sociological science

Supplement to:

Nägel, Christof, Mathijs Kros, and Ryan Davenport. 2024. "Three Lions or Three Scapegoats: Racial Hate Crime in the Wake of the Euro 2020 Final in London" Sociological Science 11: 579-599.

APPENDIX A1

A1. RDD Assumption checks

A1.1. Effects on placebo outcomes

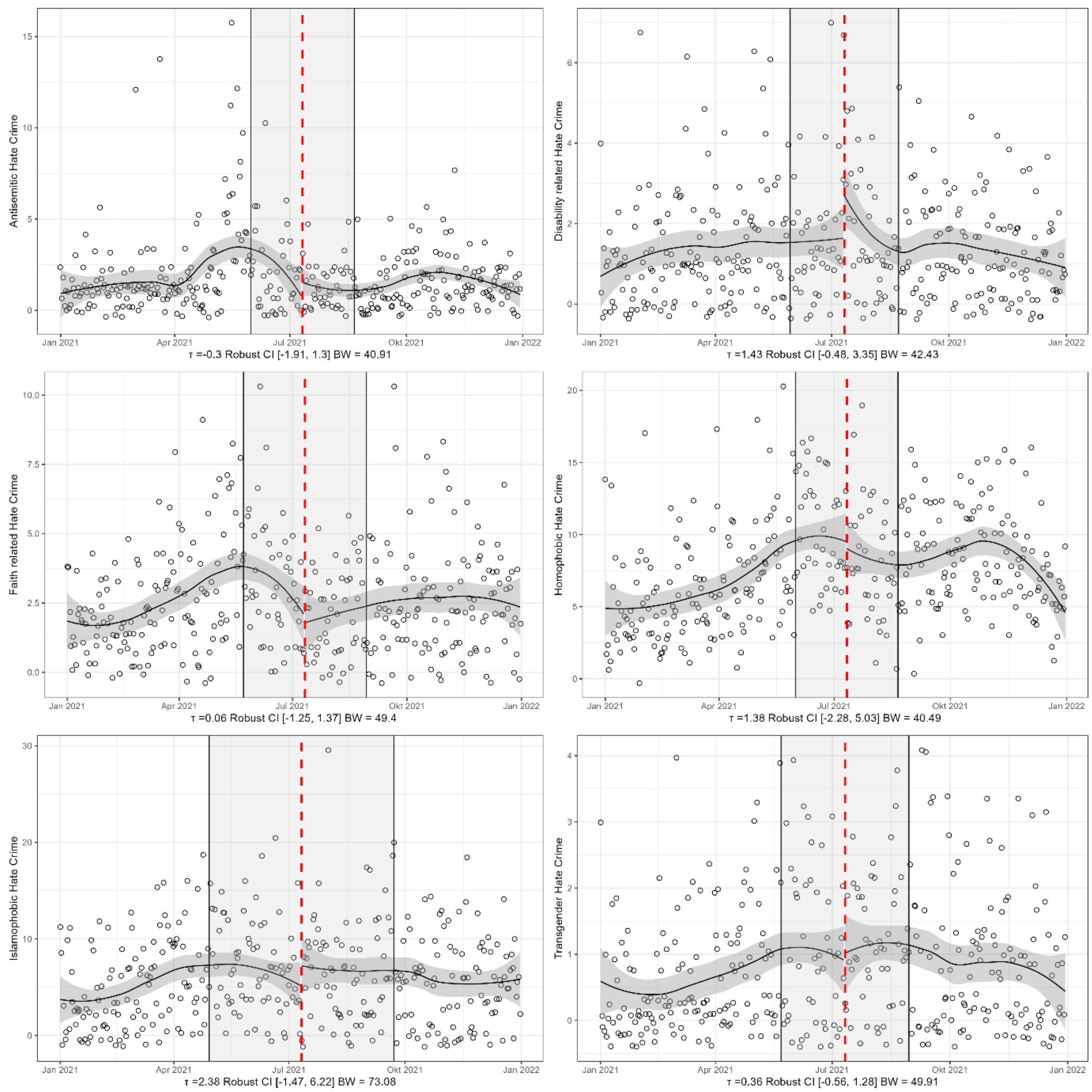

Figure A1. Placebo effects on other hate crimes

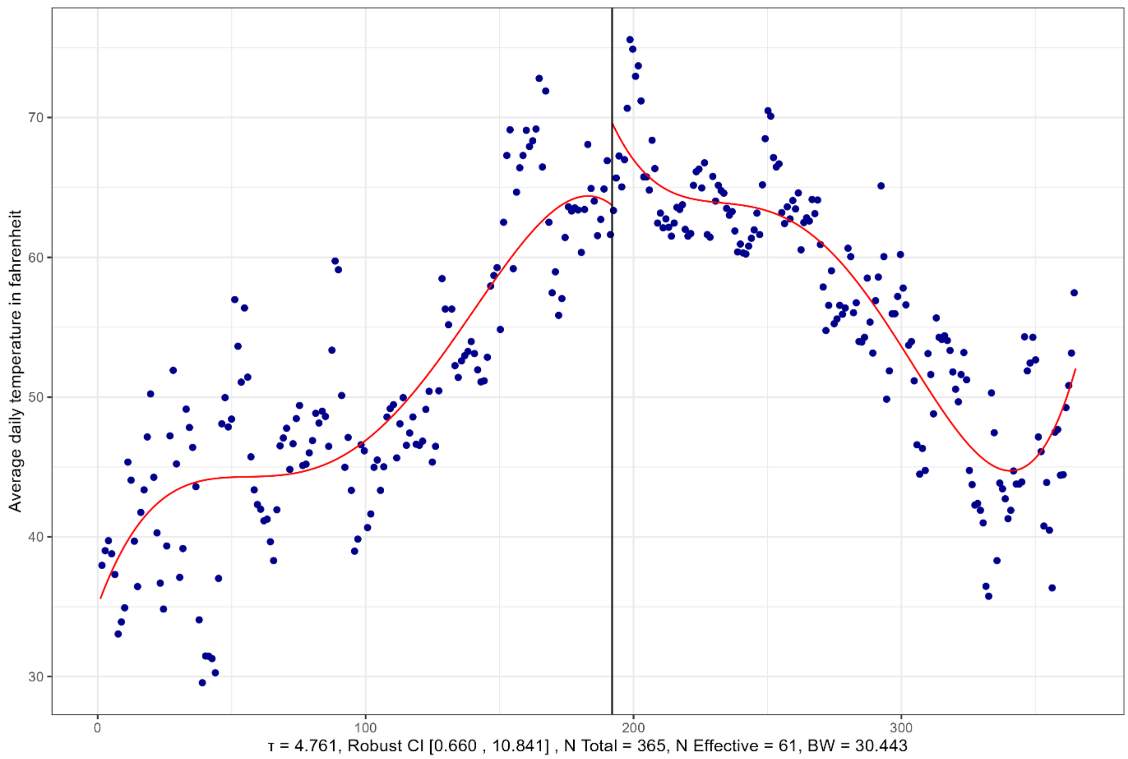

Figure A2. RD with average daily temperature as outcome.

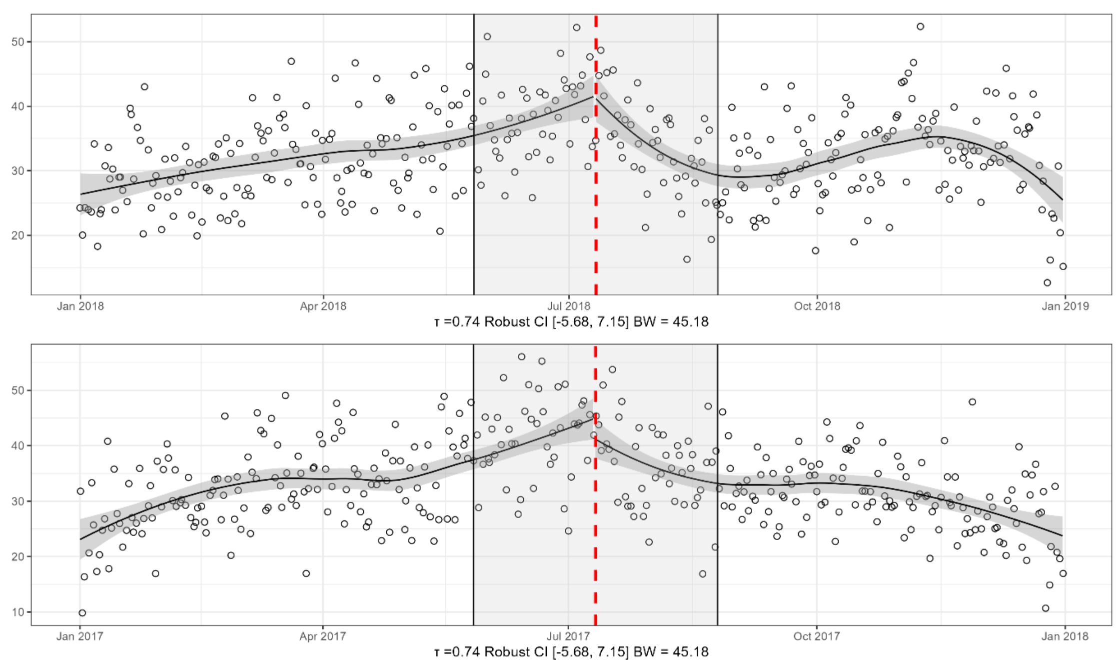

Figure A3. RD models for 2017 and 2018.

