y_\epsilon - y|^2 = u_\epsilon(y).$$

Additionally, define $\mathcal{Y}^\epsilon$ as the set of all points in $\mathcal{Y}$ that are atleast $2\left\|Du\right\|\epsilon$ distance away from $\partial\mathcal{Y}$, i.e.

$$\mathcal{Y}^\epsilon := \{y \in \mathcal{Y} : |y - w| \geq 2\left\|Du\right\|\epsilon \ \forall w \in \partial\mathcal{Y}\}.$$

We have the following properties for sup and inf-convolutions:

Lemma B2
Suppose that u is continuous. Then,

(i) *u^ϵ is semi-convex with coefficient $1/\epsilon$. Similarly, u_ϵ is semi-concave with coefficient $1/\epsilon$.*

(ii) *$|y^\epsilon - y| \leq 2\left\|Du\right\|\epsilon$ and $|y_\epsilon - y| \leq 2\left\|Du\right\|\epsilon$.*

(iii) *for all $y \in \mathcal{Y}^\epsilon$, $|u^\epsilon(y) - u(y)| \leq 4\left\|Du\right\|^2\epsilon$ and $|u_\epsilon(y) - u(y)| \leq 4\left\|Du\right\|^2\epsilon$.*

We show the above properties for u^ϵ and y^ϵ. The claims for u_ϵ and y_ϵ follow in an analogous manner.

For (i), observe that

$$u^\epsilon(y) + \frac{1}{2\epsilon}|y|^2 = \sup_{w \in \bar{\mathcal{Y}}} \left\{ u(w) + \frac{1}{\epsilon}w^\intercal y - \frac{1}{2\epsilon}|w|^2 \right\}.$$

The right hand side of the above expression is in the form of a supremum over affine functions, which is convex. Hence (i) follows by the definition of semi-convex functions.

For (ii), by the definition of y^ϵ and u^ϵ,

$$\frac{1}{2\epsilon}|y^\epsilon - y|^2 \leq u(y^\epsilon) - u(y) \leq \left\|Du\right\||y^\epsilon - y|.$$

Rearranging the above inequality we get $|y^\epsilon - y| \leq 2\left\|Du\right\|\epsilon$.

For (iii), by the definition of y^ϵ (which exists for $y \in \mathcal{Y}^\epsilon$ in view of part (ii)),

$$\begin{aligned}|u^\epsilon(y) - u(y)| &= \left|u(y^\epsilon) - u(y) + \frac{1}{2\epsilon}|y^\epsilon - y|^2\right| \\ &\leq \|Du\|\,|y^\epsilon - y| + \frac{1}{2\epsilon}|y^\epsilon - y|^2 \\ &\leq 4\,\|Du\|^2\,\epsilon,\end{aligned}$$

where the last inequality follows by (ii).□

Lemma B3

Assume that u is uniformly continuous. Suppose that $\phi \in C^2(\mathcal{Y})$, such that $u^\epsilon - \phi$ has a local maximum at $y_0 \in \mathcal{Y}^\epsilon$. Define $\psi(y) = \phi(y + y_0 - y_0^\epsilon)$. Then $u - \psi$ has a local maximum at $y_0^\epsilon \in \mathcal{Y}$, and

$$D\psi(y_0^\epsilon) = D\phi(y_0) = \frac{1}{2\epsilon}(y_0^\epsilon - y_0).$$

Since $u^\epsilon - \phi$ has a local maximum at y_0, this implies there is a ball $B(y_0, r)$ of radius r around y_0 for which

$$u^\epsilon(y_0) - \phi(y_0) \geq u^\epsilon(w) - \phi(w)$$

for all $w \in B(y_0, r)$. Hence,

$$\begin{aligned}u(y_0^\epsilon) - \frac{1}{2\epsilon}|y_0^\epsilon - y_0|^2 - \phi(y_0) &\geq u^\epsilon(w) - \phi(w) \\ &\geq u(y) - \frac{1}{2\epsilon}|w - y|^2 - \phi(w)\end{aligned}$$

for all y and $w \in B(y_0, r)$ (note that $y_0^\epsilon \in \mathcal{Y}$ in view of the definition of $\mathcal{Y}^\epsilon$ and Lemma B2). This implies that (y_0^ϵ, y_0) is the local maximum of the function

$$\Upsilon(y, w) := u(y) - \frac{1}{2\epsilon}|w - y|^2 - \phi(w).$$

In other words,

$$\Upsilon(y_0^\epsilon, y_0) \geq \Upsilon(y, w) \quad \forall y \text{ and } w \in B(y_0, r). \tag{B.29}$$

In view of (B.29), we have $\Upsilon(y_0^\epsilon, y_0) \geq \Upsilon(w - y_0 + y_0^\epsilon, w)$ for all $w \in B(y_0, r)$, which implies

$$u(y_0^\epsilon) - \frac{1}{2\epsilon}|y_0^\epsilon - y_0|^2 - \phi(y_0) \geq u(w - y_0 + y_0^\epsilon) - \frac{1}{2\epsilon}|y_0^\epsilon - y_0|^2 - \phi(w).$$

Hence, for all $w \in B(y_0, r)$,

$$u(y_0^\epsilon) - \phi(y_0) \geq u(w - y_0^\epsilon + y^\epsilon) - \phi(w).$$

Now set $y^* = w - y_0 + y_0^\epsilon$ and observe that $|y^* - y_0^\epsilon| = |w - y_0| \leq r$ for all $w \in B(y_0, r)$. We thus obtain that for all $y^* \in B(y^\epsilon, r)$,

$$u(y_0^\epsilon) - \phi(y_0) \geq u(y^*) - \phi(y^* + y_0 - y_0^\epsilon).$$

In view of the definition of $\psi(.)$, the above implies

$$u(y_0^\epsilon) - \psi(y_0^\epsilon) \geq u(y^*) - \psi(y^*) \ \forall y^* \in B(y_0^\epsilon, r).$$

Hence $u - \psi$ has a local maximum at y_0^ϵ.

For the second part of the lemma, observe that by (B.29), $\Upsilon(y_0^\epsilon, y_0) \geq \Upsilon(y_0^\epsilon, w)$ for all $w \in B(y_0, r)$, which implies (after some rearrangement)

$$\frac{1}{2\epsilon}|y_0^\epsilon - w|^2 + \phi(w) \geq \frac{1}{2\epsilon}|y_0^\epsilon - y_0|^2 + \phi(y_0), \ \forall w \in B(y_0, r).$$

Hence the function $\theta(w) := \frac{1}{2\epsilon}|y_0^\epsilon - w|^2 + \phi(w)$ has a local minimum at $w = y_0$. Consequently,

$$D\phi(y_0) = \frac{1}{2\epsilon}(y_0^\epsilon - y_0).$$

This proves the second claim after noting $D\psi(y_0^\epsilon) = D\phi(y_0)$.□

Our next lemma considers PDEs of the form

$$F(Du(y), u(y), y) = 0, \quad u = 0 \text{ on } \partial\mathcal{Y},$$

where the domain of the PDE is the open set $\mathcal{Y}$. We shall assume that $F(\cdot)$ satisfies the following property (here $C < \infty$ denotes some constant)

$$|F(p, q_1, y_1) - F(p, q_2, y_2)| \leq Cp\{|q_1 - q_2| + |y_1 - y_2|\}. \tag{B.30}$$

Lemma B4

Suppose that u is a viscosity solution of $F(Du, u, y) = 0$, and $\|Du\| \leq m < \infty$. Suppose also that $F(\cdot)$ satisfies (B.30). Then there exists some c depending on only C (from B.30) and m such that for all $y \in \mathcal{Y}^\epsilon$,

$$F(Du^\epsilon, u^\epsilon, y) \leq c\epsilon,$$

where the above holds in the viscosity sense.

