Hate Crime after the Brexit Vote:

Heterogeneity Analysis based on a Universal Treatment*

Claudio Schilter†

London School of Economics

November 3, 2018

Link to the Newest Version

Job Market Paper

Abstract

I investigate the change in hate crime targeting the victim's race or religion after the Brexit vote. The vote represents a public information shock about society's attitude regarding immigrants. My results reveal a substantial and transitory increase in such crime following the vote. The central 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. Differences in unemployment rates do not significantly contribute to the observed variance. More specifically, parsimonious linear prediction models show the shares of recent immigrants and people with formal qualifications as key predictors of the hate crime increase. My findings are consistent with treating the Brexit vote as an update of expected social sanctions to hate offenders. Issues of multiple hypothesis testing and model selection limit the use of classic methods; therefore I apply and adapt recent machine learning methods as well.

JEL Classification: J15, K42, C31

Keywords: Hate Crime, Heterogeneous Treatment Effect, Racial Violence

*I am deeply indebted to Maitreesh Ghatak, Tom Kirchmaier, and Gharad Bryan for their advice and guidance. I moreover thank Karun Adusumilli, Shan Aman-Rana, Dita Eckardt, Friedrich Geiecke, Ria Ivandic, Felix Koenig, Tatiana Komarova, Stephen Machin, Alan Manning, Rachael Meager, Guy Michaels, Niclas Moneke, Taisuke Otsu, Roger Pegram, and the participants of various LSE seminars for their help and feedback on this project. Any errors are my own.

†c.a.schilter@lse.ac.uk

1 Introduction

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

This paper studies hate crime targeting the victim's race or religion in the context of the United Kingdom European Union membership referendum (hereafter 'Brexit vote'). On June 23, 2016, the UK has decided to leave the EU, defying most polling expectations (e.g. Lord Ashcroft, 2016). The result implied a public information shock that society is more critical towards immigrants than expected. I argue this information was the key aspect of the vote to affect hate crime. Consequently, analyzing hate crime after the Brexit vote is insightful about the role of expected attitudes in society beyond the event itself.

News reports and politicians have associated an upsurge in hate crime with the Brexit vote (e.g. BBC, 2017a, Time, 2017, Al Jazeera, 2017, or Financial Times, 2017). Others, however, dispute the connection between the vote and the rise in hate crime (e.g. Daily Mail, 2016, or Spectator, 2017). Insight regarding underlying mechanisms is clearly sought after. Economic models of domestic conflict, social norms, or crime potentially provide such insight. However, a choice among these models poses an unresolved challenge.

This paper investigates not only whether the Brexit vote led to an increase in racial or religious hate crime, but also where and why this increase has occurred. While determining the precise causal channel behind the increase is infeasible, the temporal structure and descriptive models of the spatial heterogeneity provide an agnostic basis to evaluate mechanisms. The lack of an established theory creates ambiguity regarding the choice of descriptive variables, and in the current case, more than 10^{16} linear models are possible. I address the associated issues of multiple hypothesis testing and model selection by applying and adapting state-of-the-art machine learning based methods. A better understanding of where and why hate crimes occur is fundamental for more effective and efficient policy measures.

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[4] 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 associated with a higher increase in hate crime after the vote. These findings are robust across different methods.

The findings regarding the temporal structure support the argument that the main effect of the Brexit vote on hate crime occurred through information-updating. The individuals' information about society's attitude regarding immigrants was updated when the vote results were announced and plausibly re-updated with passing time or action. Contrariwise, other evident consequences of the Brexit vote can be ruled out to have meaningfully affected hate crime. The transience refutes mechanisms that are based on altered fundamentals, expected or actual, as they did not revert back six weeks after the vote. The most notable is the exchange rate of the pound to other major currencies, which dropped considerably as soon as the vote results were announced, and partially embeds expected future changes in other fundamentals (see e.g. Douch, Edwards & Soegaard, 2018). Furthermore, the lack of a hate crime increase prior to the vote is evidence

[1] Levin & Reitzel (2018), Monella (2018).
[2] See for example Craig-Henderson and Sloan (2003), or McDevitt et al (2001).
[3] See for example Hall (2013), or Gerstenfeld (2017).
[4] A borough is an administrative division. In the context of London and Manchester, the average population is 275000.

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).

The result regarding the relevance of a borough's migrant share points towards an opportunity channel. This is corroborated by the fact that the heterogeneity in the relative increase in hate crime is lower than that regarding the absolute increase. However, other events such as the terror attacks in 2017 led to an escalation in the number of hate crimes in both low-immigration and high-immigration boroughs. This implies that the opportunity for hate crime exists in low-immigration boroughs. Together with the result regarding income and wealth proxies, this calls the sufficiency of opportunity based explanations into question.

The association of income and wealth proxies with a more pronounced hate crime increase is arguably more surprising and concerns in particular the relative increase in hate crime. One mechanism consistent with the finding is the following: Attacks in wealthier regions are more attractive to offenders, for example to force immigrants away from attractive assets or opportunities (Mitra & Ray, 2014). People in richer areas benefit more from immigrants than others, and so social norms are more protective of them (Mayda, 2006). Consequently, a general information update of social norms has a higher effect on hate crime in richer areas. Conversely, opportunity cost theories of hate crime are unfit to explain the spatial heterogeneity implied by this finding.[5]

Another robust finding is that the unemployment share in boroughs with a high post-vote hate crime increase is neither economically nor statistically significantly different. This finding contributes to literature investigating the effect of the unemployment share on hate crime (e.g. Falk, Kuhn & Zweimueller, 2011, Krueger & Pischke, 1997).

Overall, the spatial heterogeneity of the hate crime increase after the Brexit vote is different from that of increases after terror attacks (see also Ivandic, Kirchmaier & Machin, 2018). This is in line with the Brexit vote being an information shock about attitudes in society, while terrorist attacks are information or taste shocks about subgroups of immigrants (see also Albornoz, Bradley & Sonderegger, 2018). In addition, detailed data from Manchester allows for a crude analysis of increased offending versus increased reporting of hate crimes by the victims. This analysis as well as anecdotal evidence suggest increased offending to be more important.

The above findings regarding the spatial heterogeneity were not possible to obtain with standard methods or country-wide aggregated crime data (used e.g. by Devine, 2018). The key methodological challenges are the following: First, the Brexit vote was a unique event where all regions of the UK were treated simultaneously. Second, as aforementioned, there are many potential theories regarding heterogeneous effects of the Brexit vote, which leads to issues of multiple hypothesis testing and model selection. A final challenge is to avoid seed dependency as that the number of observations is finite.[6]

To address these challenges, I use a unique data-set and apply and adapt state-of-the-art methodology. The data-set has two key components: The first is the racial or religious hate crime panel data from Greater London on the borough-month level. This was only recently made publicly available by the Metropolitan Police. The second is confidential high-frequency and geocoded information on racial or religious hate crime from the Greater Manchester Police. Joining the two data-sets on the borough-month level results in panel data of 42 boroughs over 88 months, which I am the first to use. This 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. Qualitatively, the aggregated time series of racial or religious hate crimes in these two

[5]Such theories state that bigots with a low opportunity cost of time are prone to commit hate crimes (e.g. Medoff, 1999).

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

metropolitan areas closely mirrors that of England and Wales.[7] This can be pointed out with time series hate crime data from England and Wales.

In terms of methodology, this is one of currently few papers to apply the recent advances in machine learning to obtain valid inference on heterogeneous treatment effects.[8] After evaluating the magnitude and temporal structure of the 'Brexit effect', I first use an adapted version of the approach of Chernozhukov et al (2018) to analyze the magnitude of spatial heterogeneity captured by the 'candidate variables', i.e. the variables from the census and vote data. I then apply the conditional post selection lasso of Lee et al (2016) and Tibshirani et al (2016) to obtain a linear prediction model of that heterogeneity. Next, I propose an ad hoc splitting 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 testing the individual correlations of the candidate variables with the increase in hate crime, benchmarking the results of the previous methods. To the best of my knowledge, the application of the approach of Chernozhukov et al (2018) and also the use of conditional post selection lasso 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' with standard methods. Relevant for the subsequent heterogeneity analysis on the month-borough level, I find the effect to be predominantly present in July 2016, the month following the vote.

As a first step to evaluate the spatial heterogeneity, I measure the abnormal hate crime in July 2016 as the difference between the observed value and a panel-OLS estimate using borough-specific trends, seasonalities, and means. I then focus on the cross-borough heterogeneity of that measure. A panel-OLS performs well as a predictor, outperforming for example a random forest. Moreover, the results are interpretable as a more conservative version of a standard panel-OLS regression where the treatment (July 2016) is interacted with the selected candidate variables.[9] Using the method proposed by Chernozhukov et al (2018), I show that the spatial heterogeneity captured by the candidate variables is significant. Due to the challenge that every borough is treated simultaneously, I cannot use propensity score matching[10] (as done by Chernozhukov et al, 2018) but instead use the measure obtained in the first step. This measure could potentially be comprised of noise with an arbitrary spatial heterogeneity. Evidence from permutation inference refutes this concern: Compared to results of the same procedure with the other months as placebo treatments, July 2016 is clearly a tail event.

The measure obtained in the first step allows using simple machine learning methods to predict the change in hate crime for each borough which is of direct interest for policing and policy. If the vote and census data is sufficiently explanatory, the spatial heterogeneity in predicted change is more informative than the heterogeneity in the measure itself. While the previous step showed that the measure contains signal, it also contains some noise, whereas a perfect prediction would show the change devoid of noise.[11]

[7] The spike after the Brexit vote might be more pronounced in London and Manchester. The difference is not significant (using permutation inference) and it would be in line with my results (more recent immigrants in London and Manchester).

[8] Other examples include Bertrand et al (2017), or Knaus, Lechner & Strittmatter (2017).

[9] In the standard setup, trends, seasonalities, and means are determined simultaneously with the treatment effect. My proposed setup implies an arguably more conservative question. I proof that the produced estimator is weakly attenuated compared to the standard estimator for the one-variable case, and also show at an example that the differences are small.

[10] All candidate variables are constant (decennial census data). Combined with the fact that every borough is treated simultaneously, this implies the overlap condition, which is necessary for propensity score matching, fails.

[11] In other words, the actual observations are severely over-fitted. See e.g. Hastie, Tibshirani and Friedman (2009).

What remains to be explained is the cross-borough heterogeneity in the single time period July 2016, using a subset of only candidate variables. Consequently, the direct application of the conditional post selection lasso method of Lee et al (2016) and Tibshirani et al (2016) is attainable, despite the fact that it was not designed for analyzing heterogeneous treatment effects. In brief, this method first uses the standard lasso method to select a model.[12] For the variables contained in the chosen model, confidence intervals are created and the lasso parameter estimate is adjusted. This second step addresses the issues of multiple hypothesis testing, model selection, and the lasso's inherent attenuation bias. As a result, a parsimonious linear prediction model of the conditional average treatment effect (CATE) is obtained with valid inference for its parameters.

Moreover, an ad hoc multiple splitting estimation method developed in this paper is used (building on Chernozhukov et al, 2018, Athey & Imbens, 2017, and Rinaldo et al, 2018). It is useful to confirm the linear model obtained by the previous method. The assumptions are different, and while some are stronger, it allows for example to correct for heteroskedasticity which is unattainable in the previous approach. In addition, interactions of candidate variables can be included.[13] The splitting estimation uses lasso on a random half of the sample to obtain a model, and the other half to estimate that model (alike Athey and Imbens, 2016). For a single sample split, this leads to valid inference and overcomes the issues of multiple hypothesis testing and model selection (see e.g. Athey & Imbens, 2017). However, in my setting with a finite number of boroughs, different models are obtained depending on the split. Therefore, the process is repeated 1000 times to overcome this seed-dependency challenge. Unfortunately, aggregating the estimates across iterations results theoretically in ambiguous bias. A series of simulations using this paper's data suggest that this bias is negligible in the current context.

The splitting estimation and the conditional post selection lasso method arrive at highly similar linear prediction models, providing evidence in favor of their validity in this setting. When the splitting 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.