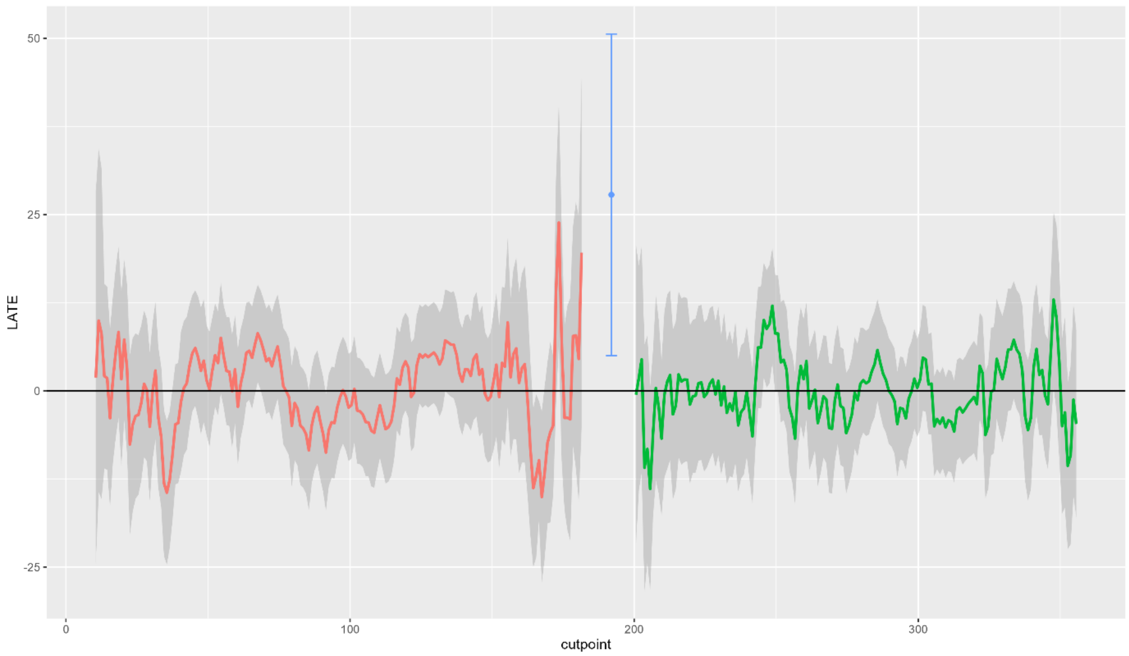

Figure A4. Fake cut-offs.

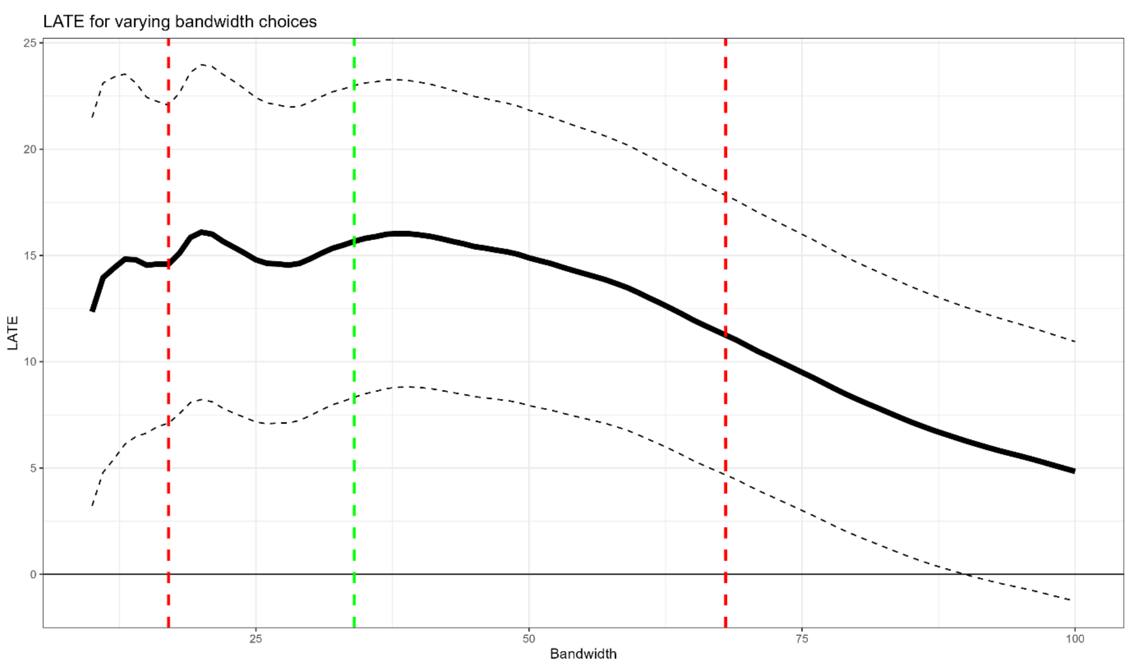

Fig A5. LATEs for varying bandwidth choices. We start at bw = 10 for reasons of statistical power. Dashed green line is the optimal bandwidth chosen in the main models. Dashed red lines indicate double and half that bandwidth.

Parametric RDiT analysis

Following (Hausman and Rapson 2018), we extend the local linear RDiT approach by a traditional parametric approach which involves using the entire available data and estimating the discontinuity by choosing the appropriate polynomial (Angrist and Pischke 2009). One drawback of this design is that the estimates are highly sensitive to the degree of the polynomial and the choice is usually quite arbitrary (Gelman and Imbens 2019). The best practice is to choose the polynomial order based on the smallest Bayesian Information Criterion (BIC) of the respective models (Hausman and Rapson 2018). In our case, this means estimating a model with a polynomial of order 4. Again, high order polynomials lead to poor coverage of confidence intervals (Gelman and Imbens 2019). For the sake of transparency, we also ran regressions with smaller and higher order polynomials. Results are graphically displayed in the appendix in Figure A5. The discontinuity estimate is not significant in the linear and quadratic model. However, for all models with higher order polynomials, the effect is positive and significantly different from zero. While the local linear RD clearly outperforms the parametric model in this application, it is reassuring to know that the effect is not dependent on the particular approach taken. In the case of time-varying treatment effects, the parametric models might not be able to estimate the initial local impact (in time) to the threshold (Hausman and Rapson 2018). We therefore assess the duration of the effect by applying an interrupted time series analysis in appendix A2.