Take any $\varphi \in C^2(\mathcal{Y})$ such that $u^\epsilon - \varphi$ has a local maximum at $y_0 \in \mathcal{Y}^\epsilon$. Set $\psi(y) := \varphi(y + y_0 - y_0^\epsilon)$. Then by Lemma B3, $u - \psi$ has a local maximum at $y_0^\epsilon \in \mathcal{Y}$. Hence by definition of the viscosity solution

$$F(D\varphi(y_0^\epsilon), u(y_0^\epsilon), \varphi) \leq 0. \tag{B.31}$$

Recall also from Lemma B3 that

$$|D\psi(y_0)| = |D\phi(y_0^\epsilon)| = \frac{1}{2\epsilon}|y_0^\epsilon - y_0| \leq \|Du\| \leq m.$$

We then have

$$\begin{aligned}
|F(D\psi(y_0), u^\epsilon(y_0), y_0) - F(D\varphi(y_0^\epsilon), u(y_0^\epsilon), y_0^\epsilon)| \\
\leq Cm\left\{|u^\epsilon(y_0) - u(y_0^\epsilon)| + |y_0 - y_0^\epsilon|\right\} \\
\leq Cm\left\{|u^\epsilon(y_0) - u(y_0)| + (1+m)|y_0 - y_0^\epsilon|\right\} \\
\leq Cm\{4m^2 + 2m(1+m)\}\epsilon := c\epsilon,
\end{aligned}$$

where the first inequality follows from (B.30) and the last inequality from Lemma B2. We thus obtain in view of the above and (B.31) that

$$F(D\psi(y_0), u^\epsilon(y_0), y_0) \leq c\epsilon. \tag{B.32}$$

Since c is a constant, we have thus shown that if $u^\epsilon - \varphi$ has a local maximum at some $y_0 \in \mathcal{Y}^\epsilon$, then (B.32) holds. This implies that in a viscosity sense

$$F(Du^\epsilon, u^\epsilon, y) \leq c\epsilon.$$

□

B.4 JTPA Application: Cluster Descriptions

We employ k-median clustering (a well-established method, for full details see Anderberg, 1973). The aim is to divide the candidates into k clusters. The clusters are chosen such that the characteristics of each candidate (age, education, and previous earnings in our JTPA example) are as close as possible to the characteristic-medians of their cluster. Formally, the clusters are chosen such that the squared sum of Euclidean distances between the vector of characteristics of each candidate and the vector of characteristic-medians of their cluster is as small as possible.

In practice, we use Lloyd's algorithm, as is usual for k-median clustering. First, we start with k randomly selected candidates (k can be chosen freely, $k = 4$ in our JTPA example), which are the 'founding members' of each cluster. All other candidates that are then allocated to the cluster with the smallest Euclidean distance between the vector of characteristics of the candidate and the founding member. Second, the median of each cluster's characteristics is computed and denoted 'centroid'. Each candidate is then re-allocated to the cluster with the smallest Euclidean distance between the vector of characteristics of the candidate and the centroid. The second step is repeated until convergence, i.e until no more re-allocations occur.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Age: Mean	31.8	44.9	31.3	26.9
Age: Min.	22	34	22	22
Age: Max.	63	78	57	34
Prev. Earnings: Mean	8999	1439	1413	1231
Prev. Earnings: Min.	3600	0	0	0
Prev. Earnings: Max.	63000	12000	9076	5130
Education: Mean	12.1	12.1	9.0	12.3
Education: Min.	7	8	7	11
Education: Max.	18	18	10	18
Observations	2278	2198	1698	3049

Table B.1: Cluster Descriptions

Table B.1 provides summary statistics for the variables in each cluster. Cluster 1 appears to contain predominantly candidates with high previous earnings. Cluster 2's distinguishing factor is the high age, and for cluster 3 it is few years of education. Cluster 4 contains young educated candidates with low previous earnings.

C Supplementary Material and Proofs to Chapter 3

C.1 Proof of Theorem 4

Part 1)

For the boots market, condition (3.7) is satisfied (see condition (3.8) and Lemma 4). Consequently, when focusing on the boots market in isolation, the equilibrium cost of signaling (by producing high-quality at a loss) is $s_{bet}^{isolation} = \frac{D_t}{1-\frac{1+g}{1+r}}\left(\frac{k-c_{beh}}{N_{bht}^{isolation}+1}\right)^2$, where $N_{bht}^{isolation}$ is defined according to N_{ht}^* in the single-variant case outlined in the proof of Lemma 4. In the case of two variants, signaling via producing high-quality boots at a loss is still possible. There may be a cheaper (and hence superior) way to signal initially. Moreover, even if not, there may be a cheaper way to signal in the future, so that N_{bht} could be increasing in the long run. At $t = 0$, this is (virtually[121]) equivalent to replacing g with $\hat{g} < g$. Consequently, $s_{be0}^* \leq s_{be0}^{isolation}$. If $D_0 < M\left(1 - \frac{1+g}{1+r}\right)\left(\frac{N_{bh0}^{isolation}+1}{k-c_{beh}}\right)^2$, then $s_{bet}^* < M$ as $s_{bet}^{isolation} < M$, and so any form of high-quality production in the sandals market is inferior (since more expensive) to signaling in the boots market.

Part 2)

As shown in part 1, the initially optimal way of signaling is producing high-quality boots at a loss. If this remained to be the optimal way, then s_{bet}^* would be equal to $s_{bet}^{isolation} = \frac{D_t}{1-\frac{1+g}{1+r}}\left(\frac{k-c_{beh}}{N_{bht}^{isolation}+1}\right)^2$. This is increasing linearly in D_t, which is growing at rate g. Therefore, there exist a point in time $\tilde{t}$ for which $s_{be\tilde{t}}^{isolation} > M$ if $t > \tilde{t}$. This contradicts the notion that signaling via producing high-quality boots at a loss remains to be the optimal way of signaling. Due to Lemma 5, the other relevant way of signaling is via producing high-quality sandals with a NFP status and then entering the boots market with a FP status (entering the boots market with a NFP status is not optimal due to the assumed condition (3.15), as shown in part 4 below). This is optimal when $s_{bet}^* > M$. As shown in part 1, this is not the case for $t = 0$. Once this "NFP signaling" is optimal, the period-profit is constant by Lemma 7. It remains to be shown that if "NFP signaling" becomes optimal for $t \geq \tau$, and consequently firms anticipate period-profits to become stagnant, s_{bet}^* is still increasing for $t < \tau$, and reaching M at $t = \tau$.[122] At $t = 0 < \tau$ (the latter following from part 1), N_{bh0}^* is given by using $t = 0$ in:

$$\frac{1-\left(\frac{1+\hat{g}}{1+r}\right)^{\tau-t}}{1-\frac{1+\hat{g}}{1+r}}\frac{D_t}{\left(N_{bht}+1\right)^2}\left(k-c_{beh}\right)^2 + \frac{\left(\frac{1}{1+r}\right)^{\tau-t}}{1-\frac{1}{1+r}}\frac{D_\tau}{\left(N_{bh\tau}+1\right)^2}\left(k-c_{beh}\right)^2 = s_{bdt}\frac{c_{beh}-c_{bdl}}{c_{bdh}-c_{bdl}}, \tag{C.1}$$

[121] The growth of N_{bht} can no longer be constant in that case. So $\hat{g}$ is the "average" (constant) growth rate of the process viewed at $t = 0$, defined such that the discounted profit is equivalent to that under the (real) non-constant growth rates.