Relation to Literature

This paper contributes to the crime-literature dating back to the seminal paper of Becker (1968). The aforementioned hate crime literature is part of it and broadly, also the literature of (domestic) conflict (e.g. Esteban & Ray, 2011, Esteban, Mayoral & Ray, 2012, Caselli & Coleman, 2013, or Mitra & Ray, 2014). More specifically related, Blair, Blattman and Hartman (2017) use machine learning methods to forecast local violence (namely lasso, random forests, and neural networks). Their setting allows for the predictions to be tested. The lasso is shown to be the best performing method which acted as a motivation for this paper to make use of lasso based methods. Ivandic, Kirchmaier and Machin (2018) use the Manchester part of this paper's data and analyze the effect of terrorist attacks on Islamophobic hate crime (see also Hanes and Machin, 2014). As discussed above, the spatial heterogeneity of the effect of terror attacks on hate crime is different to that of the Brexit vote. The concept of a controversial vote leading to public information

[12] The lasso (least absolute shrinkage and selection operator, see Tibshirani, 1996) is related to the OLS, but imposes a penalty on the magnitude of the estimated parameters such that they are shrunken. In particular, less important parameter estimates are shrunken to zero. Those variables with non-zero coefficients can be interpreted as a data-driven choice of a linear model. The caveat that lasso-based method cannot guarantee the selected variables to be the true data generating process (DGP) due to correlations in the candidate variables is limited. The true variables are likely unobserved and hence the realistic and achieved goal is an insightful descriptive model.

[13] To be more precise, it allows to include interactions and enforce the condition that the non-interacted candidate variables must be included in the chosen model if the interaction is included. This condition makes the interactions interpretable. Technically, this is achieved by using hierarchical lasso (see Bien, Taylor & Tibshirani, 2013; Lim & Hastie, 2015) instead of regular lasso to obtain the model.

about the preferences of society regarding migrants is discussed by Bursztyn, Egorov and Fiorin (2017), with focus on the Trump election. Their laboratory setting allows for a better analysis of the precise mechanisms induced by such an information update. However, it is impossible to develop strong statements pertaining to real life decisions, especially violence. Mueller and Schwarz (2018) analyze anti-Muslim hate crime during the Trump campaign and ascertain the effect of Trump's tweets.

Regarding specifically the effect of the Brexit vote on hate crime, Devine (2018) also finds that the Brexit vote had a significant impact on increasing racial or religious hate crimes. He relies on time series intervention models and does not address the heterogeneity of the effect. In a paper complementary to mine, Albornoz, Bradley & Sonderegger (2018) focus exclusively on the heterogeneity in the hate crime increase with regard to the Brexit vote shares. They develop a detailed model based on Bursztyn, Egorov and Fiorin (2017). Further regarding the Brexit vote, Becker, Fetzer and Novy (2017) analyze the spatial variation in the vote with similar candidate variables as this paper. Other than the role of income proxies, which are positively associated with both the remain vote share and the hate crime increase, the important variables are different. Moreover, 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 of receiving benefits. These measures are not of central importance in my case.

Finally, this paper is related to the literature about finding a (ideally interpretable) model for the conditional average treatment effects (CATE) (see Chernozhukov et al, 2018, for a recent review). In addition to the aforementioned methodological literature, Wager and Athey (2018) also address specifically heterogeneous treatment effects. They use a 'causal forest', but emphasize that the method relies on large samples. Furthermore, possible interpretations following their approach are limited as forests are considerably harder to interpret than lasso results.

The remainder of this paper is structured in the following way: The 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 that effect is measured and shows that it is substantial. The methodology 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.

2 Data-Sets and Summary Statistics

Regarding crimes, panel data from the police forces of the 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.[14] A subset of these incidents classified 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 of racial or religious hate crimes are racial hate crimes.

[14]While the data is very extensive, I do not posses any information about the victims.

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, aggregated time series data for 38 police forces from England and Wales is available from the Home Office (2017).[15] Monthly data is available from April 2013 to August 2017, and daily data from April 2016 to August 2017. Its level of aggregation makes this data unsuitable to use 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.[16] Summary statistics about this data are provided in Table 1, and further visualizations of the aggregate data (including details available for Manchester only) in section 3. Table 1 shows the variance in hate crimes is considerable and already suggests that while there are more hate crimes in July in general, there were even more in July 2016.

The hate crime classification is by definition 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 7.1).[17]

The Brexit vote data on the counting area level is publicly available from the Electoral Commission (2016), and was widely reported in the media the day after the Brexit vote. This data includes all 382 areas 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 counting areas are boroughs. The overall outcome of the Brexit vote is commonly considered as surprising. For example, data is available from a survey in May 2016 by Lord Ashcroft (2016). It 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 highly correlated variables. Taking into account these correlations, and aiming to lose as little information as possible, I have selected 67 variables.[18]

The vote and census data make up the candidate variables I consider in the main analysis regarding the

[15] 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.

[16] 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. the 7 months of data for 2017), the 2016 population data is used.

[17] In terms of additional data, I conducted a brief non-representative survey with four police officers. Three believe that different offender behavior was mostly responsible for the increase in racial hate crime after the Brexit vote. Two of them also indicate that they have also experienced different reporting by victims or witnesses (i.e. more sensitive towards the incident being racially motivated). The fourth officer believes the differential reporting to be mostly responsible for the increase.

[18] 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 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.

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 2. The statistics confirm that there is cross-borough variance in the vote and census data and show to what extent the variables are correlated.

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 question of interest, which is 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, football results, 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':[19] The approaches clearly indicate that racial or religious hate crime has increased after the Brexit vote for a period of approximately six weeks. No permanent effect 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.

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 "Brexit" indicates July 2016 and a clear, unprecedented high. Following a synthetic control approach leads to the qualitatively same result (see appendix), providing further 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 for them I possess the relevant data for the heterogeneity analysis.[20] Figure 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 somewhat more pronounced (which is in line with my results in section 6).

There is a second clear spike in Figure 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, Kirchmaier and Machin (2018) use the same Manchester data-set to analyze the effect of terror on hate crime (and find clear effects, especially strong ones after attacks on English soil).

While the 'Brexit effect' appears rather strikingly overall in the 38 forces, especially in London, it is less pronounced for Manchester. This is a first sign for the importance of heterogeneity. Moreover, at least for

[19] 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 the appendix.

[20] 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, the data was simply added.

Manchester where I possess the data needed to make this distinction, the 'Brexit effect' was almost exclusively driven by racial (but not religious) crimes. The respective figures can be found in the appendix.

In terms of temporal structure, the monthly data in Figure 1 shows an increase in crime numbers mainly in July 2016. Figure 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 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).[21] This average effect is ranked compared to all other possible (overlapping) periods[22] 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).

Comparing the behavior of racial or religious hate crime to some key factors of the environment, no obvious explanation can be found. If anything, Figure 4 shows that July 2016 was a little dryer than other Julys in England, and that immigration of EU job-seekers was at a high just before the Brexit vote.[23] The latter is certainly related to the topic at hand, but the increase was rather gradual, and resembles in no way the clear spike after the vote. More importantly, the increase happened before the vote, and thereafter, there was a decrease that was steady but not immediately dropping. Regarding the aforementioned football, no club-level football was played around the time of the vote but the 2016 European Championships. England dropped out in the round of 16 only four days after the Brexit vote. However, this is highly similar both in terms of date and England's achievement to the 2012 European Championships, and the 2010 and 2014 World Championships (where no relevant crime spikes are observed).

In terms of the types and severity of the crimes that are flagged as racial or religious hate crime, the differences are marginal. As shown in Figure 5 for Manchester (for which I possess the relevant data), if anything, violence against the person is relatively more common, presumably at the expense of criminal damage.[24] In terms of 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.

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 data can be used in a way related to block bootstrapping (see Kirch, 2007).[25] I consider July 2016 as potential short term effect, and the period thereafter as potential long term effect. As the data is on the monthly level, the duration is therefore only

[21] 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.

[22] Periods are blocks of weeks. This visualization is related to the permutation tests discussed in section 3.2.

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

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

[25] 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).

one period (and the remainder of the data respectively) and hence no blocks are necessary.[26] A common use in current economics for permutation tests in the time dimension is in the setting of synthetic control studies, where they are also called 'in-time placebos' (see appendix and Abadie, Diamond & Hainmueller, 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 gives an idea of 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 strong assumption that frequency and distribution of such shocks was constant across the period of data, the p-value can be interpreted as probability of the treatment (in my case the Brexit vote) having a causal effect. In order weaken this 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.

The results are presented in Table 3. There is a significant temporary, but no permanent effect. The point estimate for the temporary effect for London and Manchester is that a total of 550 racial or religious hate crimes occurred due to the Brexit vote, and a relative effect of approximately 21% (a difference-in-difference estimation leads to a highly similar result, see appendix). 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. Classical confidence intervals are not defined in permutation inference, however 'consonance intervals' are (see e.g. Kempthorne & Folks, 1971). Instead of the mean, these take the 95th percentile of the shocks as benchmark and report the size of the significant parameters in that scenario, resulting in a lower bound of the consonance interval.

The effects are qualitatively alike in both data-sets, but the relative magnitude of the short-term effect is larger in London and Manchester. Using the permutation-testing-based consonance interval boundaries, the difference is not significant. Should there be a difference, this is in line with the result of the fraction of recent immigrants being a key predictor (see section 6).

4 Spatial Heterogeneity: Challenge, Measure, and Relevance

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

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 estimating a regression like the following:

$$crime_{at} = \alpha + B_t\gamma + B_t * R_a\beta + T_{at}\delta + A_a + \varepsilon_{at} \tag{1}$$

In the above regression, B_t represents a dummy for the selected period after Brexit vote (July 2016), A_a is the borough fixed effect, R_a stands for the share of remain votes, T_{at} controls for time effects, and $crime_{at}$ is

[26] Blocks were used in the related visualization in Figure 3.

the sum 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}.[27]

As any other candidate variable could have been chosen by the researcher, there is a multiple hypothesis problem. Even absent of any true heterogeneity 5% of all variables will in expectation be significant for explaining heterogeneous treatment effects. 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 is 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 later introduced machine learning based methods. To alleviate the multiple hypothesis problem, I use the FDR correction method by Benjamini and Yekutieli (2001).[28] 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 hate crime increase 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, 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.

[27] The parameter of interest is estimated to be highly significant at 2.85. The term is also economically large: 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 are at 60.7 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 the appendix.

[28] 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 to reject null hypotheses becomes minuscule. Concepts relying on the false discovery rate (FDR) are often regarded as an improvement (see Austin, Dialsingh, & Altman, 2014, for a review). The standard method is the Benjamini and Hochberg (1995) FDR correction. However, this relies on the assumption of 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.

4.2 Obtaining a Measure: Detrending, Deseasonalizing, and Demeaning

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

$$crime_{at} = \alpha + B_t\gamma + T_{at}\delta + A_a + \varepsilon_{at} \tag{2}$$

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 no dimension of heterogeneity is included. Based on section 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.[29] Consequently, the analysis of this measure ($\varepsilon_{a,July16}$) is fully cross sectional.[30] 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 from regression 3:

$$log(crime_{at}) = \alpha^l + B_t\gamma^l + T_{at}\delta^l + A_a + \varepsilon^l_{at} \tag{3}$$

There are several other ways how the abnormal crime could be measured. The main use of this measure is to capture and model the spatial heterogeneity. All these methods build on a cross-borough lasso estimation in July 2016. Using all months but July 2016, all months but June, July, and August 2016, or only the months before July 2016 and then predicting 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.[31] Details can be found in the appendix. 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. The panel-OLS arguably adds reasonable structure to support out of sample predictions. A random forest, for example, performs worse.[32]

A further advantage of the here proposed measure is that the results are interpretable as panel-OLS results using a different (but closely related) dependent variables: 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.[33] 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. The approach is arguably more conservative than the standard setup where the estimators of treatment and control variables are obtained

[29] The number of observations in this cross section (42 overall, in some sense reduced to 21 in the splitting approaches) is not uncommon in the relevant literature. In two papers that are fundamental in 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.

[30] 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 2, and moreover for permutation tests in section 4.4.

[31] A final approach would be not to generate a separate measure but instead use a partially penalized lasso in a version of regression 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 C.P.S.L. method in section 5.1 (circumnavigating the problem via differentially weighted parameter standardizatioin does not provide a satisfactory result). Given this disadvantage, I generally refrain from using this approach. The exception is section 6.3, which includes an illustrative comparison of the this approach with the here proposed measure.

[32] 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).

[33] The census data is constant as the census is only taken every 10 years. The census from 2021 has not yet been conducted, and the census from 2001 is hardly relevant in my setting where the crime data starts in 2008 and 2010 respectively. Finally, Brexit vote data is unique.

simultaneously. This is shown formally for the case of a one-variable model.