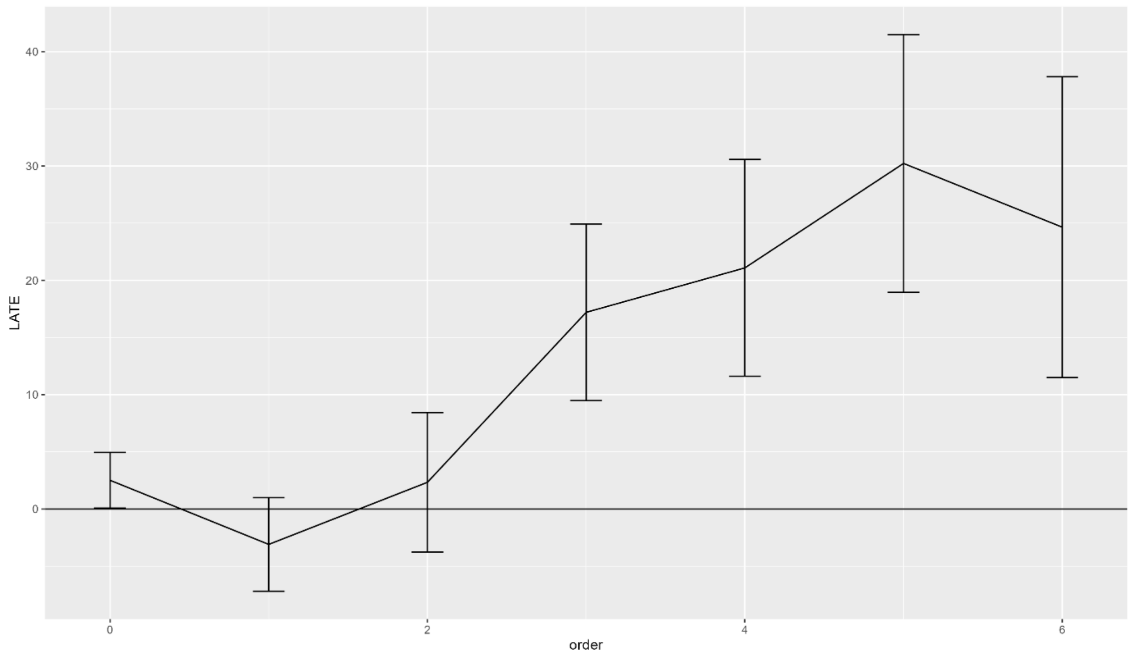

Figure A6. LATE dependent on chosen polynomial order.

RDiT hourly analyses

Since the data we were provided by the Met police is extremely granular, we are also able to assess the LATE in an hourly analysis. The Euro 2021 final ended at 22:53 British Summer Time (BST). We thus used 23:00 BST as the cut-off in the hourly RDiT analyses. Results can be seen in Figure A7. The analysis suggests an increase of 0.78 racial hate crimes on average in the post treatment period per hour compared to the pre-treatment period (Robust CI [0.19, 1.14]).

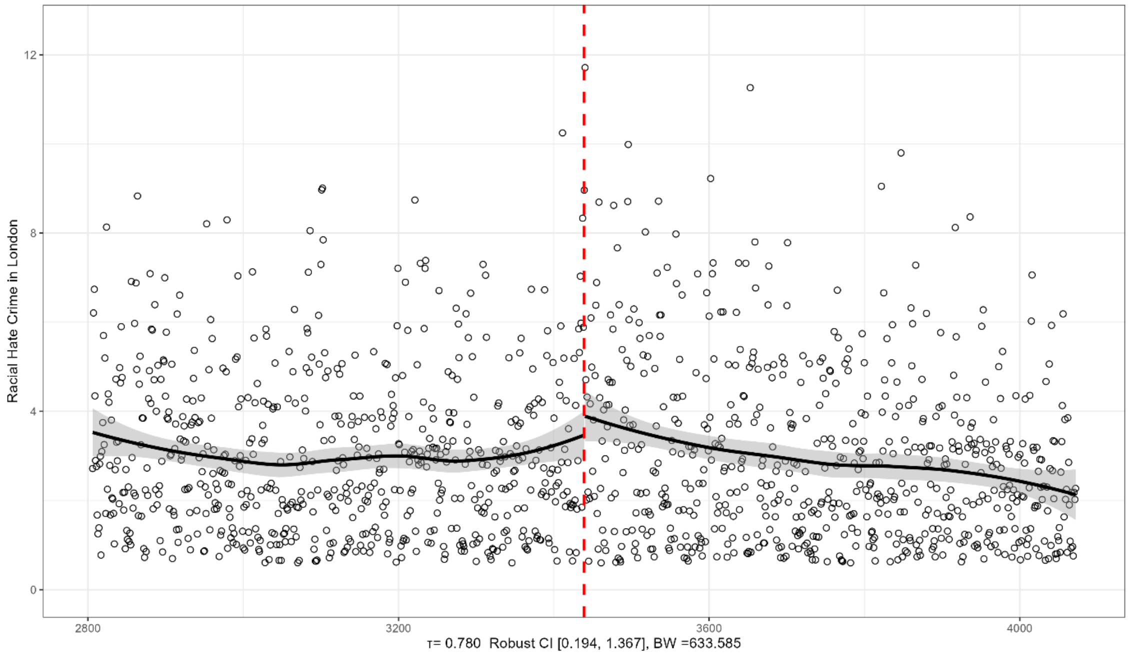

Figure A7. Discontinuity estimate for a RD model with optimal bandwidth in an hourly analysis.