[122] Specifically, at $t = \tau$, the profit from signaling in the previous period to now enter is zero (by the definition of τ). In the subsequent period, it is larger than zero. Consequently, I assume that the NFP firm is created at $t = \tau$ and entry into the boots market occurs one period thereafter. See below regarding the number of periods needed for signaling (which I assume to be one).

$$\left(k - c_{bdl} - \frac{N_{bht}}{N_{bht}+1}\left(k - c_{beh}\right)\right)\frac{D_t}{N_{bht}+1}\left(k - c_{beh}\right) = s_{bdt}. \tag{C.2}$$

The latter equality does not include any profits from milking the sandals market, which is verified in part 4. Note that at any t for which $\tau \geq t > 0$, equation (C.1) no longer holds as condition (3.9) is slack. The discounted profits of entry via producing high-quality boots (left-hand side of equation (C.1)), can be expressed as

$$\tfrac{\varphi_t}{(N_{bht}+1)^2} = \sum_{x=t}^{\tau-1}\left(\tfrac{1+g}{1+r}\right)^x \tfrac{D_t}{(N_{bht}+1)^2}\left(k - c_{beh}\right)^2 + \left(\tfrac{1}{1+r}\right)^{\tau-t}\sum_{x=\tau}^{\infty}\left(\tfrac{1}{1+r}\right)^x \tfrac{D_\tau}{(N_{bh\tau}+1)^2}\left(k - c_{beh}\right)^2.$$

Since $N_{bht} \leq N_{bh\tau}$ for $t < \tau$, this expression is, for a given N_{bht}, increasing in $\tau - t$, and hence decreasing in t (i.e. φ_t is decreasing in t). Moreover, for a given N_{bht}, the milking profits (left-hand side of equation (C.2)) are increasing in t, and so are the cost of signaling $s_{bet} = s_{bdt}\frac{c_{beh}-c_{bdl}}{c_{bdh}-c_{bdl}}$. Consequently, condition (3.9) is slack: a quality expert considering to enter the market at time t in $\tau \geq t > 0$ faces lower profits and higher signaling costs compared to $t = 0$, where quality experts were indifferent between entering or not. More formally, for $t \leq \tau$ and if both (C.1) and (C.2) were binding, then:

$$N_{bht}^{(C.1)binding} = \frac{\varphi_t\frac{c_{bdh}-c_{bdl}}{c_{beh}-c_{bdl}} - \left(k - c_{beh}\right)\left(k - c_{bdl}\right)D_t}{\left(c_{beh} - c_{bdl}\right)D_t}.$$

This expression is decreasing in t because φ_t is decreasing and D_t is increasing in t, which violates condition (3.14). Therefore, condition (3.9) is slack, condition (3.14) is binding, and $N_{bht}^* = N_{bh0}^*$ for $\tau \geq t > 0$. This result implies that $\hat{g} = g$ and $N_{bh\tau} = N_{bh0}^*$,[123] and thus

$$N_{bh0}^* = \frac{\left(c_{bdh} - c_{bdl}\right)\left(k - c_{beh}\right)}{\left(c_{beh} - c_{bdl}\right)^2}\left(\frac{1+r}{r-g} + \left(\frac{1+g}{1+r}\right)^\tau\left(\frac{1+r}{r} - \frac{1+r}{r-g}\right)\right) - \frac{k - c_{bdl}}{c_{beh} - c_{bdl}}.$$

Equation (C.2) is not affected by condition (3.9) being slack, so

$$s_{bet}^* = \frac{c_{beh} - c_{bdl}}{c_{bdh} - c_{bdl}}\left(k - c_{bdl} - \frac{N_{bh0}^*}{N_{bh0}^*+1}\left(k - c_{beh}\right)\right)\frac{D_t}{N_{bh0}^*+1}\left(k - c_{beh}\right)$$

for t in $\tau \geq t > 0$, which is linearly increasing in D_t. In sum, for $t \leq \tau$, s_{bet}^* is increasing over time with a constant rate and it will reach M eventually. The point in time where it reaches M is by definition τ.

Part 3)

Follows directly from Lemma 7. As $\frac{D_t}{(N_{bht}+1)^2}$ is constant when $\tilde{s}_{et}^{NFP} < s_{bet}$, and the latter is true iff $t > \tau$, $(N_{bht}+1)^2$ must be increasing at rate g when $t > \tau$.

[123] I again assume a constant rate of change of $\pi_{et}^{cournot,period}$ for $\tau \geq t > 0$. See subsequent appendix-section for a discussion.

Part 4)

For $t \geq \tau$, $p_{sht} = c_{seh}$ due to this price being offered by a quality expert with NFP status. For $t < \tau$, the profit is $\sum_{x=t}^{\tau-1} \left(\frac{1+g}{1+r}\right)^x \frac{d_t}{(N_{sht}+1)^2} (j - c_{seh})^2$. Under the weak assumption that $\frac{1+g}{1+r} > g$, this profit is decreasing in t. Moreover, it is decreasing in N_{sht}. So the highest profit $\pi_{max}^{benefit}$ is achieved when only one firm enters as soon as it has established a high reputation in the boots-market, i.e. at $t = 1$, and therefore $\pi_{max}^{benefit} = \sum_{x=1}^{\tau-1} \left(\frac{1+g}{1+r}\right)^x \frac{(1+g)d_0}{4} (j - c_{seh})^2$. This profit is increasing in τ. The lowest τ under which non-negative profit can be made is given by $\left(\frac{1+g}{1+r}\right) \frac{1-\left(\frac{1+g}{1+r}\right)^{\tau_{lowest}-1}}{1-\left(\frac{1+g}{1+r}\right)} \frac{(1+g)d_0}{4} (j - c_{seh})^2 = M$, hence

$$\tau_{lowest} = \frac{ln\left(1 - \frac{4M(r-g)}{d_0(1+g)^2(j-c_{seh})^2}\right)}{ln\left(\frac{1+g}{1+r}\right)} + 1.$$

The point in time τ is defined by s_{bet}^* reaching M. Note that N_{bh0}^* is a function of τ, and use

$$N_{bh0}^* = \alpha + \left(\frac{1+g}{1+r}\right)^\tau \beta$$

with $\alpha = \frac{(c_{bdh}-c_{bdl})(k-c_{beh})}{(c_{beh}-c_{bdl})^2} \left(\frac{1+r}{r-g} - \frac{k-c_{bdl}}{c_{beh}-c_{bdl}}\right)$ and $\beta = \frac{(c_{bdh}-c_{bdl})(k-c_{beh})}{(c_{beh}-c_{bdl})^2} \left(\frac{1+r}{r} - \frac{1+r}{r-g}\right)$. Then τ^* is given by the following equation:

$$k - c_{bdl} - \frac{1}{1 + \frac{1}{\alpha + \left(\frac{1+g}{1+r}\right)^{\tau^*} \beta}} (k - c_{beh}) = \frac{M\left(c_{bdh} - c_{bdl}\right)}{D_0\left(k - c_{beh}\right)\left(c_{beh} - c_{bdl}\right)} \left(\frac{\alpha + 1}{(1+g)^{\tau^*}} + \frac{\beta}{(1+r)^{\tau^*}}\right).$$

If $max\left(\tau^*\right) < \tau_{lowest}$, no for-profit firm will enter the high-quality sandals market. As outlined in part 2 above, τ^* must be bounded away from infinity, and by part 1, it is bounded away from zero as well.

The condition $max\left(\tau^*\right) < \tau_{lowest}$ is the necessary and sufficient condition that firms that have obtained a high-quality reputation in the boots-market (with FP or NFP status) do not benefit from it in the sandals market. It is not straightforward to interpret though (note that $ln\left(\frac{1+g}{1+r}\right) < 0$). The much easier condition outlined in condition (3.15) above is obtained by the following argument. The profit $\sum_{x=1}^{\tau-1} \left(\frac{1+g}{1+r}\right)^x \frac{(1+g)d_0}{4} (j - c_{seh})^2 < \frac{(1+g)d_0}{4} (j - c_{seh})^2 \frac{1}{1-\frac{1+g}{1+r}}$. If the latter is smaller than M, then this is also a sufficient (albeit too strong) condition.□

C.2 Non-Constant Rate of Change of the Period-Profit

So far, I have assumed a constant (possibly zero) rate of change of $\pi_{bet}^{cournot,period}$ ($\pi_{et}^{cournot,period}$ in section 3.2 respectively). Since D_t grows at a constant rate, this implies the assumption that $(N_{bht} + 1)$ ($N_{ht} + 1$ respectively) grows at a constant rate as well. This assumption does not affect Theorem 4 (or the underlying lemmas) qualitatively for the following reasons.