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 $y = x'\beta + Z\gamma + \nu$.[34]

Proof: See appendix.

Arguably, the intuition also applies to the multi-variable case. The standard method uses variation in X free of Z in both the numerator and the denominator, i.e. in both regression steps of the Frisch-Waugh-Lovell approach. In the here proposed approach, the full variation of X is used in the denominator. In the current setting, the interaction between a candidate variable and the treatment dummy take the role of x in Lemma 1, and all controls that of Z (i.e. all regressors in regression 2).

The other difference is that 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 6.3, both the here proposed measure as well as a standard regression (alike regression 1) are used. As expected, the differences are small (not significant).

4.3 Mapping the Spatial Heterogeneity

A first use of the measure is to obtain predictions on the borough level. This is directly relevant for policymakers and the police. However, plainly mapping $\varepsilon_{a,July16}$ and $\varepsilon^l_{a,July16}$ of the regressions 2 and 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 also visualize the predicted changes in racial or religious hate crime in July 2016 on maps of Greater London and Greater Manchester. 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. Consequently, these maps can be more relevant.

The results are shown in Figure 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 predicted values, the heterogeneity is considerable. This heterogeneity is now analyzed more formally.

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. Therefore, this section shows the heterogeneity is indeed significant and that it can be mostly attributed to the Brexit vote. This also verifies that the measure contains relevant signal.

For a first illustration, I use a measure analogue to that proposed in section 4.2 but constructed for each month in the data (not only July 2016). Figure 7 depicts the variance of this measure across the 42 boroughs of Greater Manchester and Greater London for each month. 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.

[34] As usual, $M_Z = I - Z(Z'Z)^{-1}Z'$.

To formally analyze the *captured* heterogeneity, I build directly on the approach of Chernozhukov et al (2018) 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, 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 the approach of Chernozhukov et al (2018), using the previously introduced measure that encompasses the 'Brexit effect' but also noise in July 2016. Consequently, Chernozhukov et al (2018) 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 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 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 (2018) 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.[35] Given the focus on lasso-based methods in this paper (consistent with the findings of Blair, Blattman & Hartman, 2017), I choose to use a simple lasso, using 3-fold[36] 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 can be used to conduct the following analyses:

1) Dividing the main sample in K bins according to the predicted treatment. While Chernozhukov et al (2018) use K=5, I have a smaller sample and hence choose to use 3 bins instead. The difference in hate crime between the top and bottom bin reflects the importance of the heterogeneity captured by the candidate variables.[37]

2) Regressing 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.[38]

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 (2018).

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 2 (regression 3 respectively),

[35]Chernozhukov et al (2018) 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 (2018) stress that while using the "best" method is helpful, their approach only requires a useful proxy of a treatment prediction.

[36]3-fold cross validation was chosen since after the split, 21 observations remain, so 3 is a divisor of the sample size. Moreover, this entire section focuses on tercile comparisons, so using 3-fold cross validation seems consistent.

[37]Chernozhukov et al (2018) 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 the appendix. 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 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, Chingos & West, 2018, for more details).

[38]Since measure proposed in section 4.2 is used, the constant should result in a zero estimate. This can be used as an additional check. (Successful in this paper.)

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. Analog to the aforementioned consonance interval, I can compare the estimated heterogeneity to the 90^{th} or 95^{th} percentile of heterogeneities in other months to make a statement about how much of the heterogeneity is due to the Brexit treatment rather than common noise.

The results of the first measure of heterogeneity are shown in Table 4. The absolute heterogeneity is high as expected (from Figures 6 and 7). The top (predicted) tercile has experienced an increase in racial or religious hate crime which is 82 crimes per population-million higher than the bottom terciles (or 14 percentage points respectively). This is considerable. 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. In the relative case, the difference in tercile means is not statistically significant. While this is likely due to a lower heterogeneity, the relatively low number of observations also affects the significance (due to splitting, the estimates obtained in each iteration rely on 21 observations). De facto, there are 88 months times 21 (or overall 42) observations, which is exploited by the permutation inference. It becomes visible that the relative heterogeneity was unusually large compared to the other months in the sample. 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 rows that benchmark the findings using permutation inference make use of the concept of the lower consonance interval boundary (see section 3.2). They indicate how large the heterogeneity induced by the 'Brexit effect' is if the other spatial heterogeneity in July 2016 is the 90^{th} or 95^{th} percentile of the observed values in other months. These other months are then essentially used as placebos. In brief, if the other heterogeneity in July 2016 was within the first 9 deciles, the total observed spatial heterogeneity is mainly due to the 'Brexit effect'. As shown in the appendix, the results regarding the second measure of heterogeneity are highly similar.

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.

5.1 Obtaining Linear Models: Conditional Post-Selection Lasso

The conditional post-selection lasso concept (see Lee et al, 2016 and Tibshirani et al, 2016) was designed for valid inference after model selection (by lasso or lars[39]), but not specifically heterogeneous treatment effects. Due to the plain setup following the proposed measure (see section 4.2), however, it is directly applicable to the problem at hand.

In a first step, the model is selected using a regular lasso.[40] In a second step, the estimates are adjusted and confidence intervals generated by conditioning on the fact that the chosen model was selected in that

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

[40] 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.

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.

The first step implies choosing a penalty term, the hyperparameter commonly labeled λ.[41] I employ repeated 10-fold cross validation (following Kim, 2009),[42] and use the common rule to 'err on the side of parsimony' (Hastie, Tibshirani & Friedman, 2009), i.e. choose the most parsimonious model within one standard error.[43]

A main reason to use repeated cross validation is also the overarching topic of (virtual) seed independency, which is violated by a single 10-fold cross validation. Unlike the methods in section 4.4 and 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 the appendix), 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 in order 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. The resulting estimated models in section 6.1 and 6.2 therefore 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.

A potential concern with the second step of the approach is its theoretical basis on i.i.d. normal errors. The fact that I apply the method to a cross section as well as the findings of Tibshirani et al (2018) that normality is (asymptotically) not required and that simulation results remained unaffected by some heterogeneity make me confident that this approach is suitable for my application. Further details about this approach can be found in Lee et al (2016) and Tibshirani et al (2016).

5.2 Obtaining Linear and Quasi Linear Models: Splitting Based Estimation

This section is inspired by Athey and Imbens (2017), Chernozhukov et al (2018), and Rinaldo et al (2018). In order to avoid any form of harmful (possibly collective) data mining, I propose an approach that is not seed-dependent at the cost of theoretically ambiguous bias. After developing the approach, I show in

[41] 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 that for which the correlation with the dependent variable is the highest. The results of this exercise are displayed in the appendix.

[42] Specifically, I use 1000 repetitions, overall around 10% of possible splits are use which was optimal in simulations by Kuhn & Johnson (2014) (although the differences to using e.g. 100 repetitions are marginal).

[43] While this is standard, it makes little sense to do so in the other lasso applications of this paper: In case of the agnostic inference outlined in 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, adding more parsimony does not seem warranted. In the appendix, I will also list results using the plain minimum of the cross validation (i.e. discarding this rule for parsimony).

simulations the extent of this concern. Within my sample and the results I observe, the bias seems small and the coverage high.

5.2.1 Empirical Approach

This approach combines repeated splitting similar to section 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. As in section 4.4, I employ the lasso in half of the sample to select a model. The inference is then made using an OLS regression on the other half. Intuitively, the lasso selection in one half of the data provides a model to which can be committed in the other half. This is related to the ex-ante commitment through a pre-analysis plan (see e.g. Olken, 2015).

Using OLS for inference means not only following most of the standard economics literature, but also that any form of standard error correction becomes feasible. In fact, the choice of inference is entirely free, so in other applications, other techniques might be more appropriate (e.g. logit). Similarly, the "mining" part is flexible. Given the high frequency of relevant publications in recent months and years, a number of mining procedures are imaginable.[44] Among the rather recent and still state of the art methods, I choose again the arguably most established, which is the standard lasso (on standardized data) in order to obtain a linear model.[45]

Due to the flexibility of this approach, I also allow for a quasi-linear structure with interactions between candidate variables. For this part, I will consider the hierarchical lasso approach of Bien, Taylor and Tibshirani (2013). It guarantees that the levels of each chosen interaction-variable are also included in the model, allowing the result to be interpretable.

The key caveat of the approach is that it only delivers consistent and unbiased estimates when the sample is split randomly once (i.e. the approach is seed-dependent). Since it is impossible for a single researcher, and especially for the profession of social sciences as a whole, to commit ex post to a certain random split, I suggest the following procedure (inspired by Chernozhukov et al, 2018): Split the sample randomly and use the mining and inference sample as described, record the estimates, and then repeat the procedure a total N times (I set N=1000). The most common model (subject to constraints, see below) is selected and the median estimates of those iterations that resulted in the selected model are used.[46] I do not use the p-value doubling of Chernozhukov et al (2018). This is rather conservative and the subsequent simulations do not indicate that it would be required for this method and setting.

While this procedure eliminates seed-dependency, and to some extent the common criticism of sample splitting methods that half of the data is not used in any way for inference, it introduces bias. As all splits influence the choice which splits are generating the final estimates, there is no longer complete independence between the inference sample and the data used for model selection. Rinaldo et al (2018) point out the lack of a theoretical foundation to aggregate different models in the current literature. Therefore, the resulting net bias remains unclear. Intuitively, there are both forces to amplify and attenuate the estimator.[47] It is a

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

[45] In order to pick the hyperparameter λ, I employ 10-fold cross validation, reporting again the 3-fold cross validation results in the appendix. The exception are the computationally intensive hierarchical lasso models that allow for interactions and square terms. For those only the computationally less intensive 3-fold cross validation results are generated.

[46] Contrary to section 4.4, it is not the case that each iteration produces estimates for the same parameters.

[47] On the one hand, there are at least two forces that understate the true effect: Being constrained to the inference samples for which the mining counterpart chose a given model implies that observations that drive this model selection are in the mining part (and hence lacking from the model part). The intuitive consequence is attenuation bias. Moreover, the computed confidence intervals are standard OLS intervals that are based on half of the observations. The fact that other splits of the data resulted in the same variable(s) being correlated with the outcome variable implies that such intervals are too large. They are, de facto,

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 produce insight on that specific issue.

Another use of the simulations is to compare model selection approaches.[48] Longer models have the issue of featuring more 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 bears the problem that there is no obvious choice within those models and estimators. I suggest instead using the most common model subject to a constraint. The constrained choice I suggest is to use models which include at least any W variables. I allow W to take the following values: One, two, the floor of the mean of the model lengths minus one standard deviation, and the respective ceiling. The simulations in the following section will be consulted to make a choice that is appropriate for the specific data, questions, and results at hand.

5.2.2 Simulations

By using the real data-set, I guarantee that the correlation structure of the candidate variables is identical to that in the actual regressions in section 6. In order to have the most relevant correlation structure for the results, an important part of the simulations will consider effects driven by those regressors that are in fact detected in section 6, or such that are highly correlated with them. To a smaller extent, I also include random regressors for artificially constructing is the dependent variable. As shown in the appendix, the residual from the regressions 2 and 3 appear approximately normally distributed (to the extent this can be judged from 42 observations). The dependent variable for the following simulations is therefore constructed 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. 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, and Reid, Taylor & Tibshirani, 2017), as well as the characteristics of the results obtained in section 6.

As shown and described in Table 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 the appendix and also use $y = \epsilon$; $y = X_1 + \epsilon$.

These DGP were chosen for the following reasons: Variable X_1 is chosen rather clearly in section 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 next model follows again the results obtained in section 6. The final model 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 6.[49]

built on more than half of the observations. On the other hand, there are countering forces: A key issue is that those variables are chosen which have the highest correlation due to signal plus noise. The resulting estimates would encompass a high absolute value of signal plus noise. Therefore, such noise that produces amplification bias is prone to be in the chosen estimators.

[48] As for section 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 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 the appendix.