APPENDIX A2

A2. Interrupted Time Series Analysis

We start our analyses by running a preliminary Ordinary Least Squares (OLS) model to assess the potential for temporal autocorrelation in the complete daily time series. Here, we regress daily racial hate crimes on a treatment dummy, a simple linear time indicator, and a trend variable (an interaction of the former two variables). First, we run a Durbin-Watson-test which indicates presence of autocorrelation at the first lag (DW = 0.17, $p < 0.001$). Secondly, we graph the residuals from a preliminary OLS regression to check for serially correlated errors. As can be seen in Figure A8, there is no clear pattern that would indicate temporal autocorrelation. Since these two tests come to dissimilar conclusions, we finally assessed both an autocorrelation function plot and a partial autocorrelation function plot. In the first plot in Figure A9, we see an exponential decay of significant lags. The partial autocorrelation plot furthermore suggests one significant lag before dropping to zero, indicating an AR (autoregression) 1 process. We also tested alternative model specifications of autoregression and/or moving averages in likelihood-ratio tests which did not provide a better model fit.

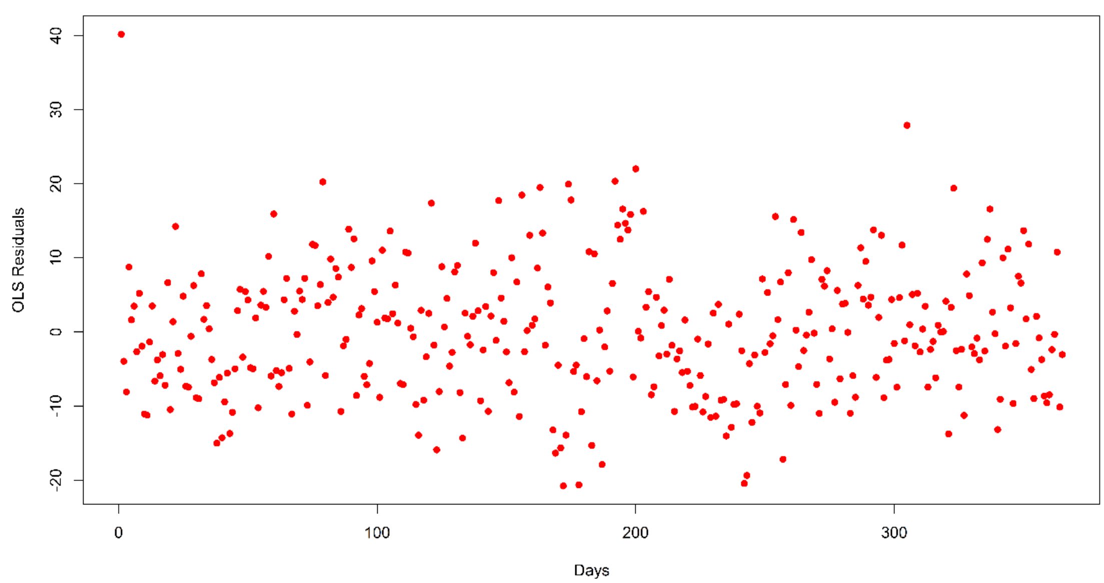

Figure A8. Simple OLS model residuals plotted against weekly time intervals.

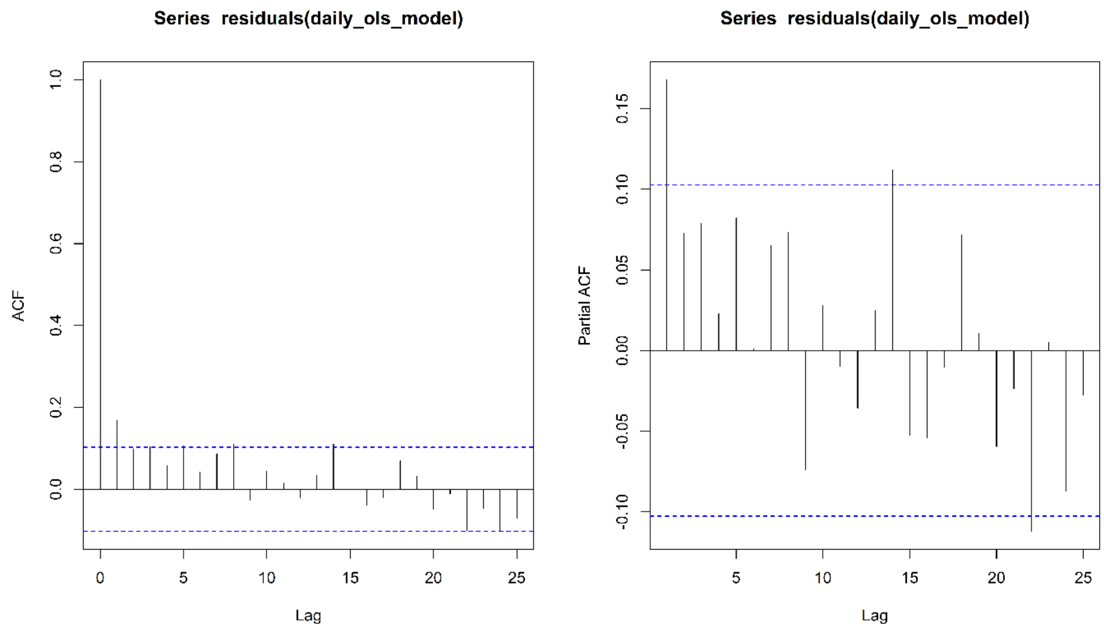

Figure A9. Autocorrelation Function (ACF) plot and Partial Autocorrelation (PACF) plot

We then performed Augmented-Dickey-Fuller (ADF) tests for stationarity (constant mean and variance over time). The overall time series is non-stationary (Dickey-Fuller = -2.48, $p = 0.37$). We account for this by using first differencing of the outcome in a SARIMA (Seasonal Autoregressive Integrated Moving Average) model with an AR (1), MA (0) process and a seasonal pattern of s = 7. This weekly pattern allows to model the weekly plummet in racial hate crimes on Sundays.[1]