Due to condition (3.14), $(N_{bht} + 1)$ cannot be decreasing (condition (3.4) regarding $(N_{ht} + 1)$ respectively).

Moreover, $\big(N_{(b)ht}+1\big)$ cannot be increasing at an increasing rate. If the growth rate of $\big(N_{(b)ht}+1\big)$ is low at time t and higher at time t', then $N^*_{(b)ht} > N^*_{(b)ht'}$. This is due to the fact that having a high reputation is more profitable in a state where the growth rate is still relatively low (as future profits are discounted). For $t > \tau$, the cost of the signal is constant, concluding the argument. While the cost of the signal is eventually determined by $\pi^{cournot,milk}_{(b)et}$ for $t \leq \tau$ (or the case in section 3.2), the cost of the signal is decreasing in $\big(N_{(b)ht}+1\big)$ (see equation (3.6), and (3.13) accordingly), but $\pi^{cournot,period}_{(b)et}$ decreases faster as $\big(N_{(b)et}+1\big)$ increases (at least for large values of $\big(N_{(b)et}+1\big)$).

For $t > \tau$, $(N_{bet}+1)$ cannot be increasing at a decreasing rate with $(N_{bet}+1)$ converging to a constant. Since the cost of signaling is constant and the demand is increasing, this would imply that entrants could make a positive profit, which refutes $(N_{bet}+1)$ being/approaching a constant.

Should there be equilibria for $t > \tau$ with $(N_{bet}+1)$ increasing at a decreasing rate but not converging, none of the results change qualitatively. Theorem 4 as well as the Lemmas 3 to 7 still hold. The only difference was that $(N_{bet}+1)$ would grow a rate faster than $\sqrt{1+g}-1$ initially.

Should there be equilibria for $t \leq \tau$ (or the case in section 3.2) with $\big(N_{(b)et}+1\big)$ increasing at a decreasing rate, again none of the results change. For a sufficiently large t, the growth rate of $\big(N_{(b)et}+1\big)$ becomes constant. Following the proof of Theorem 4.2 and Lemma 4 respectively, it must be zero (i.e. in this case, non-converging $\big(N_{(b)et}+1\big)$ cannot be an equilibrium). At least eventually, as the growth rate of $\big(N_{(b)et}+1\big)$ is sufficiently low, Lemma 3 and 4 hold. The Lemmas 5 to 7 are not affected, and consequently Theorem 4 holds.

Finally, I assume $\big(N_{(b)et}+1\big)$ not to be increasing at a fluctuating rate. Each period has identical exogenous conditions, except for D_t that increases at a constant rate, and at least for $t \leq \tau$ and $t > \tau$ taken separately, either identical or smoothly developing endogenous variables. A different reaction to the same event (D_t increasing) could only be caused by beliefs that reactions to "D_t increasing" are different in the future (affecting the expected discounted profits). If there are multiple equilibria (potentially introduced by $(N_{bet}+1)$ increasing at a decreasing rate or τ^* not being unique), I disregard cases that alternate between equilibria.

C.3 Proof of Theorem 5

Part 1) & Part 2)

Denote $M(N_{sh})$ the expected cost M as a function of the number of firms with a high-quality reputation operating in the sandals market. As a shorthand, I use M_0 as the expected cost when there are no firms producing high-quality sandals and M_1 as the cost when there is at least one such firm. In the mechanism presented in section 3.3, a key feature is that NFP signaling is free to quality experts, except for the constant expected cost M_0, and $M_1 = M_0$. If I imposed that $M_1 = 0$, quality experts would be constantly incentivized to create NFP firms in the sandals market, leading to an explosion of such NFP firms.

However, with motivated workers, even quality experts now face a positive cost of operating a NFP firm producing high-quality sandals, provided it is not the first firm in that market. The price of the first NFP firm is, by definition of its NFP status, given by price equal to costs: $p_{sht} = A_{sfh}(\bar{u} - \alpha i)$, with $f = e$ for the same reason as in Theorem 4: in equilibrium, only quality experts enter the high-quality markets. Define $\bar{u}$ as reservation utility of as many workers as needed to produce one pair of sandals. Accordingly, the influence of that many workers on the consumer surplus is $i = \frac{j-p_{sht}}{2}$ and the price is given by:

$$p_{sht} = c_{seh}^{1^{st}} = \frac{A_{seh}\left(2\bar{u} - \alpha j\right)}{2 - \alpha A_{seh}}.$$

Due to the assumption of constant returns to scale, this first NFP firm can serve the entire market at $p_{sht} = c_{seh}^{1^{st}}$. Any potential later quality experts face production costs per unit of $c_{seh}^{later} = A_{seh}\bar{u} > c_{seh}^{1^{st}}$. Consequently, all later firms make a loss of $\frac{\alpha A_{seh}(j-\bar{u}A_{seh})}{2-\alpha A_{seh}}$ per unit sold.

This implies that, as before, $\tilde{s}_{et}^{1^{st}NFP} = M_0$. However, imposing $M_1 = 0$ no longer results in $\tilde{s}_{et}^{laterNFP} = 0$, but $\tilde{s}_{et}^{laterNFP} = \tilde{h}_t \frac{\alpha A_{seh}(j-\bar{u}A_{seh})}{2-\alpha A_{seh}} > 0$. Regarding discounters, $\tilde{s}_{dt}^{laterNFP} = \tilde{h}_t \frac{2\bar{u}(A_{sdh}-A_{seh})+\alpha A_{seh}(j-\bar{u}A_{sdh})}{2-\alpha A_{seh}}$.

The characteristics of the equilibrium are generally akin to section 3.3 and proven analogue to Theorem 4. The key difference is that Theorem 4.3 only applies until one NFP firm is created. Thereafter, $\tilde{s}_{ft}^{NFP} = \tilde{s}_{ft}^{laterNFP}$, as outlined above. There is also no longer a "threat" of any form of entry that is not either high-quality boots production at a loss or high-quality sandals production at a loss (the latter with a NFP status).

According to the proof of Lemma 4, the resulting number of firms if the former is optimal is

$$N_{bht}^{bsignal} = \left(\frac{\frac{c_{bdh}-c_{bdl}}{c_{beh}-c_{bdl}}}{1-\frac{1+g}{1+r}} - 1\right)\frac{k - c_{beh}}{c_{beh} - c_{bdl}} - 1 \;\; (with\, c_{bfa} = \bar{u}A_{bfa}).$$

If the latter is optimal, the qualitative argument remains identical (and $N_{bht}^{ssignal}$ is constant for the same reason) and the key condition are (analogue to (3.5) and (3.6)):

$$\frac{1}{1-\frac{1+g}{1+r}}\frac{D_t}{\left(N_{bht}+1\right)^2}\left(k - c_{beh}\right)^2 = \tilde{s}_{dt}^{laterNFP}\frac{\alpha A_{seh}\left(j - \bar{u}A_{seh}\right)}{2\bar{u}\left(A_{sdh} - A_{seh}\right) + \alpha A_{seh}\left(j - \bar{u}A_{sdh}\right)},$$

$$\left(k - c_{bdl} - \frac{N_{bht}}{N_{bht}+1}\left(k - c_{beh}\right)\right)\frac{D_t}{N_{bht}+1}\left(k - c_{beh}\right) = \tilde{s}_{dt}^{laterNFP}.$$

Consequently,

$$N_{bht}^{ssignal} = \left(\frac{\frac{2\bar{u}(A_{sdh}-A_{seh})+\alpha A_{seh}(j-\bar{u}A_{sdh})}{\alpha A_{seh}(j-\bar{u}A_{seh})}}{1-\frac{1+g}{1+r}} - 1\right)\frac{k - c_{beh}}{c_{beh} - c_{bdl}} - 1 \;\; (with\, c_{bfa} = \bar{u}A_{bfa}).$$

Neither $N_{bht}^{bsignal}$ nor $N_{bht}^{ssignal}$ are unbounded, which proofs part 1.

