Final Paper Causal Inference* – Explaining Public Attention through Electoral Success with RDD†

Lukas Warode‡

Abstract

This paper uses Regression Discontinuity Design (RDD) to estimate the effect of close district elections on digital public attention, measured as the dynamic of Wikipedia page views. By using different operationalizations of the outcome in linear RDD models with both common and separate slopes, the author finds stable positive estimations of the local average treatment effect (LATE) across constituency candidates from three German federal elections (2009, 2013 and 2017).

Introduction

The last few years have been characterized by increasing overall digitalization in general and "mediatisation" of politics in particular (Marcinkowski 2014). This trend reveals new opportunities for political science, while also challenging existing frameworks and established theories. One of the most visited webpages in the internet – the only encyclopedia *Wikipedia* – presents a controversial pillar in the research community. While facing a lot of rejection in academia since its establishment, recent studies have shown the general potential of Wikipedia as a far-reaching data source (Brown 2011; Göbel and Munzert 2018), which is also capable of being a competitor or replacement of existing established data sources in political science (Herrmann and Döring 2021). Wikipedia is not only delivering encyclopedic information with its articles per se, but also offering the potential to analyze obtained meta information, such as page editing dynamics and page views. The level of observation plays a crucial part in this case: With more general units, such as parties or elections, the number of observations is rather low and the scope of potential analyses might be limited, whereas MP level analyses do not face the problem of observational scarcity (as least not in established democracies). Göbel and Munzert (2021) introduce the possibilities of analyzing MPs based on scraped data from Wikipedia and present a databsase that covers biographical, political and article meta-information for parliamentarians in several countries.
Election research in political science is classically dominated by "genuine" electoral variables, such as vote shares and turnout. With the rise of regression discontinuity design (RDD) in economics and political science, the most common application of this technique is also falling in the realm of classical electoral analyses (Lee 2008; Hainmueller and Kern 2008; Eggers et al. 2015). With the goal of combining a topic that is relevant for an increasingly digitalized society and polity by making use of an established causal inference method (RDD), this paper is trying to make a contribution in assessing the relationship between MPs' digital public attention and electoral success, while putting an emphasis on technical and data-related specifications, as well as potential analysis problems that are inherently part of the underlying data structure itself.

*https://www.hertie-school.org/en/study/course-catalogue/course/course/causal-inference
†Link to Project Repository
‡Hertie School, Berlin

Theoretical and Empirical Foundations

Public Attention in Digital Spheres

As already introduced, public attention in a digitalized society and polity is an increasingly important concept that offers various possibilities for modern social science applications. When trying to operationalize public attention on the level of parliamentarians, several options are feasible. Wikipedia page editing dynamics represent one potential dynamic that is able to reveal politically motivated behavior, but also has the demerit of an overall lack of generalizability, since page edits do not necessarily reflect a representative measure of public attention across the whole society (Yasseri and Bright 2016; Göbel and Munzert 2018). In fact, they are often motivated by internal political dynamics: Göbel and Munzert (2018, 165) find that 51% of German MPs' Wikipedia page edits can be traced back to IP addresses that are linked to the German Bundestag.
A more suitable and representative metric that serves as an assessment indicator of societal based public attention can be found in the general article traffic of respective MPs (Yasseri and Bright 2016). Yasseri and Bright (2016) find the suitability of page views as a explanatory variable for electoral outcomes. Daily Wikipedia page views on MP level are also part of the *Comparative Legislators Database* (CLD) from Göbel and Munzert (2021), which is marking the transition of Wikipedia based information into the realm of traditional empirical political science.