Our goal is to arrive at a more precise estimation of the duration of the treatment effect. We therefore follow Piatkowska and Stults (2022) and include 20-day lags of the triggering event variable.[2] These models allow to study the effect on each of the 20 days in the wake of the EURO 2020 final. Similar to the models in the paper, we add weekday dummies as well as daily average temperature as additional covariates. Figure A10 plots the daily lags from models with (grey dots) and without (black dots) these additional covariates. Racial hate crimes appear to be significantly and substantially elevated by roughly 25 additional counts per day for 7 days (including the day of the event). After this period, the following day dummy is not significantly

[1] Note that simpler models (such as ARMA (Autoregressive Moving Average) models with AR (1) and MA (0) processes lead to very similar point estimates as those presented in figure A10. The same applies to different seasonal patterns such as s = 30 or s = 14

[2] Note that the study by Piatkwoski use 15-day lags. Our findings are largely non-sensitive to the number of daily dummies.

different from zero, which is likely to be attributed to the fact that Sundays generally have the lowest count for hate crimes in our data. After this day, the treatment dummies are not consecutively statistically significant with only two outliers (day 8 after the final and day 11 after the final).

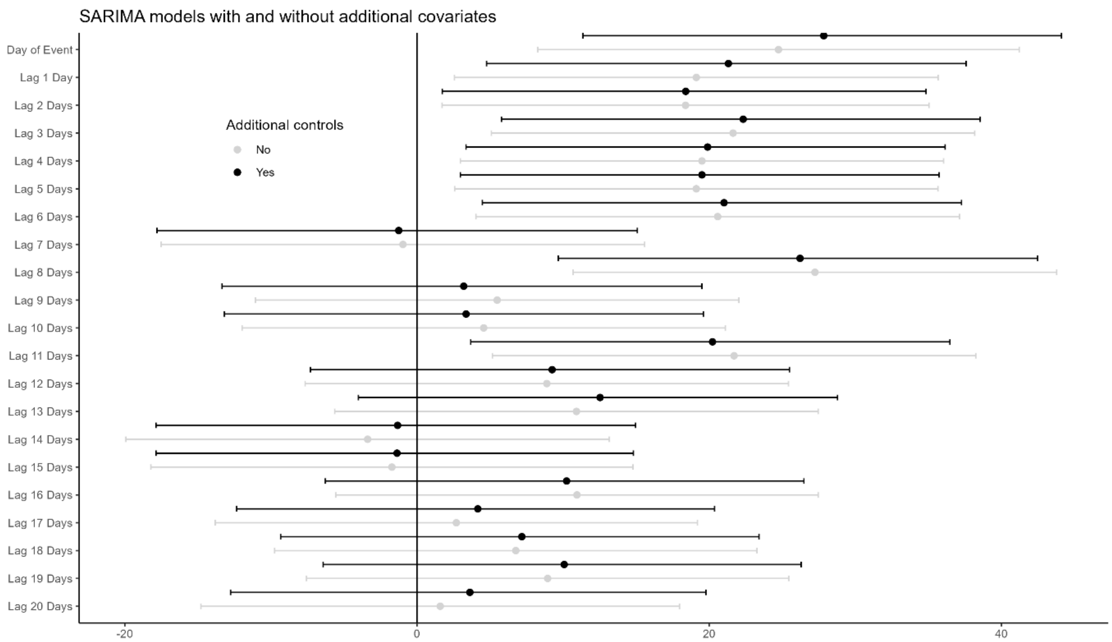

Figure A10. Daily treatment dummies for SARIMA models with and without additional covariates.