If signaling through the production of high-quality boots is superior to signaling through the production of high-quality sandals, then $N^* = N_{bht}^{bsignal}$. Formally, signaling in the boots market is superior if a quality expert's loss of producing a high-quality boot and selling it at p_{bdl} relative to a discounter's loss of doing the same is lower than the analogue fraction for sandals, i.e.:

$$\frac{c_{beh} - c_{bdl}}{c_{bdh} - c_{bdl}} > \frac{\alpha A_{seh}\left(j - \bar{u} A_{seh}\right)}{2\bar{u}\left(A_{sdh} - A_{seh}\right) + \alpha A_{seh}\left(j - \bar{u} A_{sdh}\right)}. \tag{C.3}$$

Under condition (C.3), $max\left\{N_{bht}^{bsignal}, N_{bht}^{ssignal}\right\} = N_{bht}^{bsignal} = N_{bht}^*$. This implies that after one firm has entered via "NFP signaling", the situation is as outlined in section 3.2. Prior to entry via "NFP signaling", the only difference to section 3.2 is the expected entry of one firm, which will have created the NFP firm, at $t = \tau + 1$. The exception of this case is when an "inside firm" creates a NFP firm producing high-quality sandals (see part 3 below).

In sum, under condition (C.3) and absent of an "inside firm" creating the NFP firm, N_{bht}^* is constant for $t \leq \tau$ as outlined in the proof of Theorem 4.2. After one firm has entered via "NFP signaling", the situation is akin to section 3.2 (possibly with (3.4) binding instead of (3.1)) and N_{bht}^* is constant again. The only rational belief prior to the firm entering "via NFP signaling" is that exactly one firm will enter after $t = \tau$, and N_{bht}^* is determined accordingly. Finally, if it is an "inside firm" that creates the NFP firm, it is rational to expect this, and hence N_{bht}^* is constant (and does not even increase by one after $t = \tau$).[124]$\square$

Part 3)

I denote the firms that have a high-quality reputation and sell boots "inside firms". Inside firms rationally expect that if there is no NFP firm selling high-quality sandals, an outside quality expert will create such a NFP firm at $t = \tau$ and enter the market for high-quality boots thereafter. This reduces the net present value of the stream of profits they receive. Specifically, τ is defined such that $M_0 = \frac{\pi(N_{bh\tau}+1)}{1+N_{bh\tau}}$, where $\pi(.)$ is the net present value of the combined profits of all inside firms, taking into account that the equilibrium Cournot price (and hence profits) depends on N_{bht}. Each inside firm obtains $\frac{\pi(N_{bh\tau}+1)}{1+N_{bh\tau}}$ if the outsider joins and $\frac{\pi(N_{bh\tau})}{N_{bh\tau}}$ if it is an inside firm that creates the NFP.

If $N_{bh\tau} = 1$, then $\frac{\pi(N_{bh\tau})}{N_{bh\tau}} - \frac{\pi(N_{bh\tau}+1)}{1+N_{bh\tau}} = \pi(1) - \frac{\pi(2)}{2} > \frac{\pi(2)}{2} = M_0$ (at $t = \tau$) since $\pi(1) > \pi(2)$ as a monopolist sets prices higher than two Cournot competing firms and is therefore able to extract more surplus from the consumers. Consequently, a monopolist will always create the first NFP firm to produce high-quality sandals.

For $N_{bh\tau} > 1$, a coordination problem arises. Since the profit per firm and period is given by $\pi_t^{cournot,period} = \frac{D_t}{(N_{bht}+1)^2}\left(k - c_{beh}\right)^2$, I express $\pi(N_{bh}) = \frac{N_{bh} C}{(N_{bh}+1)^2}$, where C is a constant that does not depend on N_{bh}. For $N_{bh\tau} = 2$, $\frac{\pi(N_{bh\tau})}{N_{bh\tau}} - \frac{\pi(N_{bh\tau}+1)}{1+N_{bh\tau}} = \frac{C}{9} - \frac{C}{16} < \frac{C}{16} = \frac{\pi(N_{bh\tau}+1)}{1+N_{bh\tau}} = M_0$.

[124]Note: Much of this proof is due to constant returns to scale implying that a second NFP firm has no impact on consumer surplus given the first is operating (and serving the entire market). More generally, if NFP firms can choose to supply any given price and quantity as long as at least this quantity is demanded at the price (i.e. NFP firms are allowed to be out of stock), then one NFP firm can create at least half as much consumer surplus as two NFP firms. Consequently, a similar argument is potentially possible without relying on CRS, but still obtaining only a limited number of NFP firms created after $t = \tau$ and allowing for $M(N_{sh})$ to be decreasing. The formal details are left to future research.

The same holds generally for $N_{bh\tau} \geq 2$ since

$$max_{N_{bht} \geq 2} \left(\frac{C}{(N_{bht}+1)^2} - \frac{2C}{(N_{bht}+2)^2} \right) < 0.$$

Consequently, inside firms only create the NFP firm themselves if they are able to coordinate whom to creates it and how to split the cost. Note that for inside firms, an even split of costs is always preferable to an additional firm entering since $\frac{\pi(N_{bh\tau})}{N_{bh\tau}} - \frac{M_0}{N_{bh\tau}} = \frac{\pi(N_{bh\tau})(N_{bh\tau}+1)}{N_{bh\tau}(N_{bh\tau}+1)} - \frac{\pi(N_{bh\tau}+1)}{N_{bh\tau}(N_{bh\tau}+1)} > \frac{N_{bh\tau}\pi(N_{bh\tau}+1)}{N_{bh\tau}(N_{bh\tau}+1)} = \frac{\pi(N_{bh\tau}+1)}{1+N_{bh\tau}}$. In sum, it is an inside firm that creates the sandal-producing NFP firm if the inside firm is a monopolist or if the inside firms are able to coordinate.□

C.4 Equilibrium of the Market for Low-Quality Sandals with Motivated Workers

Should several firms be active in the sandals market, then none of them makes a differential impact to the poor and $p_{slt} = A_{sdl}\bar{u}$. What remains to be checked is whether in a dynamic setting, the first firm optimally charges a price that is lower than $A_{sdl}\bar{u}$ to benefit from the motivation of the workers.

A monopolist maximizes:

$$max_{p_{slt}}\ \hat{d}_t(y - p_{slt}) \left(p_{slt} - A_{sdl} \left(\bar{u} - \alpha \frac{y - p_{sht}}{2} \right) \right)$$

If $\alpha < \frac{2}{A_{sdl}}$, then the above function is concave and $p_{slt}^{Monopolist} = \frac{y}{2} + \frac{A_{sdl}(2\bar{u} - \alpha y)}{4 - 2\alpha A_{sdl}}$. If, in addition, $p_{slt}^{Monopolist} < A_{sdl}\bar{u}$, then this price constitutes a dynamic equilibrium. The latter condition can be rewritten as $y < A_{sdl}\bar{u} - A_{sdl}\alpha\left(A_{sdl}\bar{u} - y\right)$, which must fail if $y > A_{sdl}\bar{u}$.

In sum, if a firm can sell low-quality sandals without motivated workers ($y > A_{sdl}\bar{u}$), then the only equilibrium price is $p_{slt}^* = A_{sdl}\bar{u}$. I assume throughout the paper that the market for low-quality sandals is served. A sufficient assumption is $y > A_{sdl}\bar{u}$ (which is akin to the assumption that $y > c_{sdl}$ in section 3.3). However, this assumption can be replaced by any assumption that ensures that the market for low-quality sandals is served, such as $y < A_{sdl}\bar{u}$ and $\alpha < \frac{2}{A_{sdl}}$, in which case $p_{slt}^* = p_{slt}^{Monopolist}$.