[49] My setting consists of 42 observations and 68 vote and census variables (in total 69 candidate variables if a dummy for

Table 5 summarizes the results of different true models and constraints. While I generally average biases across both models and regressors, I make the distinction between biases of positive and negative parameters.[50] Due to that, column 3 and 4 are informative about the overall bias being potentially attenuating or amplifying. Column 5 indicates how often the "true" parameter of the relevant approximation model is contained in the 95% interval, and column 6 states the average length of the chosen model (excluding the intercept). The mean frequency is the mean across the 1000 simulation-iterations of the frequency of the chosen model in the 1000 splits. This is a measure that will also be available for the results in section 6. As the true model becomes longer, the mean frequency approaches 0.1%. This is problematic for the key argument of seed-independency. If the frequency equals 0.1%, that means that each of the 1000 splits led to a different model choice based on the mining sample. Consequently, the approach is forced to chose a model randomly, reverting back to a random-split approach, which is fully seed-dependent. Since this is fortunately not the case in general in section 6, I do not focus on these simulations (they highlight a limitation of the general applicability of this approach though). The eighth column indicates what fraction of the true DGP is on average encompassed in the selected approximation model. 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. In cases of such parsimonious and linear true DGP, it appears optimal to include the full true DGP. However, due to the violated irrepresentability condition, it is not necessarily found. Moreover, in some cases there can be an informative model so parsimonious that it is shorter than the true DGP. Finally, the last column indicates how often the weakest constraint (to include at least one candidate variable) 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 i.i.d. Normal errors are assumed). The model with the constraint to include at least two variables performs almost identically to that of one variable, with a slightly better performance of the latter. The other constraints that force choices of longer models suffer from mean frequencies that approach 0.1 faster.

In sum, the simulation results are reassuring for the applicability of this section's approach in the current setting. If anything, the confidence intervals seem too conservative. Between the constraints regarding the model selection, that requiring at least one variables seems to have the most desirable attributes.

6 Heterogeneity Models: Results

Both approaches outlined in the previous sections lead to the same main results. Boroughs with a wealthy and/or immigration-heavy population are more affected. The borough-share of recent immigrants (for the

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, Taylor and Tibshirani (2017) outline a somewhat 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 a 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 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.

[50] Positive parameters are such in which the mean of the parameter of 1000 repeated OLS regression of the relevant approximation model with 42 observations randomly drawn from the true data generating process 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.

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 insight, specifically regarding the absolute 'Brexit effect'. That model remains unchanged even with the possibility of interactions.

6.1 Linear Heterogeneity Models

The results of finding a linear (prediction) model regarding absolute crime levels are summarized in Table 6. The first three columns report the result regarding the absolute increase in racial or religious hate crime, and the other columns that regarding the relative increase. The two approaches outlined in the sections 5.1 and 5.2 arrive at highly similar results. As a benchmark, the standard lasso is unsurprisingly considerably lower as the lasso penalty term (λ) leads to attenuation bias.[51] 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 for 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 formal qualifications.[52] Following the previous simulations of the splitting approach, I choose the model-selection constraint to include at least one variable for the splitting method (column 2). This constraint is not binding (the intercept-only model is selected only twice out of 1000 splits).[53] Areas with a high share of immigrants that have arrived within 2 years before participating in the 2011 census are the most affected by the 'Brexit effect'.[54] Further data provides strong evidence that (recent) immigration areas are virtually the same in 2011 and 2016.[55] 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 4 and 5 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 parentheses indicate that the significance disappears once heteroskedasticity robust errors are used.[56] The fact that the lasso agrees with the conditional post selection lasso (C.P.S.L.) in terms of selected variable is guaranteed by construction.

Another benchmark to the two approaches 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. As shown in the appendix, the here proposed splitting approach is stable across different seeds (dropping and/or adding an insignificant variable in case of the full model as only main difference), while the single split approach reaches dramatically different models with different seeds.

[51] 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).

[52] This is the only direct measure of qualifications I include (i.e. no differentiation between different qualifications is included; see footnote 18).

[53] In fact, even the "floor"-constraint was not binding, and hence resulting in the identical model. The "ceiling"-constraint did bind and resulted in 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.

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

[55] 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.

[56] In general, the effect of using HC errors is small, this being the only parameter whose significance is affected by it (including the appendix). 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).

In terms of magnitude, this means that boroughs that have one percentage point more recent immigrants experienced a 'Brexit effect' that was 12 racial or religious hate crimes per million borough-population and month higher. The differences in the fraction of recent immigrants across boroughs; the fraction is 0.7% at the 25$^{\text{th}}$ percentile and at the 75$^{\text{th}}$ it is 3.3%.[57] 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.[58] 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 2 is 52 (see appendix for regression output).

Regarding the result for the relative increase, the chosen model contains only the share of people with no formal qualifications.[59] The constraint to include at least one variable in the splitting method (column 5) was binding. The low variance across boroughs caused the splitting approach to choose the intercept-only model in 225 of the 1000 splits. As outlined in section 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 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). For the share of people without formal qualifications, the coefficient is 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 3 is 20.6% (see appendix for regression output).

The findings are visualized in Figure 8. There are considerably more racial or religious hate crimes in the top tercile of boroughs with respect to share of recent immigrants, and the spike after the Brexit vote is very pronounced. In the bottom tercile, the increase was virtually absent. The subsequent spike around the terror attacks moreover indicates that it is hardly the case that there was a lack of opportunities or victims for a spike to occur. 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 the appendix. The lack of formal qualifications, for example, is highly correlated with indicators of the social grade.

[57] 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.

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

[59] 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 the appendix.

6.2 Quasi Linear Heterogeneity Models

As outlined in section 5.2, I also use the hierarchical lasso method of Bien, Taylor and Tibshirani (2013) in the mining part of the splitting estimation in order to obtain a (prediction) model that is allowed to include well interpretable interactions. As this sets 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 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 7 provide further robustness evidence regarding the absolute model and further reason for cautious interpretation of the relative effect (but support for the share of people with formal qualifications being important).

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 magnitude of the increase in crime after the Brexit vote. It complements the previous results and shows the link to approaches that are more standard. Moreover, it allows to use both the proposed model setup in section 4.2 and the standard approach.

The C.P.S.L. approach with one variable (see appendix) chooses the candidate variable with the highest correlation with the dependent variable. Consequently, when weighted with an appropriate measure of the variance, the same variable as chosen by the C.P.S.L. results as the most important individual variable. Different to the C.P.S.L. approach, it becomes directly visible, which variables are nearly as correlated. It is similar to consulting the pairwise correlations (see appendix) after conducting C.P.S.L.. The disadvantage is the arbitrary choice of a one variable model.

To assess the economic significance and compare the results to the previous findings, I weight 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 will refer to it as tercile span. The estimated coefficient is then multiplied with this tercile span. Table 8 shows the 5 variables with the highest effect as well as in interesting cases where the null hypothesis could not be rejected.[60]. Regarding the p-values, I follow the procedure described in section 4.1 and use Benjamini-Yekutieli (2001) FDR adjustments.

I conduct this approach both following regression 1 (replacing the remain vote share with each candidate variable), and also simply regressing the residual from regression 2 on the candidate variable (and the respective analogue when using logarithmic crime as dependent variable).[61] The former is the standard procedure, while the latter follows the model setup used for all previous results. As expected, differences are small

[60] A complete list of all candidate variables can be found in the appendix.

[61] For the former, the whole panel is used and I cluster on the borough level (which includes robustness to heteroskedasticity) as described in section 4.1. Regarding the latter, I use accordingly heteroskedasticity robust errors, using only the cross section of residuals for July 2016 (see section 4.2).

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 4.2).

The most important variable of the residual based method is unsurprisingly that chosen by the C.P.S.L. method and the magnitudes 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 according results are shown in Table 9.

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 to asses, specify, and reframe various existing theories. The obtained results are also of direct interest for policing and policy.

7.1 Economic Mechanisms

The first result relevant for analyzing economic mechanisms is the temporal structure discussed in section 3. The effect is transitory and absent before the vote. This finding is insightful for evaluating which aspects of the Brexit vote were important for causing the increase in racial or religious hate crime. Several possible aspects can be ruled out, but not that the Brexit vote provided information that society is more critical towards immigrants than expected.

A first candidate is that a change in fundamentals led to an equilibrium with a different level of conflict (see e.g. Esteban & Ray, 2011, Esteban, Mayoral & Ray, 2012, or 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 effected (see e.g. Douch, Edwards & Soegaard, 2018), arguably implying expectations about future trade and other economic factors. However, the data is not in support of 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 one or two months.

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 of England and Wales show increased numbers of hate crime for virtually every day for weeks after the vote (see Figure 2).

Furthermore, the aggressive rhetoric in the media or from campaigning is a third candidate 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.[62]

Mechanisms in which the Brexit vote represents information about the expected opinions or norms in society are generally consistent with the temporal structure. Information about vote results can significantly

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

affect whether people publicly commit xenophobic actions, which is in line with a model of expected social norms (see Bursztyn, Egorov and Fiorin, 2017). Building on this, an example of one possible mechanism is the following: Committing a hate crime has a social cost about which offenders are imperfectly informed. The Brexit vote provided information to potential offenders about the expected social norms. This information lowered the expected social cost of hate crime, resulting in more offenses. Potential offenders receive further updates about the social norms with passing time or especially the action of offending. The transitory nature is therefore not in contrast with the above.

The other key results of my analysis concern the spatial heterogeneity of the 'Brexit effect'. This heterogeneity was high. The effect was virtually absent in places with a low share of recent immigrants or people with formal qualifications. The heterogeneity was lower regarding the relative effect, implying that hate crime increased in absolute numbers especially where it was already common: Areas with a high share of immigrants. The latter finding is in line with an opportunity channel. While I have no access victim data, this finding suggests that immigrants are major victims of the additional hate crime after the Brexit vote, in line with anecdotes in the media (e.g. BBC, 2017b, or Independent, 2016). More such victims represent more opportunities for post-Brexit hate crime. It is unlikely that this is the only channel though.

The role of wealth and income proxies (the share of people with formal qualifications) leading to a higher increase is perhaps more surprising and cannot be explained with the opportunity channel. This finding concerns predominantly the relative increase. Figure 8 shows that pre-vote hate crime levels are comparable in regions with a high and a low share of formally qualified people. However, the share of qualified people is highly correlated with the share of immigrants (correlation: 0.64). The previous mechanism-example can be enriched to take the above findings regarding the heterogeneity into account: First, targeted violence against an opposing group is more lucrative in wealthier areas if the objective is to drive immigrants away to obtain access to economic opportunities (e.g. jobs) or assets (e.g. flats) (see Mitra & Ray, 2014). This is in line with immigration-critical arguments of immigrants taking away jobs or flats. Moreover, economic opportunities and assets are generally more desirable in wealthier areas. Still, as shown in the appendix, controlling for the share of immigrants, there is less hate crime in wealthier areas before the Brexit vote. According to the social norms mechanism, what is hindering offenders are expected social sanctions. Mayda (2006) finds in surveys 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 and therefore, expected social sanctions are more important to prevent hate crime. The Brexit vote then provided an information shock about social norms. This having a larger effect in areas where wealthier and more skilled people live is fully in line with this mechanism.

A related mechanism is to build on the fact that the outcome of the vote was according to pre-vote polls more surprising in remain voting areas (e.g. Lord Ashcroft, 2016). As shown in section 6.3, the remain vote share is individually correlated with the increase in hate crime. A bigger information shock then leads to a larger effect on hate crime (this is modeled in detail by Albornoz, Bradley & Sonderegger, 2018). This is fully in line with the findings of this paper. There are two caveats though. First this mechanism requires that offenders commit the crimes in the same borough they live. Since most hate-crime offenders are not caught, this is difficult to evaluate.[63] Second, the remain-share was not the most important variable in my

[63] 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.

analysis, but rather income and migration proxies. However, they are correlated with the remain vote. The main correlation used by Albornoz, Bradley and Sonderegger (2018) is not direct but via the remain vote (i.e. no data shows directly the correlation between pre-vote beliefs and the increase in hate crime). It remains unclear weather the direct correlation between pre-vote expectations and crime would have been the strongest predictive relationship (or weather it would be weaker). It is certainly a plausible mechanism in light of my results.

The result on spatial heterogeneity can also be used to rule out mechanisms that ex-ante seemed plausible. Given their prominence in the hate crime literature, I discuss the relative depreviation theory and the opportunity cost mechanism.

The relative depreviation theory states that the unemployed (or their children) feel deprived and are consequently more likely to follow extreme ideologies and develop violent predispositions (Falk, Kuhn & Zweimueller, 2011, Siedler, 2011). The unemployment share has consequently been a major variable in previous hate crime analyses (e.g. Falk, Kuhn & Zweimueller, 2011, or 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 generally higher correlations between candidate variables, that of the 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.[64] As with the theory by Albornoz, Bradley and Sonderegger (2018), the key caveat is again that offenders may not commit crimes in the borough they live. Nevertheless, relative depreviation theory is not a suitable model to explain the spatial heterogeneity in the hate crime increase observed 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.