Electoral Districts and MPs

Analyzing and explaining electoral outcomes is at the core of empirical political science. Vote shares play a prominent role, either on a federal or district level. The former case makes it normally harder to analyze the performance of individual candidates, while the latter presents clear electoral metrics that indicate the performance of candidates and incumbents in elections. Assessing the electoral success of a district candidate can simply be measured by the vote share. In order to make it comparable and to have a clear indication on whether and how much a candidate won or lost the district election, the *Margin of Victory* (MOV) can be used. Hainmueller and Kern (2008, 219) define it as the difference between the vote share of the party and the party with the highest vote share apart from the party itself. For winning parties, it is simply the vote share difference between them and the party with the second highest vote share, while the MOV for all other parties is the difference between their vote share and the result of the winning party. By definition, all winning candidates have a positive MOV, whereas all loosing competitors have a negative MOV, which also yields a dichotomous differentiation.
Electoral systems that at least include a single member district (SMD) element, such as the German (federal) system, are suitable for analyzing incumbency via MOVs (Nohlen 2013). Another important factor for the suitability of the case selection lies in the stability of political systems with regards to the direct candidates and its party affiliations. In emerging democracies and thus emerging party systems, observing consistent incumbency and party affiliation is less likely to happen, with the implication of a rather vague research design structure. Analyzing the effect of incumbency in established democracies, such as the United Kingdom (Eggers and Hainmueller 2009) or Germany (Hainmueller and Kern 2008), is backed up by validated studies on the other hand.

Regression Discontinuity Design

The aim of this papers is to measure the effect of electoral success on digital public attention in a causal inference framework. As already mentioned, measuring electoral success via MOV is a common procedure in political science, which is also offering the potential of treatment allocation in a regression discontinuity design (RDD). But what is a RDD and why should one use it?
Cook (2008) reflects on the history of RDD and concludes that RDD was "waiting for life," marking the emergence of very influential econometrical articles at the turn of the millennium (J. D. Angrist and Lavy 1999; Hahn, Todd, and Van der Klaauw 2001). Lee (2008) and Hainmueller and Kern (2008) confirm the applicability of RDD in political science. The central idea of RDD is to assign treatment according to another variable (either called *running* or *forcing* variable). Figure 1 (Steiner et al. 2017, 264) is showing the directed acyclic graph (DAG) that explains the causal structure of RDDs. Graph A is presenting the

structure before the limiting procedure of the RDD is applied: The running (continuous) variable X is causing both the treatment D and the outcome Y, while being potentially related to a set of variables U. Thus, X presents an observable confounder.

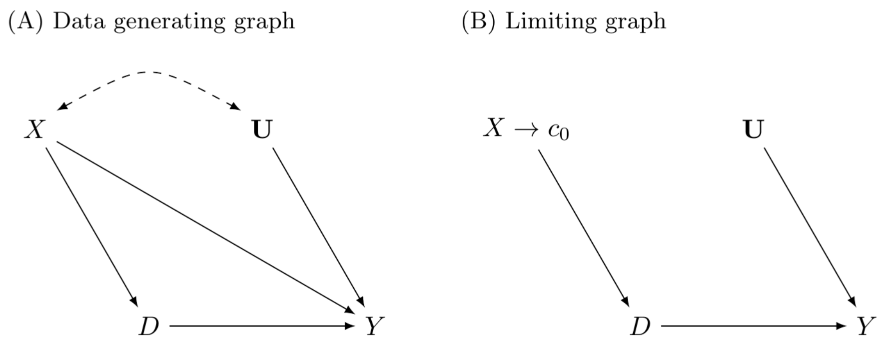


Figure 1: DAG of RDD

The limiting graph is indicating how we can identify causal effects in the given structure: X determines the value of treatment with the cutpoint c_0. If X is greater than the chosen cutpoint c_0, the observation is assigned i with treatment ($D_i = 1$), otherwise (X equal or less than c_0) the observation i is assigned to the control group ($D_i = 0$):

$$D_i = \begin{cases} 1 \text{ if} & X_i > c_0 \\ 0 \text{ if} & X_i \leq c_0 \end{cases}$$