C.5 Number of Periods Needed for Signaling

As mentioned in the footnotes 96 and 100, I assume throughout the paper that one period is always sufficient to signal, i.e. to sell the quantities $\frac{h_t}{1+r}$ or $\frac{\tilde{h}_t}{1+r}$ respectively. Regarding $\tilde{h}_t$ for $t > \tau$ in the case with motivated workers and h_t in general, this assumption is only for notational convenience. It can easily be changed, as outlined in footnote 96, without having any effects on the results presented in this paper.

Except for these cases, the issue arises that the expected cost M is realistically spent in the period where the NFP firm is created, not the period in which the FP firms enter the boots market. The cost M is therefore $(1+r)^x$ times the true development costs, where x refers to the number of periods needed for signaling. I assume x to always be equal to one, and hence M to be constant. Due to (3.10) and (3.12), this implies that I assume that:

$$Q^D_{sht}(c_{seh}) > \tilde{h}_t = max\left\{0, \left(\frac{Q^D_{bht}(p_{bht})}{c_{sdh} - c_{seh}} - M\right)\right\}, \; with$$

$$\frac{Q^D_{bht}(p_{bht})}{c_{sdh} - c_{seh}} = \frac{D_t(k - c_{beh})\left(k - c_{bdl} - \frac{N_{bht}}{N_{bht}+1}(k - c_{beh})\right)}{(N_{bht} + 1)(c_{sdh} - c_{seh})}.$$

Note that $\frac{Q^D_{bht}(p_{bht})}{c_{sdh}-c_{seh}}$ is strictly decreasing in N_{bht}, which is non-decreasing over time. Therefore, $\frac{Q^D_{bht}(p_{bht})}{c_{sdh}-c_{seh}}$ grows over time at a rate weakly less than g. Since $Q^D_{sht}(c_{seh}) = d_t(j - c_{seh})$, it grows at the rate g over time. Consequently, a sufficient condition is to assume that the "market for the poor" is large relative to the "market for the rich", such as:

$$d_0 > D_0 \frac{(k - c_{beh})\left(k - c_{bdl} - \frac{N^*_{bh0}}{N^*_{bh0}+1}(k - c_{beh})\right)}{\left(N^*_{bh0} + 1\right)(c_{sdh} - c_{seh})(j - c_{seh})}.$$

I make the above assumption throughout this paper.

References

Abadie, A., Angrist, J., and Imbens, G. (2002). Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. *Econometrica*, 70(1):91–117.

Abadie, A., Chingos, M. M., and West, M. R. (2018). Endogenous Stratification in Randomized Experiments. *The Review of Economics and Statistics*, (forthcoming).

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, 93(1):113–132.

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2017). Income and wealth distribution in macroeconomics: A continuous-time approach. Technical report, National Bureau of Economic Research.

Agarwal, A., Hsu, D., Kale, S., Langford, J., Li, L., and Schapire, R. (2014). Taming the monster: A fast and simple algorithm for contextual bandits. In *International Conference on Machine Learning*, pages 1638–1646.

Al Jazeera (2017). Hate Crimes Rise around Brexit Vote, Recent Attacks. *Retrieved from https://www.aljazeera.com/news/2017/10/hate-crimes-rise-brexit-vote-attacks-171018110119902.html*.

Albornoz, F., Bradley, J., and Sonderegger, S. (2018). The Brexit Referendum and the Rise in Hate Crime. *mimeo*.

Anderberg, M. R. M. R. (1973). *Cluster Analysis for Applications*. Academic Press, New York.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7353–60.

Athey, S. and Imbens, G. (2017). The State of Applied Econometrics - Causality and Policy Evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

Athey, S. and Wager, S. (2018). Efficient Policy Learning. *arXiv:1702.02896*.

Austin, S. R., Dialsingh, I., and Altman, N. S. (2014). Multiple Hypothesis Testing: A Review. *mimeo*.

BBC (2017a). Rise in Hate Crime in England and Wales. *Retrieved from https://www.bbc.co.uk/news/uk-41648865*.

BBC (2017b). The Truth about Hate Crime and Brexit. *Retrieved from https://www.bbc.co.uk/news/av/magazine-41584532/the-truth-about-hate-crime-and-brexit.*

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76(2):169–217.

Becker, S. O., Fetzer, T., and Novy, D. (2017). Who voted for Brexit? A comprehensive district-level analysis. *Economic Policy*, 32(92):601–650.

Belloni, A., Chernozhukov, V., Fernández-Val, I., and Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233–298.

Benitez-Silva, H., Hall, G., Hitsch, G. J., Pauletto, G., and Rust, J. (2000). A comparison of discrete and parametric approximation methods for continuous-state dynamic programming problems. *manuscript, Yale University.*

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*, 57(1):289–300.

Benjamini, Y. and Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics*, 29(4):1165–1188.

Bertrand, M., Crépon, B., Marguerie, A., and Premand, P. (2017). Contemporaneous and Post-Program Impacts of a Public Works Program: Evidence from Côte d'Ivoire. *World Bank Working Paper.*

Besley, T. and Ghatak, M. (2007). Retailing Public Goods: The Economics of Corporate Social Responsibility. *Journal of public Economics.*

Bhatnagar, S., Sutton, R. S., Ghavamzadeh, M., and Lee, M. (2009). Natural Actor-Critic Algorithms. *Automatica*, 45(11):2471–2482.

Bhattacharya, D. and Dupas, P. (2012). Inferring Welfare Maximizing Treatment Assignment under Budget Constraints. *Journal of Econometrics*, 167:168–196.

Bien, J., Taylor, J., and Tibshirani, R. (2013). A Lasso for Hierarchical Interactions. *The Annals of Statistics*, 41(3):1111–1141.

Blair, R. A., Blattman, C., and Hartman, A. (2017). Predicting Local Violence. *Journal of Peace Research*, 54(2):298–312.

Bonhomme, S., Lamadon, T., and Manresa, E. (2017). Discretizing unobserved heterogeneity. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2019-16).

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience Theory of Choice under Risk. *Quarterly Journal of Economics*, 127(3):1243–1285.

Borjas, G. J. (1995). The Economic Benefits from Immigration. *The Journal of Economic Perspectives*, 9(2):3–22.

Bursztyn, L., Egorov, G., and Fiorin, S. (2017). From Extreme to Mainstream: How Social Norms Unravel. *NBER Working Paper No. 23415*.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, 90(3).

Caselli, F. and Coleman, W. J. (2013). On the Theory of Ethnic Conflict. *Journal of the European Economic Association*, 11(Suppl. 1):161–192.

Chakraborty, B. and Murphy, S. A. (2014). Dynamic Treatment Regimes. *Annual Review of Statistics and Its Application*, 1(1):447–464.

Chamberlain, G. (2011). Bayesian Aspects of Treatment Choice. In Geweke, J., Koop, G., and Van Dijk, H., editors, *The Oxford Handbook of Bayesian Econometrics*, pages 11–39. Oxford University Press.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018a). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2018b). Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments. *NBER Working Paper No. 24678*.

Chudik, A., Kapetanios, G., and Pesaran, M. H. (2018). A One Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models. *Econometrica*, 86(4):1479–1512.

Craig-Henderson, K. and Sloan, L. R. (2003). After the Hate: Helping Psychologists Help Victims of Racist Hate Crime. *Clinical Psychology: Science and Practice*, 10(4):481–490.

Crandall, M. G., Ishii, H., and Lions, P.-L. (1992). User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American mathematical society*, 27(1):1–67.