APPENDIX A3

A3. Additional robustness checks

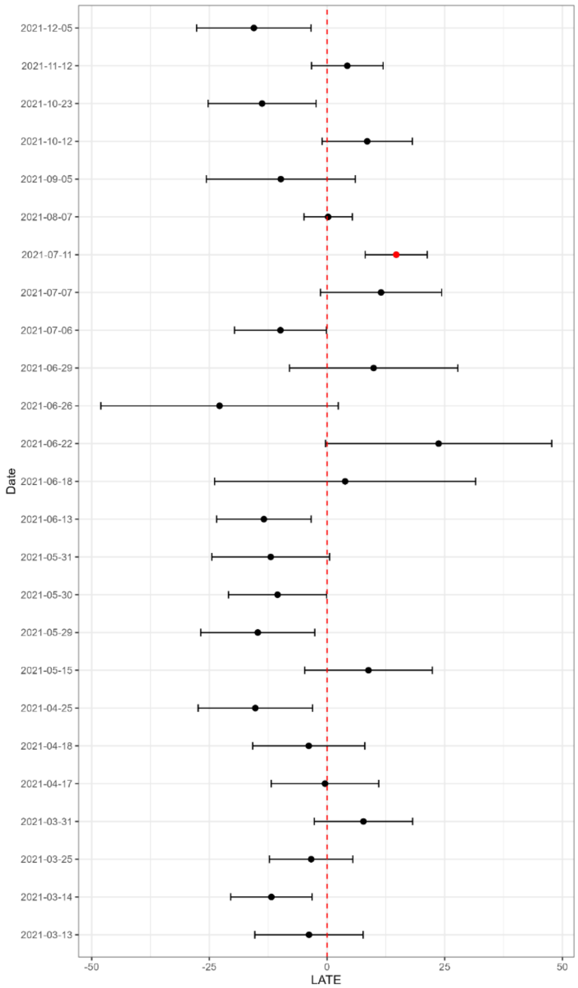

Fig A11. Placebo RD estimates on all games in Wembley in 2021

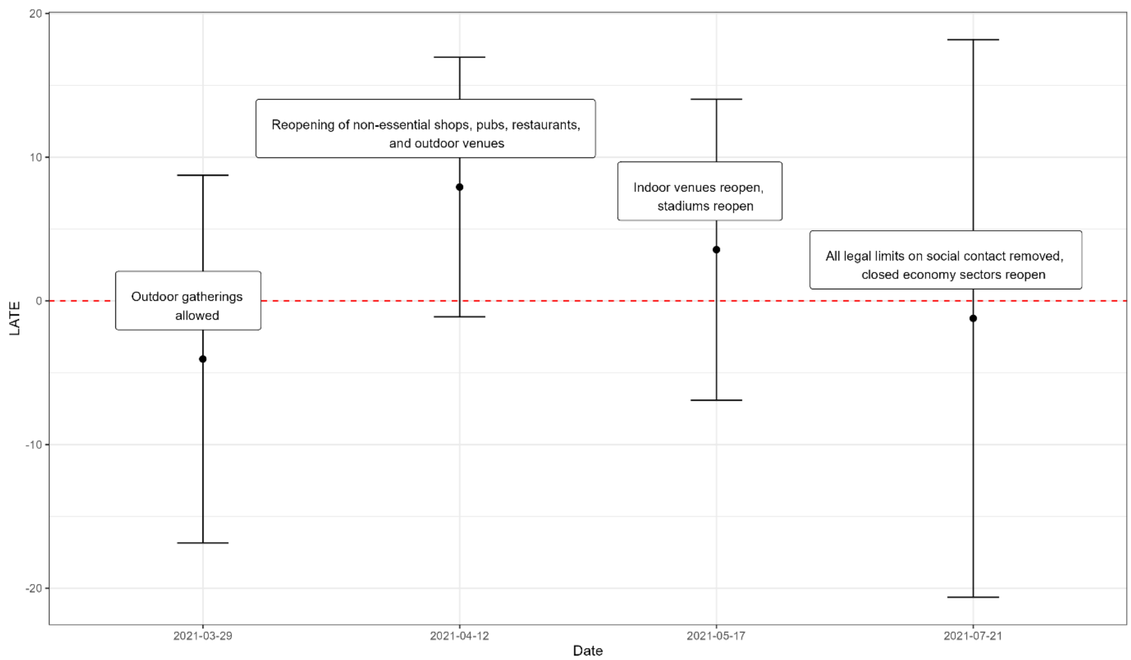

Fig A12. Placebo RD estimates on all COVID-19-restriction relaxations.

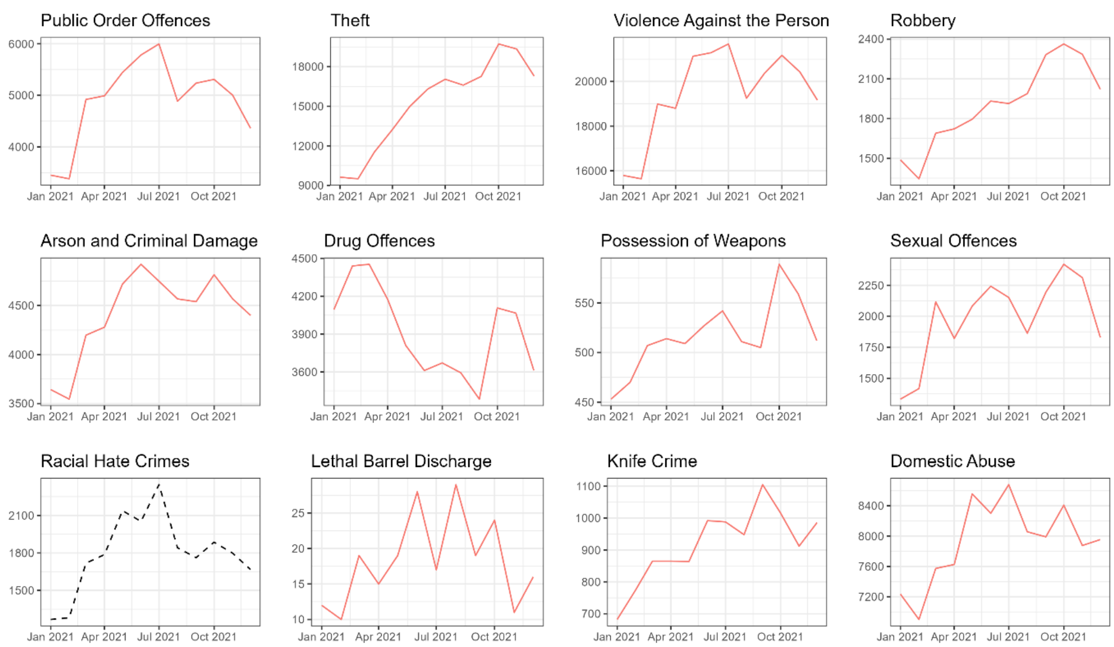

Fig A13. Monthly recorded crime trends throughout the year 2021. Y-axis in each plot is set by the default option in the software program to visualise relative spikes.