The core assumption for estimating causal effects lies in the assumed *local randomization* around the cutoff c_0. In addition, the continuity of average potential outcomes around c_0 is another key assumption: $E\big[Y_i(d) \mid X_i = x\big]$ is continuous in x around $X_i = c_0$ for $d = 0, 1$. This implies that we are able to estimate the *local average treatment effect* (LATE)[1] around a certain (optimal) bandwidth h, which can be estimated (J. Angrist and Imbens 1995; Imbens and Kalyanaraman 2012). The LATE is defined as follows:

$$\delta_{LATE} = E\big[Y_i(1) - Y_i(0) \mid X_i = c_0\big]$$

The size of the (optimal) bandwidth $h > 0$ is deciding how large the range or window of included observations around the cutoff c_0 is:

$$c_0 - h \leq X_i \leq c_0 + h$$

The size of the optimal bandwidth is conventionally calculated by using the Imbens-Kalyanaraman (IK) algorithm (Imbens and Kalyanaraman 2012). Estimating the LATE implies also necessity of choosing a certain model design. Several options are possible: Non-parametric, parametric (linear and non-linear models), while parametric models can be estimated with common and different slopes. It is often advised to avoid too complex models, such as high-order polynomials (Gelman and Imbens 2019).

Data

To be able to apply the research design implies the case selection of countries that at least partially use SMDs in their electoral system. Prominent examples of political systems in Europe that use SMDs include the United Kingdom and Germany. Germany has proven to be a suitable country for applying both election

[1] This paper is solely making use of sharp RDD, while fuzzy RDD is not playing a role in the research design. Many papers conventionally label the sharp LATE as δ_{SRD}, referring to the abbreviation of sharp regression discontinuity (SRD).

research on SMDs by using MOV and RDD (Hainmueller and Kern 2008), as well as also ensuring appropriate data quality regarding MPs on the federal level (Göbel and Munzert 2018). Hence, Germany is chosen as the country for conducting the analysis.

The CLD (Göbel and Munzert 2021) offers traffic data for German (federal) MPs from December 2007 until end of March 2020 (using the `R` API `legislatoR` version 1.0). It is possible to analyze electoral district races of MPs of every federal election that falls within this time span. Consequently, the included elections for this analysis are the federal elections 2009, 2013 and 2017. Federal election results on a constituency level can be obtained from the *Bundeswahlleiter*[2]. The distribution of the MOVs[3] is shown per election in Figure 2. The probability density plots are indicating that the MOV is roughly equally distributed across the three included elections, while also highlighting the imbalance in the variable when comparing district winners (positive MOV) and (potential[4]) district losers (negative MOV). The figure also highlights the imbalance of the MOV variable in terms of actual winners and losers: In all three elections, the large majority of candidates (n: 1046, 90%) has a positive MOV (election winners), whereas only 10% (n: 114) of the sample are election losers (negative MOV). However, in a RDD analysis, this fact will likely not cause statistical problems, since only values around the cutoff threshold are included, which effectively excludes observations with high positive MOVs.