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 was consequently in the focus of some hate crime studies (e.g. Krueger & Pischke, 1997, or 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, as mentioned in section 2, there is the question of changed reporting versus changed offending behavior. In terms of mechanisms, this distinction is important. In sum, evidence supports mainly increased offending. The major reporting issue is arguably that victims could be more sensitive about the fact that they are victims of a hate instead of a regular crime.[65] There are strict guidelines on the side of the police to minimize the influence of reporting and subjectivity. Moreover, for Manchester, crime severity data is available. As shown in Figure 4 in section 3, over 60% of hate crimes have a crime severity score above 15 (out of 20). This is not different in the period after the Brexit vote. Assuming that due to thorough investigation these most severe crimes are always recorded as hate crimes if they are such, this provides evidence against changed reporting. Even when using the statistically insignificant difference, a crude back-

[64] 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 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.

[65] Theoretically, the reporting issue has at least one more dimension, which is the reporting behavior of the police. As shown in Figure 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 not more 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.

of the envelope calculation leads to 87% being due to a change in offending.[66] There is the caveat that this argument is not completely robust to lower severity thresholds.[67] However, assuming that increases in less severe hate crime is necessarily a change in reporting is rather strict. Furthermore, there is also anecdotal evidence in favor of changed offending and not changed reporting.[68]

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 is a public information shock regarding society's preferences about 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 of the winning party or candidate, similar to the Brexit vote, and have arguably also led to increases in hate crime (e.g. the Trump election, see Bursztyn, Egorov & Fiorin, 2017, and 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 crimes 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 tackled better than an unexpected one (see Gerstenfeld, 2017). In the event of a similar event, e.g. another referendum that is tied closely to immigrants (or potentially another group of people), the best predictor for an absolute rise in racial or religious hate crime is the share of recent immigrants and the best predictor for a relative increase is the share of people with formal qualifications;[69] 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, or 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 hate crime increase 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).

[66] Focusing only on the most severe crimes (severity score of 16 or higher), non-severe crimes are around 1 percentage point more common after the Brexit vote (statistically insignificant). A simple regression for Manchester only of total logarithmic hate crime on month dummies, time, time squared, and a July-2016 dummy leads to a point estimate of 8.3% for the July dummy. If the non-severe crimes represent the sensitivity channel (which is a very strong assumption), the offender effect then caused an increase of 0.99*8.3p.p.-1p.p. = 7.2 percentage points or 87%.

[67] With lower severity thresholds, non-severe crimes are around 5 percentage points more common after the Brexit vote. The offender effect then only caused an increase of 0.95*8.3p.p.-5p.p. = 2.9 p.p., or 35% of the effect and the sensitivity story the remaining 65%.

[68] I have surveyed four police officers: Three experienced both increased sensitivity and increased offending, but two of them believe the offending to be more important for the increase in hate crime. The fourth officer has neither experienced nor heard about increased sensitivity, but only increased offending. In addition, there are reports of victim anecdotes (e.g. Independent, 2016). Victims frequently mention that the offender cited the Brexit vote outcome. This supports again a change on the side of the offenders, not the victims.

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

Campaigns to encourage reporting hate crimes seem particularly suitable. Paired with investigating hate crimes thoroughly, this increases the expected legal cost to the offender, 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. The confidence in the police can for example be strengthened, 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 of the offender.

8 Conclusion

The key concept in the criminology literature with respect to hate crime increases 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. On the side of the victims, an equivalent mechanism with regards to reporting instead of committing crimes (as hate crimes) might be at play. In this framework, my results provide insight to narrow down what the trigger caused by the Brexit vote was, where and why people got triggered, and, to a smaller extent, whether it is more consistent with a change of offending or reporting.

Regarding the trigger itself, the key finding is that the increase in racial or religious hate crime lasted for approximately six weeks after the Brexit vote and had no effect before. This temporal structure is neither in line with an 'inflammatory rhetoric' mechanism, nor one of changed fundamentals. There was no detectable effect before the Brexit vote, but the rhetoric against immigrants has started weeks if not months earlier. Moreover, while economic fundamentals, expected or actual, did change after the Brexit vote, they did not revert back within the following six weeks. In line with the temporal structure is a mechanism that treats the Brexit vote as a public information shock about society's attitude regarding immigrants. While this also requires to be counteracted after the vote to mirror the transience of the hate crime effect, there are several possibilities how this could have happened.

The spatial heterogeneity of the increase in hate crime was substantial, hence analyzing where people were triggered is important. I first show that the heterogeneity captured by census and vote data is considerable.

In a next step, linear (prediction) models were obtained. These result in predictors with valid confidence intervals. Both the conditional post-selection lasso and the suggested ad hoc multiple splitting estimation reach 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. A mechanism consistent with the latter is that the information update from the Brexit vote resulted in lower expected social sanctions of hate crime. In wealthier areas, social sanction could be more important to prevent hate offenses, so there is relatively more previously unexploited room for hate crime in places where people with formal qualifications live. These predictions and predictor estimates are also important to guide future, possibly qualitative, research to better understand the effect of the Brexit vote on hate crime, and potentially racial or religious hate crime more generally. Moreover, they are of key importance for policing and preventative measures in case of a similar event.

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

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

This proposed four-step procedure of applying state-of-the-art machine learning methods to a universal unique treatment appears reasonable for this specific setting, in part due to the presented simulation evidence using the respective data. However, at least 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 setup. In this setting, it was possible to analyze the heterogeneity in a cross sectional setup, and use the remaining panel data for detrending, demeaning, and deseasonalizing the dependent variable on the borough level. This was feasible due to the effect being temporary and because including a borough fixed effect was reasonable. If these conditions are satisfied, then step two can be applied. The third step, using the proposed multiple splitting estimation, is (currently) theoretically ambiguous. Future theoretical research would be insightful. Currently, applications rely on simulations. In my simulations, it appeared that the method becomes problematic if many variables carry a strong signal. It seems to perform well with fewer signals, or at least few signals that are strong.

Furthermore, a back-of-the-envelope calculation based on hate crime characteristics as well as anecdotal evidence provide evidence for the 'Brexit effect' to be mainly due to increased offending rather than increased reporting of offenses (as hate crimes).

Overall, this paper provides evidence on an agnostic basis to evaluate mechanisms. While this generates insight and can guide future studies, it does not provide a final answer to the specific mechanisms at play. More, especially qualitative, research is needed.

References

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.

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.*

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.

Austin, S. R., Dialsingh, I., and Altman, N. S. (2014). Multiple Hypothesis Testing: A Review. *mimeo.*

BBC (2017). Rise in Hate Crime in England and Wales. *Retrieved from https://www.bbc.co.uk/news/uk-41648865.*

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.

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.*

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.

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):414–427.

Caselli, F. and Coleman, W. J. (2013). On the Theory of Ethnic Conflict. *Journal of the European Economic Association*, 11(Suppl. 1):161–192.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2018). 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.

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.*

Devine, D. (2018). The UK Referendum on Membership of the European Union as a Trigger Event for Hate Crimes. *SSRN Electronic Journal.*

Douch, M., Edwards, H., and Soegaard, C. (2018). The Trade Effects of the Brexit Announcement Shock. *Warwick Economics Research Papers No: 1176.*

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.

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.

Glaeser, E. L. (1994). Cities, Information, and Economic Growth. *Cityscape*, 1(1):9–47.

Hall, N. (2013). *Hate Crime.* Routledge, New York, 2nd edition.

Hanes, E. and Machin, S. (2014). Hate Crime in the Wake of Terror Attacks. *Journal of Contemporary Criminal Justice*, 30(3):247–267.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* Springer, 2nd edition.

Hebiri, M. and Lederer, J. (2013). How Correlations Influence Lasso Prediction. *IEEE Transactions on Information Theory*, 59(3):1846–1854.

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*.

Ivandic, R., Kirchmaier, T., and Machin, S. (2018). Jihadi Attacks and Local Hate Crime. *mimeo*.

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.

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.

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.

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.

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.londonstockexchange.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.*

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.

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.

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*.

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). Local Area Migration Indicators, UK. *Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/migrationwithintheuk/datasets/localareamigrationindicatorsunitedkingdom.*

Office for National Statistics (2018c). 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.*

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.

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/.*

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, 58:267–288.

Tibshirani, R. J., Rinaldo, A., Tibshirani, R., and Wasserman, L. (2018). Uniform Asymptotic Inference and the Bootstrap after Model Selection. *The Annals of Statistics*, 46(3):1255–1287.

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/.*

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.

Zhao, P. and Yu, B. (2006). On Model Selection Consistency of Lasso. *Journal of Machine Learning Research*, 7(Nov):2541–2563.

List of Tables

	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: Racial or Religious Hate Crime Summary Statistics

	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 2: Candidate Variables Summary Statistics

	(1)	(2)	(3)	(4)
	RR Hate Crime	Log(RR Hate Crime)	RR Hate Crime	Log(RR Hate Crime)
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 bound of significant parameters in parentheses (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 on the 1% level, and none of the post July 2016 dummies are significant.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Overall Effect of the Brexit Vote on Racial or Religious Hate Crime Considerable but Transitory

	(1)	(2)	(3)	(4)
	Hate Crimes per Population Million		log(Hate Crimes per Population Million)	
	Top Tercile	Bottom Tercile	Top Tercile	Bottom Tercile
July 2016 Mean	91.8***	9.8	0.278***	0.142
Tercile Difference	81.99***		0.136	
Permutation Significance	**		**	
Permutation 90% Benchmark	76.30 [93%]		0.12 [86%]	
Permutation 95% Benchmark	69.50 [85%]		0.09 [67%]	
Candidate Variables	69		69	
Observations	42		42	
Placebos (Permut. Test)	87		86	

Note: 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 (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). 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 4: Captured Heterogeneity Large and Mainly Attributable to the Brexit Vote

True DGP	Constraint	Bias: Pos Param.	Bias: Neg Param.	Param. in CI	Length	Frequency	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 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; 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 6 are robust to either case.

Table 5: Simulation Results of the Splitting Method: Bias Small, Coverage High

	(1)	(2)	(3)	(4)	(5)	(6)
	Hate Crimes per Population Million			log(Hate Crimes per Population Million)		
Recent Immigrants	1237***	1150*	977			
No Relig. Stated	641***	715$^{(**)}$	405			
No Qualifications	-235	-274	-114	-2.55*	-2.67***	-0.37
Method	C.P.S.L.	Splitting	Lasso	C.P.S.L.	Splitting	Lasso
Frequency	NA	1.6%	NA	NA	2.0%	NA
Candidate Var.	69	68	69	69	68	69
Observations	42	42	42	42	42	42

Note: Method-chosen models from 68(69) candidate variables. Cross sectional analysis across 42 boroughs in July 2016 (the month after Brexit). Dependent variable is detrended, deseasonalized, and demeaned on 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 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. Mean log(hate crimes per borough population million) in July 2016: 4.567. C.P.S.L assumes i.i.d. errors by construction. Splitting allows for robust errors, parentheses indicate lost significance due to using heteroskedasticity robust erros. 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 C.P.S.L. and Splitting)

Table 6: Best Linear Model for the Absolute/Relative Increase in Hate Crime

	(1)	(2)	(3)	(4)
	Hate Crimes per Pop. Mio.		log(Hate Crimes per Pop. Mio.)	
Recent Immigrants	1426*	1309*		
No Religion Stated	595	641		
No Qualifications	-214	-211	20.69	3.50
Male			8.64	-2.46
Male * No Qualifications			-48.47	-6.14***
Square Terms Allowed	Yes	No	Yes	No
Frequency	0.8%	1.0%	1.2%	1.1%
Candidate Variables	68	68	68	68
Observations	42	42	42	42

Note: Method-chosen models from 68 candidate variables, their first order interactions, and, where indicated, their squared values. Cross sectional analysis across 42 boroughs in July 2016 (the month after Brexit). Dependent variable is detrended, deseasonalized, and demeaned on 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 share of people not stating a religion: 0.080. Mean share of people with no formal qualifications: 0.194. Mean share of males: 0.493. Mean hate crimes per borough population million in July 2016: 219. Mean log(hate crimes per borough population million) in July 2016: 4.567. Heteroskedasticity robust erros used.
* p < 0.1, ** p < 0.05, *** p < 0.01