APPENDIX A4

A4. Spatially heterogenous treatment effects

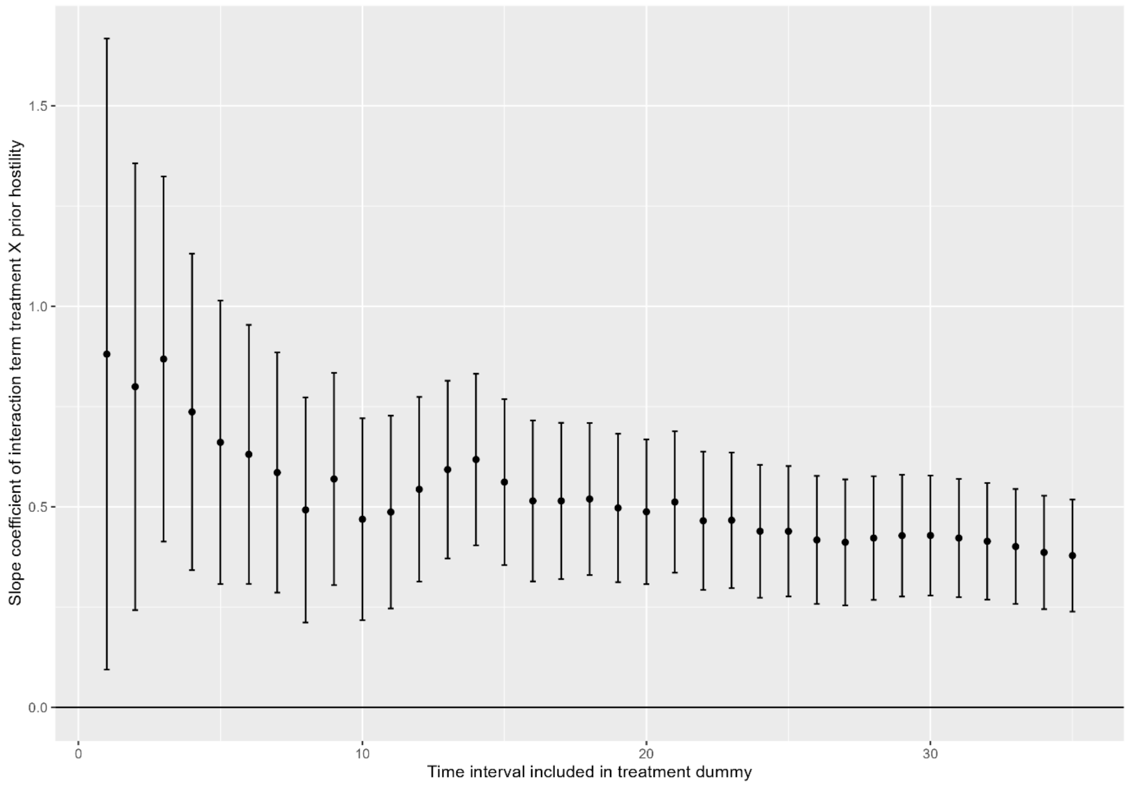

Fig A14. Interaction effect slope coefficient for choosing different lengths for the treatment dummies.

Table A1. Fixed effects regression models with logged dependent variable

	Racial Hate Crimes			
	Model 1	Model 2	Model 3	Model 4
After Euro 2020 final (OBW)	0.13***	0.03	0.13***	0.03
	(0.02)	(0.02)	(0.01)	(0.02)
After Euro 2020 final (OBW) * Prior Hostility			0.09**	0.09**
			(0.03)	(0.03)
R^2	0.01	0.02	0.01	0.03
Adjusted R^2	0.00	0.02	0.00	0.02
F Statistic	66.93***	95.21***	38.12***	33.08***
Borough fixed effects	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	No	Yes

Notes: * p<0.05; ** p<0.01; *** p<0.001, N = 11,680 in all models, Cluster-Robust Standard Errors in brackets, OBW = optimal bandwidth from RDiT Model 1 in table 1 (35 days after the final), interacted vectors are mean centred, additional controls include a linear time spline, average temperature, and weekday dummies

Table A2. Negative Binomial regressions

	Racial Hate Crimes			
	Model 1	Model 2	Model 3	Model 4
After Euro 2020 final (OBW)	1.26***	1.06	1.24***	1.04
	(0.04)	(0.03)	(0.07)	(0.03)
After Euro 2020 final (OBW) * Prior Hostility			1.21***	1.21***
			(0.07)	(0.05)
Observations	11,680	11,680	11,680	11,680
AIC	35386.8	35189.9	35376.5	35179.0
BIC	35408.8	35271.0	35406.0	35267.3
RMSE	1.29	1.28	1.29	1.28
Borough fixed effects	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	No	Yes

Notes: *** p<0.001, Displayed coefficients are incidence rate ratios, N = 11,680 in all models , Cluster-Robust Standard Errors in brackets, OBW = optimal bandwidth from RDiT Model 1 in table 1 (35 days after the final), interacted vectors are mean centred, additional controls include a linear time spline, average temperature, and weekday dummies

REFERENCES

Angrist, J. D., and J. S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37(3):447–56. doi: 10.1080/07350015.2017.1366909.

Hausman, Catherine, and David S. Rapson. 2018. "Regression Discontinuity in Time: Considerations for Empirical Applications." *Annual Review of Resource Economics* 10(1):533–52. doi: 10.1146/annurev-resource-121517-033306.

Piatkowska, Sylwia J., and Brian J. Stults. 2022. "Brexit, Terrorist Attacks, and Hate Crime: A Longitudinal Analysis." *Social Problems* 69(4):968–96. doi: 10.1093/socpro/spab005.