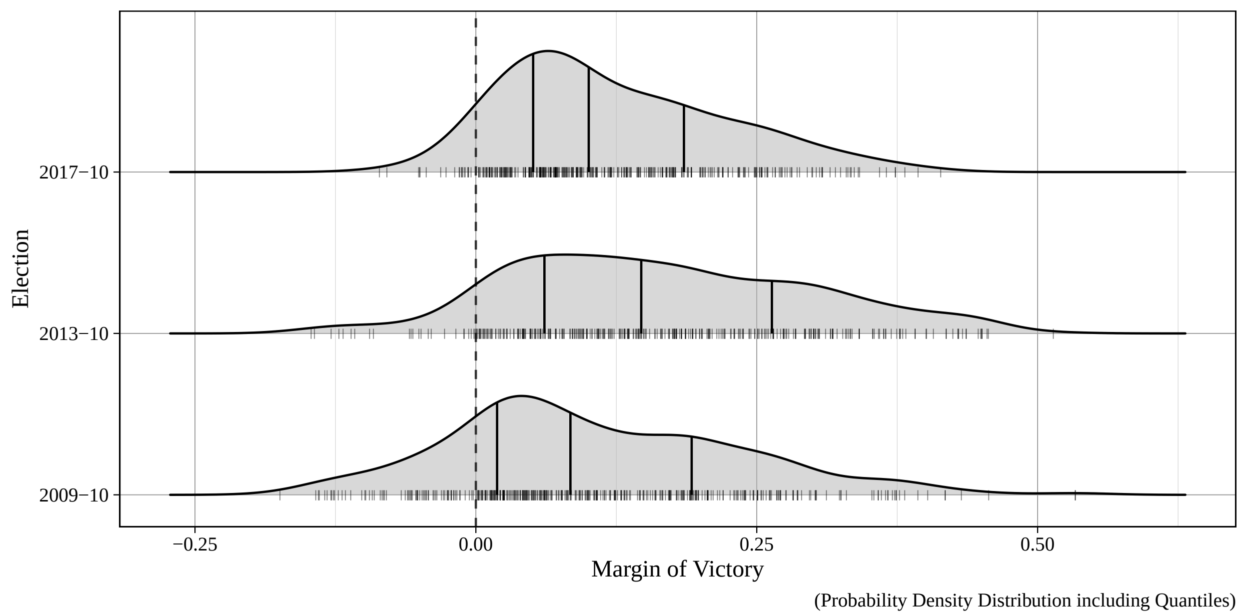


Figure 2: Distribution of Margin of Victory

To capture the dynamics of MPs' digital public attention, it is not enough to look only at a fixed value metric, but rather to measure temporal trends. To measure the trend in a given time interval, the difference between either monthly sum or monthly mean values of page views from two dates are considered. When calculating the trend, it is important to use the same period respectively:

$$Trend\ Mean_{ijd} = Mean_{ij+d} - Mean_{ij-d}$$

[2]https://www.bundeswahlleiter.de/bundestagswahlen/2021/ergebnisse.html

[3]The code of this paper, used for this and the following calculations and visualizations, can be found here: https://github.com/lwarode/ci_final_paper

[4]A demerit in using CLD for the research design is the formal scope of actual parliamentarians, while data on candidates is not part of the database. This implies that we can only calculate the MOV for a) all winning candidates, and b) for all losing candiates that are incumbent, whereby we cannot fully be sure that losing incumbents did actually compete again in the same district.

$$Trend\ Sum_{ijd} = Sum_{ij+d} - Sum_{ij-d}$$

For every every MP i in month j, the trend can be calculated by choosing a time period value d. For instance, when capturing the mean trend of MP i for the federal election in October 2009 with $d = 1\ month$, the difference between the mean monthly page views for November 2009 and September 2009 is computed, covering the dynamics of a 2 month interval.

Table 1 is showing all the relevant variables for this analysis. For both monthly mean and sum page view dynamics, every trend metric with time period values (d) 1, 3, 6 and 12 months is calculated, yielding respectively the dynamics of 2, 6, 12 and 24 months time intervals. Looking at the trend values, the large variance is striking. For instance, the 3 months sum trend metric is ranging from -70449 to 57477, while having a standard deviation (SD) that is 43 times larger than the mean of the metric. The appendix also includes scatterplots of the non-transformed trend metrics in relationship with the MOV, highlighting outliers. This in indeed no surprise: As already the non-trend monthly mean and sum values are indicating, there is a large variation in the variable, which is due to the uneven actual (public) relevance and attention that MPs receive. When prominent politicians, like former chancellor Angela Merkel did win her district elections from 2009 to 2017, her level of (digital) public attention exceeds the average first-time district winner by far. What are pertinent measures to avoid dealing with a large variance and skewness in the outcome measure, which could violate relevant statistical assumptions? Transforming and rescaling are established ways in statistical applications to deal with those problems. A prominent way is applying log-transformation (Changyong et al. 2014). While log-transforming is not solving all the problems effectively, it is also incapable in converting negative values, which leads to an effective pruning of the sample, when dealing with negative values that are prominent in the trend metrics as indicated in Table 1: The number of lost observations, when log-transforming reaches from 416 (36% of all observations) to 856 (74% of all observations), which constitutes a huge share of the original sample respectively. However, it could also be interesting to just analyze MPs that enjoyed a digital public attention boost, which would be effectively the case when working with log-transformed outcomes. When trying to obtain a similar result when transforming skewed variables, but also also being able to rescale negative values, inverse hyperbolic sine (ISH) transformation is a common measure in econometrics (Bellemare and Wichman 2020). Applying ISH to a variable is accomplished in the following way:

$$x_{ISH} = \log(x + \sqrt{x^2 + 1}),$$

where the natural logarithm (ln) is normally used. Table 1 indicates the ISH-transformed trend variables and is showing that the variance, as well as the range, including both positive and negative observations, is reduced without pruning the sample as log-transformation did.

Table 1: Descriptive Statistics of Relevant Variables

	N	Missing	Mean	SD	Min	Median	Max
Margin of Victory	1160	0	0.13	0.12	-0.17	0.11	0.53
Monthly Sum of Page Traffic	1125	35	2265.17	11105.25	0.00	287.00	141257.00
Mean of Daily Page Traffic per Month	1125	35	73.08	358.23	0.00	9.29	4556.68
Trend Mean 1 Month	1114	46	-61.37	676.92	-14267.60	-8.82	1877.47
Trend Mean 3 Months	1103	57	3.78	163.84	-2272.55	1.10	1854.10
Trend Mean 6 Months	1099	61	8.29	118.32	-789.23	1.07	2404.10
Trend Mean 12 Months	1084	76	30.18	375.38	-3172.13	2.24	6030.10
(Log) Trend Mean 1 Month	304	856	1.13	1.96	-3.40	0.98	7.54
(Log) Trend Mean 3 Months	744	416	1.06	2.19	-35.35	0.86	7.53
(Log) Trend Mean 6 Months	702	458	1.17	1.74	-3.40	0.99	7.78
(Log) Trend Mean 12 Months	738	422	1.59	2.26	-38.12	1.49	8.70
(ISH) Trend Mean 1 Month	1114	46	-1.98	2.96	-10.26	-2.87	8.23
(ISH) Trend Mean 3 Months	1103	57	0.66	2.49	-8.42	0.95	8.22
(ISH) Trend Mean 6 Months	1099	61	0.68	2.41	-7.36	0.93	8.48
(ISH) Trend Mean 12 Months	1084	76	0.96	2.71	-8.76	1.55	9.40
Trend Sum 1 Month	1114	46	-1826.45	20306.01	-428028.00	-260.00	56324.00
Trend Sum 3 Months	1103	57	118.12	5078.92	-70449.00	34.00	57477.00
Trend Sum 6 Months	1099	61	248.84	3549.52	-23677.00	32.00	72123.00
Trend Sum 12 Months	1084	76	936.90	11636.58	-98336.00	70.00	186933.00
(Log) Trend Sum 1 Month	309	851	4.52	1.95	0.00	4.38	10.94
(Log) Trend Sum 3 Months	743	417	4.54	1.74	0.00	4.29	10.96
(Log) Trend Sum 6 Months	702	458	4.58	1.74	0.00	4.39	11.19
(Log) Trend Sum 12 Months	738	422	5.08	1.72	0.00	4.93	12.14
(ISH) Trend Sum 1 Month	1114	46	-3.47	5.69	-13.66	-6.25	11.63
(ISH) Trend Sum 3 Months	1103	57	1.82	5.25	-11.86	4.22	11.65
(ISH) Trend Sum 6 Months	1099	61	1.64	5.19	-10.77	4.16	11.88
(ISH) Trend Sum 12 Months	1084	76	2.21	5.52	-12.19	4.94	12.83

[a] Negative infinite as well no real numbers that were generated due to log-scaling the trend variables were replaced with ‘NAs‘ (missing values) in order to display the descriptive statistics correctly.