Table 7: Best Quasi-Linear Model with Interactions using the Splitting Method

Variable	Estimate * Tercile Span	Variable	Estimate * Tercile Span
Recent Immigrants	85.0***	Arrived 2 Years	99.1***
Social Grade C2[70]	-77.8***	Social Grade C2	-90.8***
Mixed Ethnicity	77.1***	Mixed Ethnicity	90.0***
Industry Code E[71]	-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		*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 in the lowest tercile. Using residuals denotes a cross sectional analysis accross 42 boroughs in July 2016 (the month after Brexit) with the dependent variable detrended, deseasonalized, and demeaned on the borough level using 88 months of data. Full regression denotes 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 8: Top 5 Individually Important Variables & Unemployment (Absolute Increase in Hate Crime)

Variable	Estimate * Tercile Span	Variable	Estimate * Tercile Span
No Qualifications	-0.29**	No Qualifications	-0.34**
Industry Code E[72]	-0.27	Industry Code E	-0.32
Industry Code J[73]	0.27*	Industry Code J	0.31*
Recent Immigrants	0.25*	Arrived 2 Years	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		*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 in the lowest tercile. Using residuals denotes a cross sectional analysis accross 42 boroughs in July 2016 (the month after Brexit) with the dependent variable detrended, deseasonalized, and demeaned on the borough level using 88 months of data. Full regression denotes 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 9: Top 5 Individually Important Variables & Unemployment (Relative Increase in Hate Crime)

[70] Skilled working class: Main income from skilled manual work
[71] Water supply, sewerage, waste management and remediation activities
[72] Water supply, sewerage, waste management and remediation activities
[73] Information and communication

List of Figures

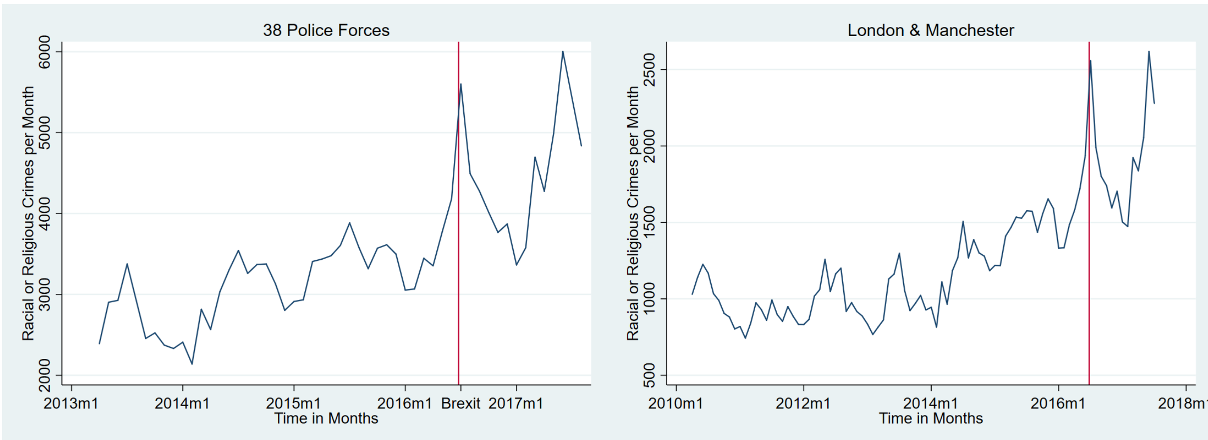


Figure 1: Racial or Religious Hate Crimes over Time: England and Wales vs London and Manchester

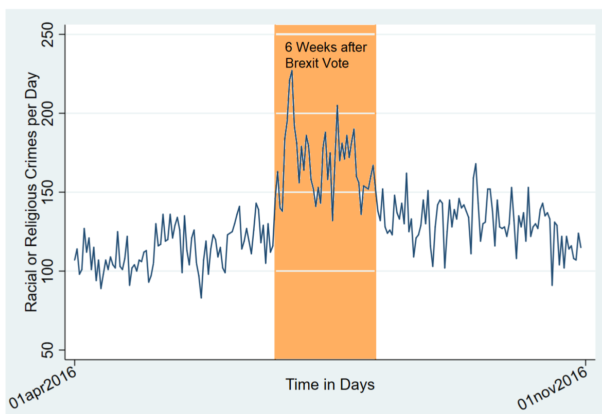


Figure 2: Concentration 6 Weeks after the Vote and No Pre-Vote Effect Visible (38 Police Forces)

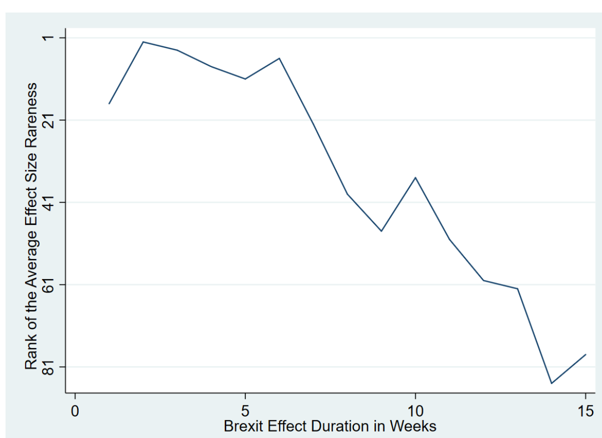
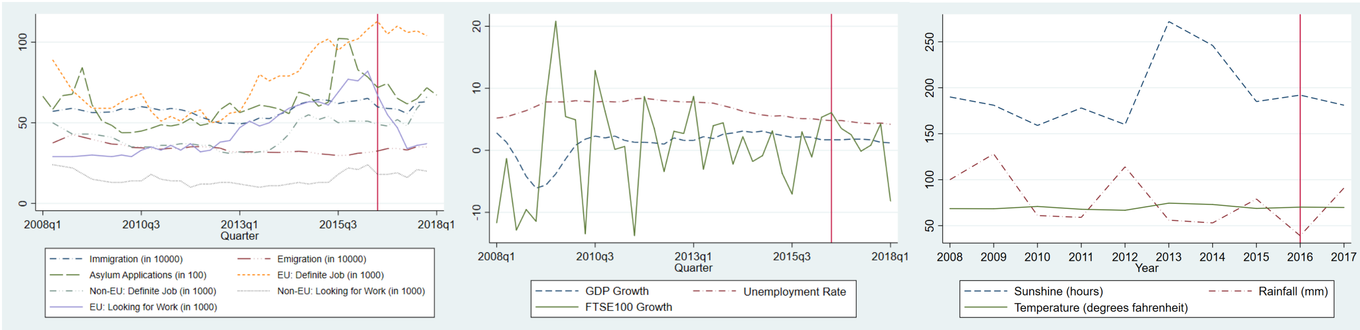


Figure 3: Rarity of Spike Depending on Duration Confirms 6 Week Effect-Duration (Manchester)



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

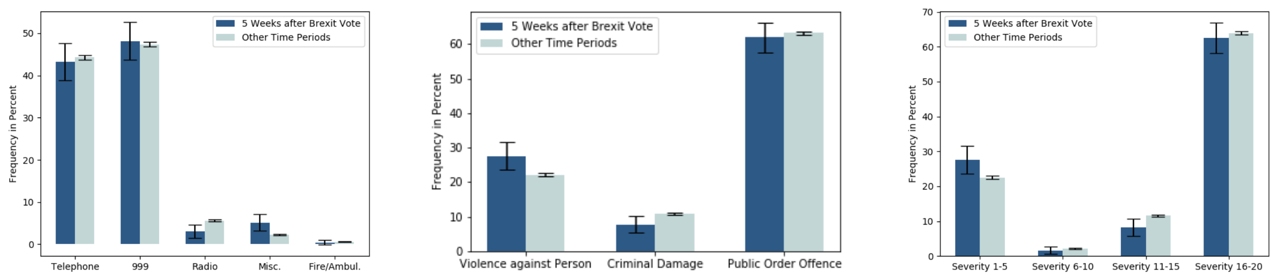
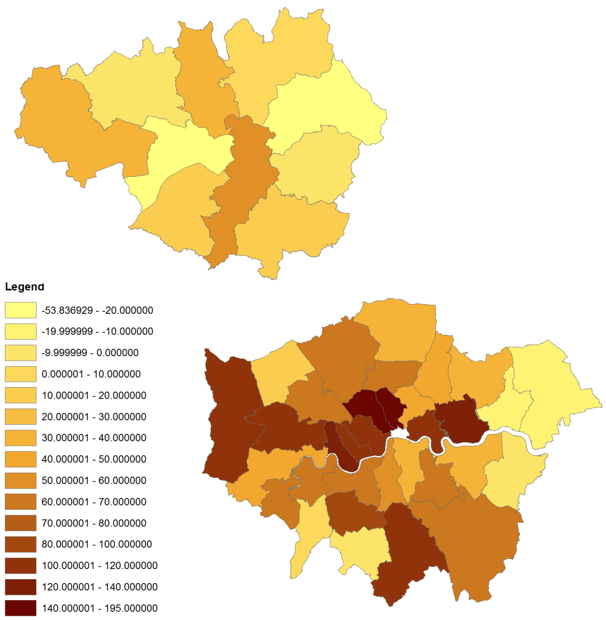
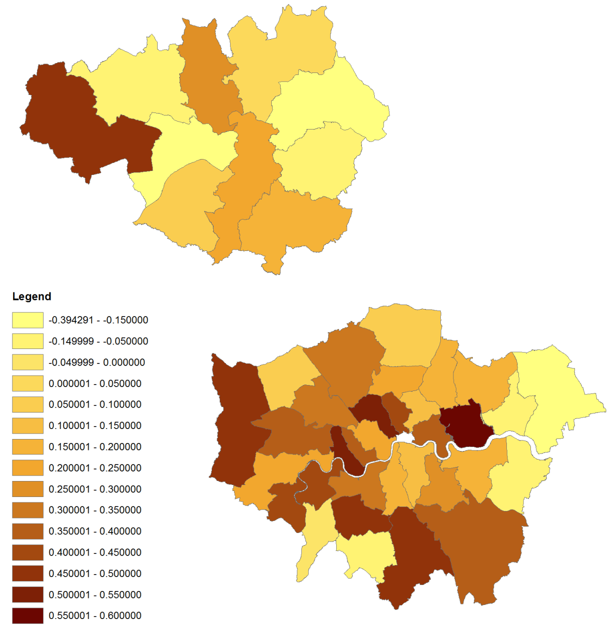


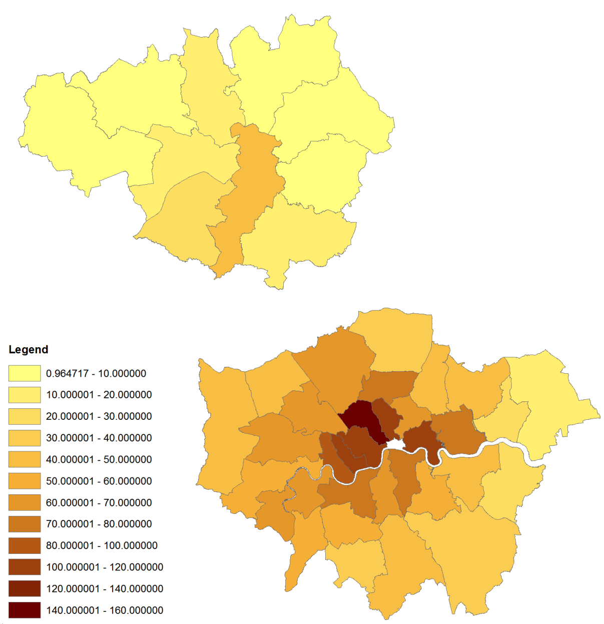
Figure 5: Only Small Differences in Method of Call, Crime Types, and Severity (Manchester)



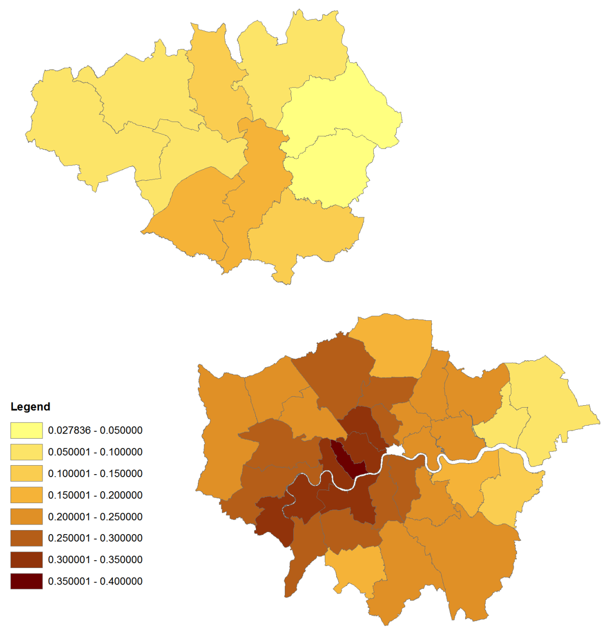
(a) Observed Absolute Hate Crime Increase