Crandall, M. G. and Lions, P.-L. (1983). Viscosity solutions of hamilton-jacobi equations. *Transactions of the American mathematical society*, 277(1):1–42.

Crandall, M. G. and Lions, p.-L. (1987). Remarks on the existence and uniqueness of unbounded viscosity solutions of hamilton-jacobi equations. *Illinois J. Math.*, 31(4):665–688.

Daily Mail (2016). Epidemic of Race Crimes since Brexit are Simply False. *Retrieved from http://www.dailymail.co.uk/news/article-3805008/The-great-Brexit-hate-crime-myth-claims-epidemic-race-crimes-referendum-simply-false.html.*

Danone (2012). 15 Years that Transformed Danone. *Retrieved from http://danone11.danone.com/en/dataviz.*

De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., Hughes, C. O., Raine, R., Hughes, J., Sim, D. A., Egan, C., Tufail, A., Montgomery, H., Hassabis, D., Rees, G., Back, T., Khaw, P. T., Suleyman, M., Cornebise, J., Keane, P. A., and Ronneberger, O. (2018). Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease. *Nature Medicine*, 24(9):1342–1350.

Degris, T., White, M., and Sutton, R. S. (2012). Off-policy actor-critic. *arXiv preprint arXiv:1205.4839.*

Devine, D. (2018). The UK Referendum on Membership of the European Union as a Trigger Event for Hate Crimes. *SSRN Electronic Journal.*

Dimakopoulou, M., Athey, S., and Imbens, G. (2017). Estimation considerations in contextual bandits. *arXiv preprint arXiv:1711.07077.*

Douch, M., Edwards, H., and Soegaard, C. (2018). The Trade Effects of the Brexit Announcement Shock. *Warwick Economics Research Papers No: 1176.*

Dudik, M., Hsu, D., Kale, S., Karampatziakis, N., Langford, J., Reyzin, L., and Zhang, T. (2011). Efficient optimal learning for contextual bandits. *arXiv preprint arXiv:1106.2369.*

Dunn, O. J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, 56(293):52.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least Angle Regression. *The Annals of Statistics*, 32(2):407–499.

Electoral Commission (2016). EU Referendum Results. *Retrieved from https://www.electoralcommission.org.uk/find-information-by-subject/elections-and-referendums/past-elections-and-referendums/eu-referendum/electorate-and-count-information.*

Esteban, B. J., Mayoral, L., and Ray, D. (2012). Ethnicity and Conflict: An Empirical Study. *The American Economic Review*, 102(4):1310–1342.

Esteban, J. and Ray, D. (2011). A Model of Ethnic Conflict. *Journal of the European Economic Association*, 9(3):496–521.

Falk, A., Kuhn, A., and Zweimüller, J. (2011). Unemployment and Right-Wing Extremist Crime. *Scandinavian Journal of Economics*, 113(2):260–285.

Fetzer, T. (2018). Did Austerity Cause Brexit? *University of Warwick Working Paper Series No.381.*

Financial Times (2017). UK Hate Crime Figures Spike after Brexit Referendum, Terror Attack. *Retrieved from https://www.ft.com/content/6ecbde7a-800a-31e8-9b12-1ce7f6897c4b*.

Fisher, R. A. (1935). *The Design of Experiments*. Hafner, New York.

Francois, P. (2000). 'Public Service Motivation' as an Argument for Government Provision. *Journal of Public Economics*, 78(3):275–299.

Francois, P. (2003). Not-For-Profit Provision Of Public Services. *The Economic Journal*, 113(486):C53–C61.

Freedman, D. and Lane, D. (1983). A Nonstochastic Interpretation of Reported Significance Levels. *Journal of Business and Economic Statistics*, 1(4):292–298.

Gerstenfeld, P. B. (2017). *Hate Crimes - Causes, Controls, and Consequences*. SAGE Publications, London.

Ghatak, M. and Mueller, H. (2011). Thanks for nothing? Not-for-profits and motivated agents. *Journal of Public Economics*.

Glaeser, E. and Shleifer, A. (2001). Not-for-Profit Entrepreneurs. *Journal of Public Economics*, 81(1):99–115.

Glaeser, E. L. (1994). Cities, Information, and Economic Growth. *Cityscape*, 1(1):9–47.

Grameen (2014a). Grameen Danone foods Ltd. *Retrieved from http://www.grameencreativelab.com/live-examples/grameen-danone-foods-ltd.html*.

Grameen (2014b). Network - Grameen Family of Businesses. *Retrieved from http://www.grameencreativelab.com/our-company/network.html*.

Grameen Uniqlo (2014a). Brand Concept. *Retrieved from http://www.grameenuniqlo.com/en/about/index.html*.

Grameen Uniqlo (2014b). Company Information. *Retrieved from http://www.grameenuniqlo.com/en/corp/index.html*.

Hall, N. (2013). *Hate Crime*. Routledge, New York, 2nd edition.

Hansmann, H. B. (1980). The Role of Nonprofit Enterprise. *The Yale Law Journal*, 89(5):835.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2nd edition.

Hauschild, H. (2014). Adidas Tested den Ein-Euro-Schuh [Adidas Tests the One-Euro-Shoe]. *Retrieved from http://www.handelsblatt.com/unternehmen/industrie/in-bangladesch-adidas-testet-den-ein-euro-schuh/3632580.html*.

Hebiri, M. and Lederer, J. (2013). How Correlations Influence Lasso Prediction. *IEEE Transactions on Information Theory*, 59(3):1846–1854.

Hirano, K. and Porter, J. R. (2009). Asymptotics for Statistical Treatment Rules. *Econometrica*, 77(5):1683–1701.

Home Office (2017). Hate Crime, England and Wales, 2016/17. *Statistical Bulletin 17/17SW*.

Home Office (2018). How Many People Do We Grant Asylum or Protection to? *Retrieved from https://www.gov.uk/government/publications/immigration-statistics-year-ending-march-2018/how-many-people-do-we-grant-asylum-or-protection-to*.

Independent (2016). Racism Unleashed: True Extent of the 'Explosion of Blatant Hate' that Followed Brexit Result Revealed. *Retrieved from https://www.independent.co.uk/news/uk/politics/brexit-racism-uk-post-referendum-racism-hate-crime-eu-referendum-racism-unleashed-poland-racist-a7160786.html*.

Kempthorne, O. and Folks, L. (1971). *Probability, Statistics, and Data Analysis*. Iowa State University Press.

Kim, J.-H. (2009). Estimating Classification Error Rate: Repeated Cross-Validation, Repeated Hold-Out and Bootstrap. *Computational Statistics & Data Analysis*, 53(11):3735–3745.

King, R. D. and Sutton, G. M. (2013). High Times for Hate Crimes: Explaining the Temporal Clustering of Hate-Motivated Offending. *Criminology*, 51(4):871–894.

Kirch, C. (2007). Block Permutation Principles for the Change Analysis of Dependent Data. *Journal of Statistical Planning and Inference*, 137(7):2453–2474.

Kitagawa, T. and Tetenov, A. (2018). Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica*, 86(2):591–616.

Knaus, M. C., Lechner, M., and Strittmatter, A. (2017). Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach. *IZA DP No. 10961*.

Krueger, A. B. and Pischke, J.-S. (1997). A Statistical Analysis of Crime against Foreigners in Unified Germany. *The Journal of Human Resources*, 32(1):182–209.

Krylov, N. V. (2005). The rate of convergence of finite-difference approximations for bellman equations with lipschitz coefficients. *Applied Mathematics & Optimization*, 52(3):365–399.

Kuhn, M. and Johnson, K. (2014). Comparing Different Species of Cross-Validation - Applied Predictive Modeling. *Retrieved from http://appliedpredictivemodeling.com/blog/2014/11/27/vpuig01pqbklmi72b8lcl3ij5hj2qm*.

Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). Penalized Regression, Standard Errors, and Bayesian Lassos. *Bayesian Analysis*, 5(2):369–412.