Figure 3 is showing the relationship between MOV and the trend dynamics for monthly page view mean and sum values (ISH-transformed). The scatterplot is supplemented by separately run linear regression curves, which are indicating a small jump respectively, especially for the 6 and 12 months trends. These jumps are potential indicators of significant LATEs, however, they have to hold locally, after pruning the sample equally around the cutoff value by using IK bandwidth (Imbens and Kalyanaraman 2012). In addition, the plot is also revealing the visual impression that there is no global positive relationship between MOV and the page view trends. In order to be able to make further plausible statistical and causal statements, the RDD needs to be applied.

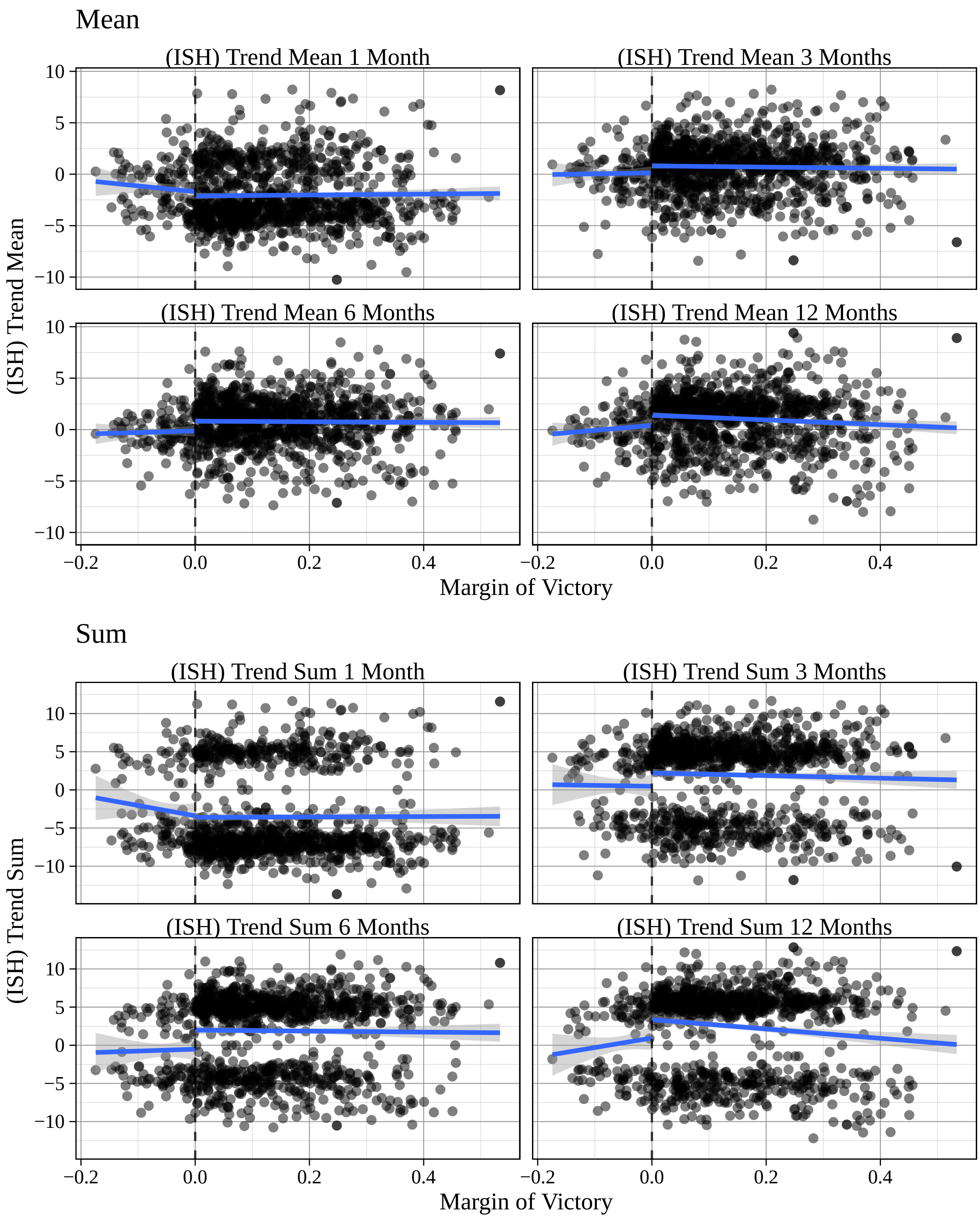


Figure 3: Scatterplot with Separate Linear Regression (ISH Transformation)