(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 6: Borough-Level Hate Crime Increase in July 2016

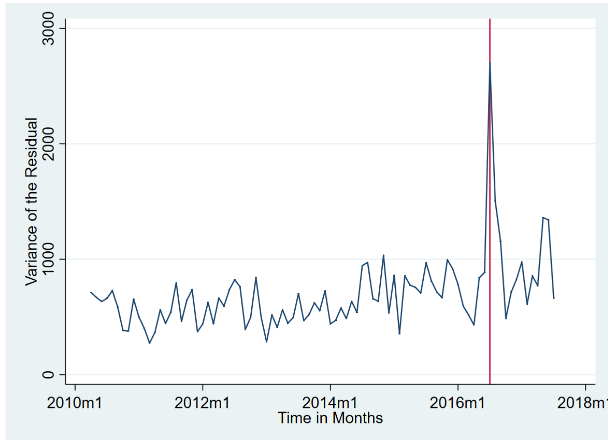


Figure 7: Variance in Detrended Deseasonalized Racial or Religious Hate Crime across Boroughs

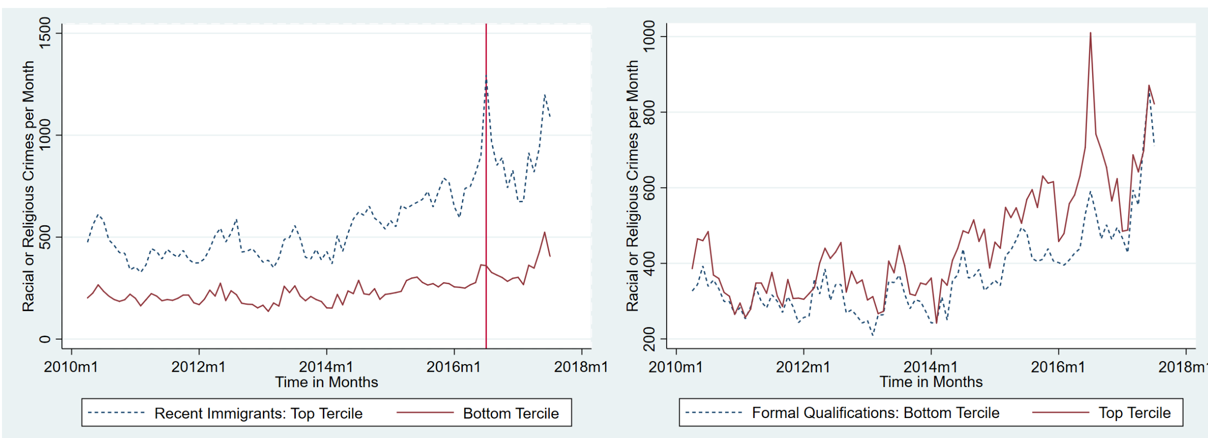


Figure 8: Heterogeneity in the Brexit Effect - Terror Spikes in 2017 Different from Brexit

Appendix

Synthetic Control Method

The synthetic control method has been introduced in the seminal papers of Abadie and coauthors (Abadie & Gardeazabal, 2003; Abadie, Diamond & Hainmueller, 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 the racial or religious hate crime. For the donor pool, I require time series of variables that are (1) affected similarly by factors that is 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.[74] 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 e.g. 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 behaviors 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.[75] That way, each entity in the panel has a generally unique predictor variables (its respective averages).

Result

The synthetic control result is fully consistent with the visual inspection in section 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 Appendix Table 1 below. The result of using the synthetic control

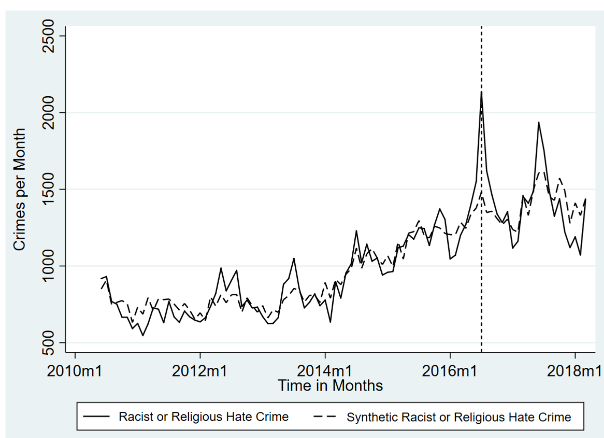
[74] Example: For Manchester, where the data is on the incident-level, I see that at most 21.5% of any given crime-type (public order offense) used in the donor pool is flagges as racial hate crime. That is an outlier in that respect, the second most being violence without injury at 3.2%. Given that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

[75] The average of the given year or calendar month was calculated for each observation based on all other observations, although using all observations lead to almost exactly the same results.

method with this panel data is illustrated in Appendix Figure 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 Offence	0.001

Appendix Table 1: Synthetic Control Weight (London)

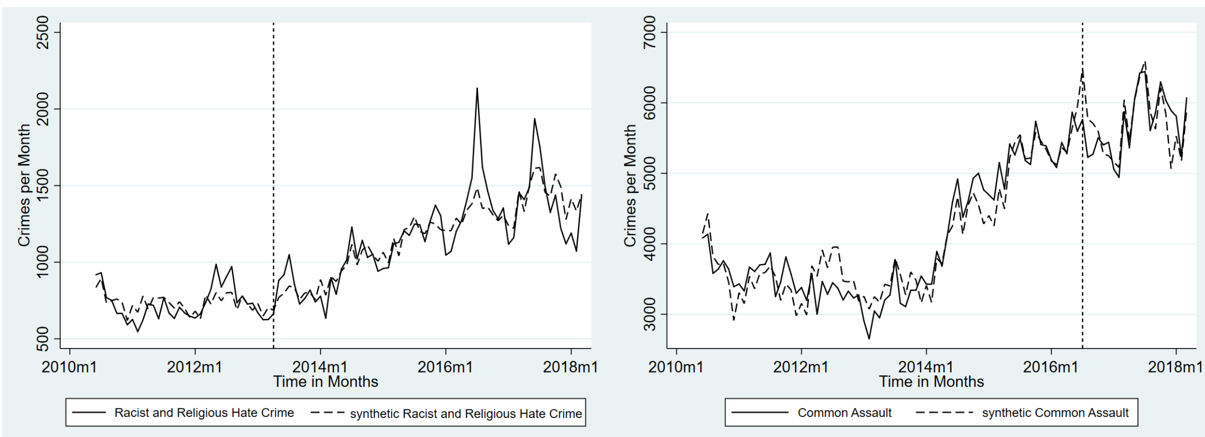


Appendix Figure 1: Synthetic Control (London)

What is rather surprising is that rape is such an important part of the synthetic construct. The weights are generally quite different from Figure 5 and footnote 24, which served (other than data availability) as guide which crime types to include. However, as shown in Appendix Figure 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 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 Appendix Figure 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 Appendix Figure 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).



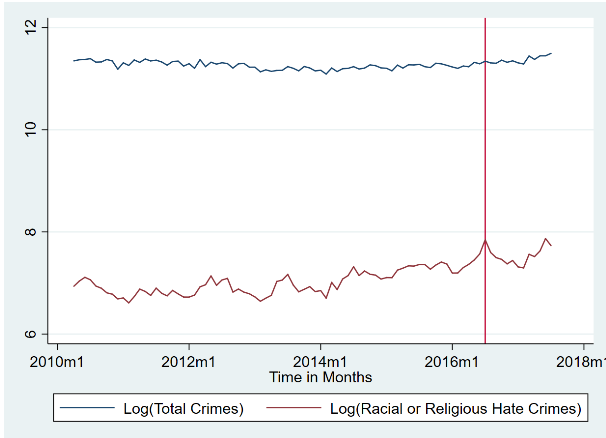
Appendix Figure 2: London: In-Time and In-Crime Synthetic Control Placebos

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.

Column 1 of Appendix Table 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 Appendix Figure 3 though, this result is likely misleading. Hate crime seems to have a slightly different trend and more pronounced seasonal effects than total crime.

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 0.25 for hate crime, highly similar to the 0.21 obtained in section 3.2.



Appendix Figure 3: Trends

	(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

Appendix Table 2: Overall Effect of Brexit: Difference-in-Difference Estimation

The standard errors are not adjusted in any way in Appendix Table 2 and consequently invalid (see e.g. Cameron, Gelbach & Miller, 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 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 Appendix Table 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 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 Appendix Figure 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, Gelbach and Miller (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

Appendix Table 3: Overall Effect fo Brexit: Difference-in-Difference Estimation

Proof of Lemma 1

Take the regression $y = X\beta + Z\gamma + \nu$ (X is a vector, but shown as capital letter to emphasize the intuition beyond the here considered single variable case). 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.[76] Since β is a scalar, $|\hat{\beta}_c| \geq |\hat{\beta}_p|$.

q.e.d.

[76]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 $X'X \geq X'M_Z X$.

Alternative Measures for the Abnormal Hate Crime in July 2016

	(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)
Recent Immigrants	16.3		19.0		19.7		15.7		16.5		17.3	0.00	0.05	
No Religion Stated	11.8		13.8		13.5		11.5		10.1		6.2			
No Qualifications	-5.8	-0.02	-6.7	-0.02	-6.9	-0.03			-4.0	-0.01		-0.00		
Born Old EU													0.07	
Method	As used in this paper and described in section 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 4.2		Using lags instead of time squared		Using partly penalized full regression (alike regression 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 on the borough level. A lack of any entries means that a constant is chosen by the lasso as prediction model. Despite the fact that no direct wealth or income proxy is chosen by the full regression method, section 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.

Appendix Table 4: Lasso Models of Different Measures

Alternative Measure for the Captured Heterogeneity

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

Appendix Table 5: Captured Heterogeneity: Alternative Measure

Binning

Another question for which the approach following Chernozhukov et al (2018) provides insight is that regarding the significant differences in candidate variables. Alike the previous section 6.3 but contrary to the following sections 6.1 and 6.2, no choice between correlated (groups of) variables is made. All significant differences are reported. Due to the rather conservative p-value adujstment (see section 4.4), I select the significance threshold to report variables at 10%. Those with the largest difference are listed in Table 6, the complete list can be found below.

Variable	Difference	Variable	Difference	Variable	Difference
Born in the UK	-27**	Born in 2000-EU	143%*	Born in 2000-EU	87%*
Remain Vote	25**	Recent Immigrants	120%**	Industry Code E	-86%**
2 Bedrooms or fewer	19*	Same Sex Marriage	104%*	Industry Code J[78]	84%**
Christian	-16*	Industry Code E[77]	-102%**	Social Grade C2	-83%**
Socal Grade AB	15*	Buddhist	90%**	Remain Vote	83%**
Percentage Points		*Percent of the Mean*		*Percent of Tercile Span*	

Appendix Table 6: Statistically Significantly Different Candidate Variables regarding the Absolute Effect: Top 5

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 rather wealthy areas with a large fraction of immigrants. Finally, the small differences in Table 6 between actual and the minimum differences emphasize that if the noise in July 2016 was not stronger than in every other month, nearly all of these effects are driven by the 'Brexit effect'.

[77] Water supply, sewerage, waste management and remediation
[78] Information and communication

Interestingly, after conducting the FDR adjustment, none of the candidate variables are significant in characterizing boroughs with the largest relative effect.

Comparing these results to those in section 6.3, there are clearly similarities. This shows that the simple mechanism there is already rather informtive, even for characterizing boroughs by effect size. There are some small but interesting differences. Boroughs with a high share of immigrants from the early EU member states, for example, seems to be a key characteristic of much affected boroughs (in terms of relative to its tercile span, and even unconditional on the significance), while there are many candidate variables with a higher individual corrlation. 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 mean drives them further up. Repeating the procedure in the previous section but with this splitting and median selection (using identical splits), the number of significant variables also decreases sharply. However, the listed top-5 variables remain the same, so a comparison of Table 6 with Table 8 is not driven by this fact (what is more likely driven by it is the absence of any significant regressors in this section regarding the relative effect).