Laber, E. B., Lizotte, D. J., Qian, M., Pelham, W. E., and Murphy, S. A. (2014). Dynamic treatment regimes: Technical challenges and applications. *Electronic journal of statistics*, 8(1):1225.

Laffont, J.-J. and Tirole, J. (1986). Using Cost Observation to Regulate Firms. *Journal of Political Economy*, 94(3):614–641.

Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3):907–927.

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., and Murphy, S. (2012). A "SMART" Design for Building Individualized Treatment Sequences. *Annual Review of Clinical Psychology*, 8(1):21–48.

Levin, B. (1999). Hate Crimes. *Journal of Contemporary Criminal Justice*, 15(1):6–21.

Levin, B. and Reitzel, J. D. (2018). *Hate Crimes Rise in U.S. Cities and Counties in Time of Division & Foreign Interference: Compilation of Official Data (38 Jurisdictions)*. Center for the Study of Hate and Extremism; California State University, San Bernardino.

Lim, M. and Hastie, T. (2015). Learning Interactions via Hierarchical Group-Lasso Regularization. *Journal of Computational and Graphical Statistics*, 24(3):627–654.

London Stock Exchange (2018). FTSE Statistics. *Retrieved from https://www. londonstock-exchange.com/statistics/ftse/ftse.htm.*

Lord Ashcroft (2016). EU Referendum Poll. *Retrieved from https://lordashcroftpolls.com/wp-content/uploads/2016/05/LORD-ASHCROFT-POLLS-EU-Referendum-Poll-Summary-May-2016.pdf.*

Manski, C. F. (2004). Statistical Treatment Rules for Heterogeneous Populations. *Econometrica*, 72(4):1221–1246.

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

McDevitt, J., Balboni, J., Garcia, L., and Gu, J. (2001). Consequences for Victims. *American Behavioral Scientist*, 45(4):697–713.

Medoff, H. (1999). Allocation of Time and Hateful Behavior: A Theoretical and Positive Analysis of Hate and Hate Crimes. *The American Journal of Economics and Sociology*, 58:959–973.

Melik, J. (2009). Danone's Yogurt Strategy for Bangladesh. *Retrieved from http://news. bbc.co.uk/1/hi/8100183.stm.*

MetOffice (2018). Temperature, Rainfall and Sunshine Time-Series. *Retrieved from https:// www.metoffice.gov.uk/climate/uk/summaries/actualmonthly.*

Metropolitan Police (2018). Hate crime or Special Crime Dashboard. *Retrieved from https:// www.met.police.uk/sd/stats-and-data/met/hate-crime-dashboard/.*

Mitra, A. and Ray, D. (2014). Implications of an Economic Theory of Conflict: Hindu-Muslim Violence in India. *Journal of Political Economy*, 122(4):719–765.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). Human-Level Control through Deep Reinforcement Learning. *Nature*, 518(7540):529–533.

Monella, L. M. (2018). Are Hate Crimes on the Rise in Italy? *Retrieved from http://www.euronews.com/2018/07/31/are-hate-crimes-on-the-rise-in-italy-*.

Moore, M. and Ramsay, G. (2017). *UK media coverage of the 2016 EU Referendum campaign.* King's College, London.

Müller, K. and Schwarz, C. (2018). Making America Hate Again? Twitter and Hate Crime Under Trump. *SSRN Electronic Journal.*

Murphy, S. A. (2005). An Experimental Design for the Development of Adaptive Treatment Strategies. *Statistics in Medicine*, 24(10):1455–1481.

Office for National Statistics (2016). 2011 Census Aggregate Data. *DOI: http://dx.doi.org/10.5257/census/aggregate-2011-1.*

Office for National Statistics (2017). Lower Super Output Area Mid-Year Population Estimates (supporting information). *Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates.*

Office for National Statistics (2018a). Gross Domestic Product: Q-on-Q4 Growth Rate CVM SA %. *Retrieved from https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyr/ukea#othertimeseries.*

Office for National Statistics (2018b). Unemployment Rate (aged 16 and over, seasonally adjusted). *Retrieved from https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/timeseries/mgsx/lms.*

Olken, B. A. (2015). Promises and Perils of Pre-Analysis Plans. *Journal of Economic Perspectives.*

Reid, S., Taylor, J., and Tibshirani, R. (2017). Post-Selection Point and Interval Estimation of Signal Sizes in Gaussian Samples. *Canadian Journal of Statistics*, 45(2):128–148.

Rinaldo, A., Wasserman, L., G'Sell, M., and Lei, J. (2018). Bootstrapping and Sample Splitting For High-Dimensional, Assumption-Free Inference. *arXiv:1611.05401.*

Rose-Ackerman, S. (1996). Altruism, Nonprofits, and Economic Theory. *Journal of Economic Literature*, 34(2):701–728.

Rose-Ackerman, S. (1997). Altruism, Ideological Entrepreneurs and the Non-Profit Firm. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 8(2):120–134.

Russo, D. and Van Roy, B. (2016). An information-theoretic analysis of thompson sampling. *The Journal of Machine Learning Research*, 17(1):2442–2471.

Sadique, M. (2013). Uniqlo Opens two Stores in Bangladesh. *Retrieved from http://www.bbc.co.uk/news/business-23234872.*

Schneider Electric (2012). *Schneider Electric and Grameen Shakti create Grameen Schneider Electric, a Social Business for Energy Services in Bangladesh [Press Release].* Schneider Electric, Rueil-Malmaison.

Schneider Electric (2014). Schneider Electric Bangladesh. *Retrieved from http://www.schneider-electric.com/site/home/index.cfm/bd/.*

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*, 174(3):737 – 758.

Souganidis, P. E. (1985). Existence of viscosity solutions of hamilton-jacobi equations. *Journal of Differential Equations*, 56(3):345–390.

Souganidis, P. E. (2009). Rates of convergence for monotone approximations of viscosity solutions of fully nonlinear uniformly elliptic pde (viscosity solutions of differential equations and related topics).

Spectator (2017). Hate Crime is Up, but it's not Fair to Blame Brexit. *Retrieved from https://blogs.spectator.co.uk/2017/10/hate-crime-is-up-but-its-not-fair-to-blame-brexit/.*

Stoye, J. (2009). Minimax Regret Treatment Choice with Finite Samples. *Journal of Econometrics*, 151(1):70–81.

Stoye, J. (2012). New Perspectives on Statistical Decisions Under Ambiguity. *Annual Review of Economics*, 4(1):257–282.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction.* MIT press.

Sutton, R. S., McAllester, D. A., Singh, S. P., and Mansour, Y. (2000). Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, pages 1057–1063.

Tetenov, A. (2012). Statistical treatment choice based on asymmetric minimax regret criteria. *Journal of Econometrics*, 166(1):157–165.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, 58:267–288.

Tibshirani, R. J., Taylor, J., Lockhart, R., and Tibshirani, R. (2016). Exact Post-Selection Inference for Sequential Regression Procedures. *Journal of the American Statistical Association*, 111(514):600–620.

Time (2017). Hate Crimes Soared in England and Wales After Brexit. *Retrieved from http://time.com/ 4985332/hate-crime-uk-2017/.*

Uniqlo (2014). Store Locator. *Retrieved from http://www.uniqlo.com/uk/shop/.*

Wager, S. and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, pages 1–15.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation Inference for the General Linear Model. *NeuroImage*, 92:381–397.

Yunus, M. (2007). History of Grameen Danone Foods [Video File]. *Retrieved from http://www.gainhealth.org/videos/mohammad-yunus-history-grameen-danone-foods.*

Yunus, M. (2008). Creating a World without Poverty: Social Business and the Future of Capitalism. *Global Urban Development*, 4(2):16–41.

Zhao, P. and Yu, B. (2006). On Model Selection Consistency of Lasso. *Journal of Machine Learning Research*, 7(Nov):2541–2563.