Analysis and Results

In this chapter, the results of the RDD are presented and interpreted. As explained above, the respective outcome measures (mean and sum trends) are ISH-transformed in order to deal with the variance and skewness of the variables. Tables 2 (mean trends) and 3 (sum trends) are showing the results for the RDD models by using the same slope for both the control and treatment group. All LATE coefficients in the mean trend models are significant with a significance level of at least 95%. Surprisingly, the LATE for the 1 month mean trend model has a negative sign, while all other LATE coefficients are positive. In the sum trend models, the LATE coefficients of the 3 month model (90% significance level), as well as the 6 and 12 month model (99.9% significance level) are also significant, while the 1 month model LATE is losing its significance. No MOV coefficient of all 8 models is yielding significance. Thus, we can conclude that there is a group level difference between treatment and control group, while the effect of the running variable (MOV) has no general effect around the cutoff value.

Table 2: RDD ISH-Transformed Mean Trends (Same Slope)

	Trend Mean 1 Month	Trend Mean 3 Months	Trend Mean 6 Months	Trend Mean 12 Months
Intercept	−1.292****	−0.041	−0.296	0.106
	(0.285)	(0.272)	(0.225)	(0.250)
LATE	−0.893**	0.864**	1.196****	1.104***
	(0.384)	(0.399)	(0.307)	(0.345)
Margin of Victory	1.704	−2.097	−1.377	−0.546
	(1.669)	(2.851)	(1.410)	(1.609)
Num.Obs.	845	586	807	782
R2	0.007	0.011	0.024	0.021
R2 Adj.	0.005	0.008	0.022	0.018
AIC	4158.5	2675.4	3576.9	3608.4
BIC	4177.4	2692.9	3595.7	3627.1
Log.Lik.	−2075.245	−1333.693	−1784.446	−1800.209
F	2.937	3.301	10.059	8.360
RMSE	2.83	2.36	2.21	2.42

Note: Standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01, **** p<0.001.

Table 3: RDD ISH-Transformed Sum Trends (Same Slope)

	Trend Sum 1 Month	Trend Sum 3 Months	Trend Sum 6 Months	Trend Sum 12 Months
Intercept	−2.918****	0.377	−0.824	0.092
	(0.568)	(0.646)	(0.524)	(0.533)
LATE	−0.419	1.726*	3.039****	2.892****
	(0.792)	(0.975)	(0.740)	(0.731)
Margin of Victory	−5.024	−0.712	−3.605	−2.632
	(4.052)	(9.193)	(4.041)	(3.385)
Num.Obs.	752	465	713	787
R2	0.007	0.016	0.035	0.028
R2 Adj.	0.005	0.012	0.032	0.025
AIC	4693.4	2834.8	4310.0	4821.9
BIC	4711.9	2851.4	4328.2	4840.6
Log.Lik.	−2342.694	−1413.399	−2150.983	−2406.955
F	2.698	3.753	12.785	11.267
RMSE	5.46	5.07	4.95	5.16

Note: Standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01, **** p<0.001.

Tables 4 and 5 show the results for the mean and sum trend models (also ISH-transformed), where the regression slopes were estimated separately. Again, no MOV coefficient is significant, also not when estimated group-wise. While the 1 and 3 months trend models are providing no significant LATE coefficient, neither for the mean nor for the sum trends, the 6 and 12 months trend models are yielding significant LATE coefficients across all model specifications. The results for the (16) corresponding models without any transformation of the trend outcome measures are part of the appendix (Tables 6, 7, 8 and 9). None of the models yield any significant results, which clearly emphasizes the necessity of applying ISH transformation.