Variable	Difference
Born in the UK	-26.50**
Remain Vote	25.28**
2 Bedrooms or fewer	19.07*
Christian	-16.02*
Socal Grade AB	15.27*
No Qualifications	-9.17**
Social Grade C2	-9.03**
Industry Code M	7.63*
Industry Code G	-5.21**
Born in 2000-EU	5.09*
Age > 64	-5.01*
Industry Code J	4.68**
Disabled	-4.56*
Ethnicity other Asian	4.00*
Industry Code F	-3.71*
Recent Immigrants	2.88**
Providing Unpaid Care	-2.60**
Industry Code R,S,T,U,Other	2.58**
Mixed Ethnicity	2.29**
Buddhist	0.76**
Industry Code E	-0.47**
Same Sex Marriage	0.38*
Born Rest World	0.01*

Absolute Effect, Percentage Points

Appendix Table 7: Complete List of Statistically Different Variables in Most vs Least Affected Terciles

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.

Appendix Table 8: Simulation Results Using a Single Variable: Bias Small, Coverage High

Table 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 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.[79]

The second column of Table 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"[80] 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 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 on 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 variable are almost equivalently informative predictors.

[79] 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.

[80] True in the sense of its predictive property (i.e. including omitted variable bias).

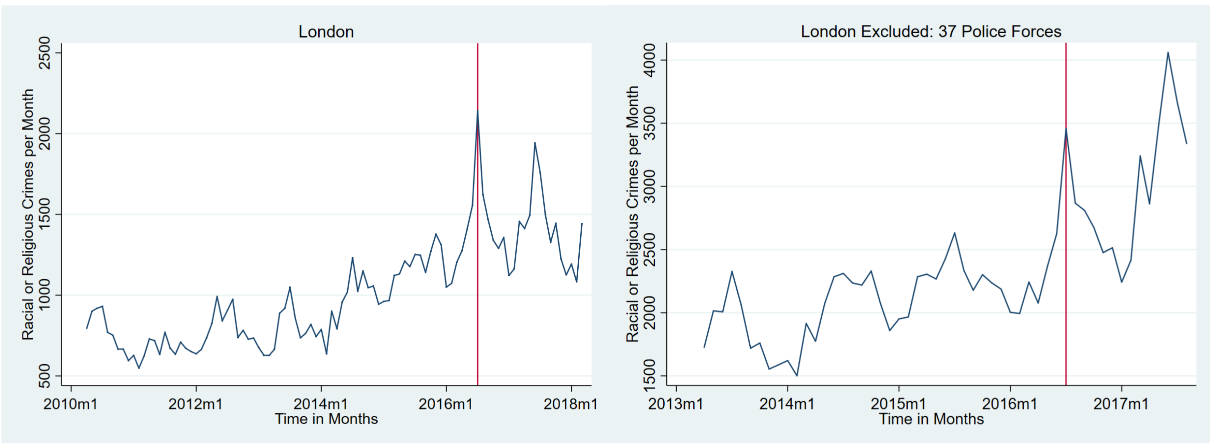
Further Results: Best Single Variable

	(1)	(2)	(3)	(4)	(5)	(6)
	Hate Crimes per Population Million			log(Hate Crimes per Population Million)		
Recent Immigrants	2308***	2321***	390			
No Qualifications				-2.55**	-2.47***	-0.79
Method	C.P.S.L.	Splitting	Lasso	C.P.S.L.	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

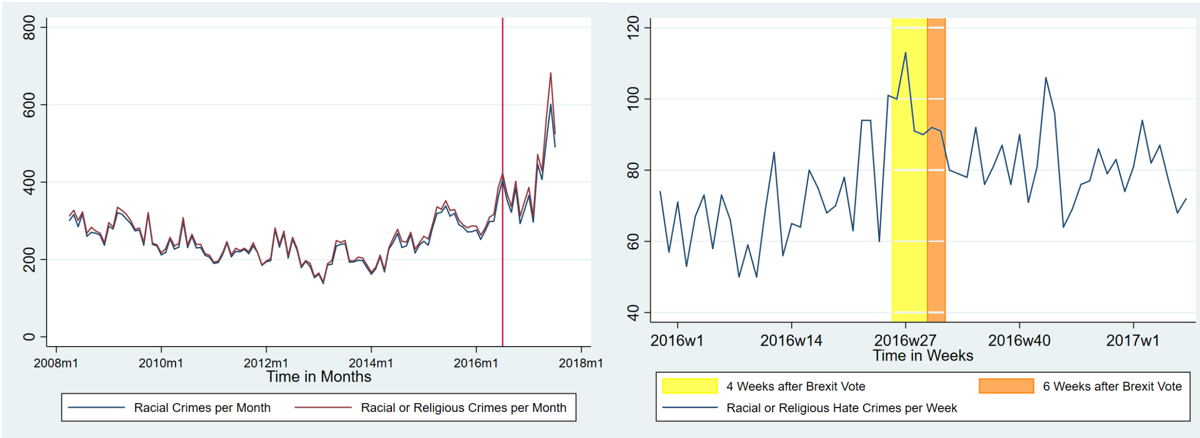
Note: Method-chosen models from 69 candidate variables. Cross sectional analysis across 42 boroughs in July 2016 (the month after Brexit). Dependent variable is detrended, deseasonalized, and demeaned on the borough level using 88 months of data. Mean share of people that have arrived in the UK within 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 erros are used. Significance not defined for plain lasso (which serves as benchmark only).
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (for C.P.S.L. and Splitting)

Appendix Table 9: Best Single Variable for the Absolute/Relative Increase in Hate Crime

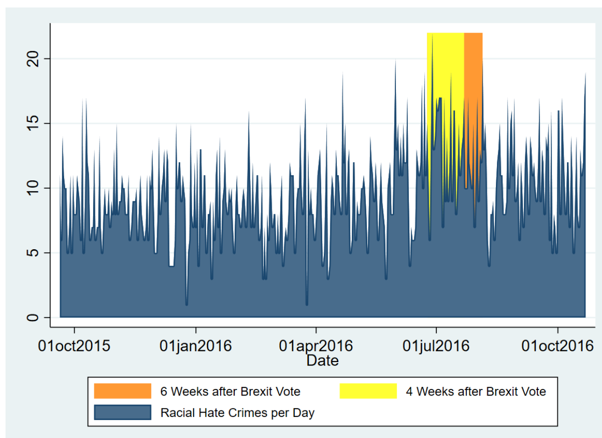
Further Time Series Visualizations



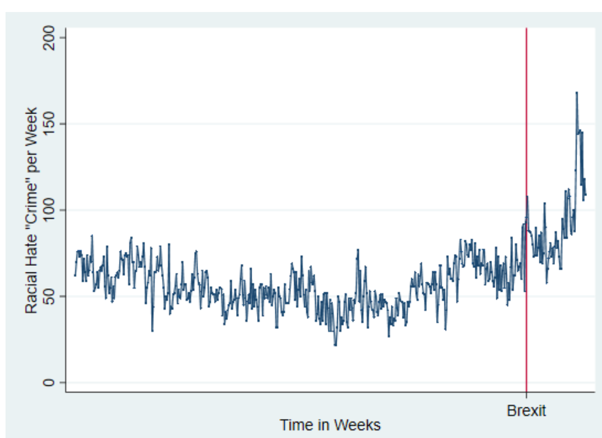
Appendix Figure 4: Racial or Religious Hate Crime over Time: London vs the other 37 Forces



Appendix Figure 5: Racial Hate Crimes over Time: Manchester



Appendix Figure 6: Racial Hate Crimes over Time: Daily Data for Manchester



Appendix Figure 7: Racial Hate Crimes over Time: Manchester Complete Weekly Data

Further Regression Tables

	(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*

Appendix Table 10: Regressing Racial or Religious Crime on the Sole Heterogeneity Dimension Remain 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*

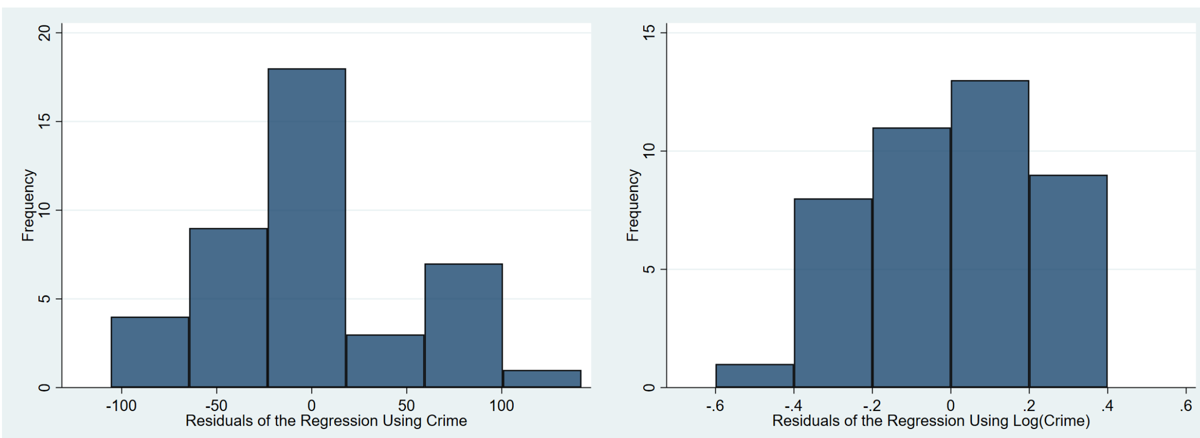
Appendix Table 11: Regression Equations 2 and 3

	(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*

Appendix Table 12: Regrssion on the Pre-Brexit Period Using Heterogeneity Variables instead of Fixed Effects

Residual Distributions



Appendix Figure 8: Sample Distribution of July 2016-Residuals of Regression 2 and 3

Seed Dependency Examples: Repeated Splitting vs Single Split

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Recent Immigrants	2321***	2273***	2310***	2261***	2492***		2790***	
No Religion Stated								2401***
Social Grade 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
Candidate Var.	68	68	68	68	68	68	68	68
Observations	42	42	42	42	42	42	42	42

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

Appendix Table 13: 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*

Appendix Table 14: Seed Dependency Issue: Full Model, Levels

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No Qualifications	-2.47***	-2.38***	-2.41***	-2.40***				
Industry D					-57.63**			
Remain Vote						0.77*	0.69*	
No Religion Stated								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
Candidate Var.	68	68	68	68	68	68	68	68
Observations	42	42	42	42	42	42	42	42

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

Appendix Table 15: 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

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

Appendix Table 16: Seed Dependency Issue: Full Model, Log

Results using Different Forms of Cross Validation

	(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	C.P.S.L.	Lasso	C.P.S.L.	Lasso	C.P.S.L.	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 C.P.S.L. and Splitting)*

Appendix Table 17: Results Regarding Absolute Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Recent Immigrants						2.02	1.09		
No Qualifications	-1.97	-2.26*	-2.51***	-2.55*	-0.34	-1.76**	-1.31	-2.18	-1.67
Industry Code D		-25.11				-22.17	-8.53	-11.17	-13.65
Industry Code E	-3.22								
Male								3.36	2.40
No Relig. Stated								1.65	0.47
No Religion								-0.88	-0.04
Method	Splitting	Splitting	Splitting	C.P.S.L.	Lasso	C.P.S.L.	Lasso	C.P.S.L.	Lasso
Folds	3	10	3	3	3	3	3	10	10
Parsimony 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 Rule	Yes	Yes	No	Never	Never	Never	Never	Never	Never
Candidate Var.	68	68	68	69	69	69	69	69	69
Observations	42	42	42	42	42	42	42	42	42

** p < 0.1, ** p < 0.05, *** p < 0.01 (for C.P.S.L. and Splitting)*

Appendix Table 18: Results Regarding Relative Effects

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[81]	0.89
Born in a 2000-EU State	0.84	Disabled	0.87
2 Bedrooms or Fewer	0.81	Industry J[82]	-0.86
Provides Unpaid Care	-0.81	Social Grade DE	0.83
		Social Grade C2	0.83
		Industry O[83]	-0.82

Appendix Table 19: Candidate Variables with an Absolute Correlation to the Model Variables Larger than 0.8 (None for No Religion Stated)

[81] Water supply, sewerage, waste management and remediation
[82] Information and communication
[83] Public administration and defence

Individually Interacted Candidate Variables

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

Appendix Table 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

Appendix Table 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

Appendix Table 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

Appendix Table 23: Relative Effect Using Residuals