Table 4: RDD ISH-Transformed Mean Trends (Separate Slopes)

	Trend Mean 1 Month	Trend Mean 3 Months	Trend Mean 6 Months	Trend Mean 12 Months
Intercept	−1.703****	0.222	−0.133	0.412
	(0.447)	(0.406)	(0.350)	(0.391)
LATE	−0.533	0.666	1.056***	0.842*
	(0.488)	(0.459)	(0.384)	(0.430)
Margin of Victory (Control)	−5.658	3.523	1.530	4.844
	(6.385)	(7.021)	(5.003)	(5.538)
Margin of Victory (Treatment)	7.902	−6.729	−3.158	−5.887
	(6.615)	(7.683)	(5.214)	(5.787)
Num.Obs.	845	586	807	782
R2	0.009	0.012	0.025	0.022
R2 Adj.	0.005	0.007	0.021	0.019
AIC	4159.1	2676.6	3578.5	3609.4
BIC	4182.8	2698.5	3602.0	3632.7
Log.Lik.	−2074.528	−1333.308	−1784.262	−1799.689
F	2.435	2.455	6.823	5.918
RMSE	2.82	2.36	2.21	2.42

Note: Standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01, **** p<0.001.

Table 5: RDD ISH-Transformed Sum Trends (Separate Slopes)

	Trend Sum 1 Month	Trend Sum 3 Months	Trend Sum 6 Months	Trend Sum 12 Months
Intercept	−3.391****	1.498	−0.574	0.950
	(0.865)	(0.989)	(0.795)	(0.833)
LATE	−0.029	0.890	2.837***	2.157**
	(0.957)	(1.123)	(0.884)	(0.914)
Margin of Victory (Control)	−13.491	29.703	0.961	12.491
	(12.355)	(22.325)	(11.629)	(11.792)
Margin of Victory (Treatment)	9.488	−36.604	−5.193	−16.480
	(13.079)	(24.491)	(12.403)	(12.310)
Num.Obs.	752	465	713	787
R2	0.008	0.021	0.035	0.030
R2 Adj.	0.004	0.014	0.031	0.026
AIC	4694.9	2834.5	4311.8	4822.1
BIC	4718.0	2855.3	4334.6	4845.5
Log.Lik.	−2342.430	−1412.275	−2150.895	−2406.055
F	1.973	3.253	8.572	8.116
RMSE	5.47	5.07	4.96	5.16

Note: Standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01, **** p<0.001.

Discussion and Conclusion

The goal of the paper was to apply the established RDD framework on a rather novel field of empirical political science: Using electoral district vote share results of 3 German federal elections as an explanatory variable in order to understand page view trend dynamics of MP's Wikipedia articles as a measure of (digital) public attention in a causal inference setup. While the original outcome measures are highly skewed with high positive and negative outlier values, rescaling (ISH transformation) the dependent variables leads to stable significant result of the models. All models with 6 and 12 months trend outcome measures (covering 12 and 24 months time spans) are yielding significant positive estimations of the LATE. While the effect of the running variable itself (MOV) is overall not meaningful, there is a significant difference between the control and treatment group in the mentioned time spans.

Future work should focus on the imbalance of the page view trend dynamics, since it directly reflects the public attention of politicians. Statistical and causal measures could involve clustering and matching strategies, which are sufficiently appropriate to deal with MPs' heterogeneity from a theoretical and empirical perspective.

In addition, the chosen bandwidth of the IK algorithm was leading to a relatively large causal discontinuity sample. This is potentially problematic for both causal and theoretical considerations. Prior work has been placing a key emphasis on "close district races," which is likely not met in this paper's analysis. However, the reliability of the results can be tested by applying artificially chosen smaller bandwidth values. In addition, this paper only analyzed district candidates that were competing in German federal elections. Comparing the results with analyses from other countries that use SMDs should be a future goal, as well as linking the approach to current MP level research (Rauh and Schwalbach 2020; Sältzer 2020). While party affiliation is an already existing metric, further more specific political variables that vary on the MP level could really benefit our understanding of political elite behavior and its connection to the public sphere.

References

Angrist, Joshua D, and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *The Quarterly Journal of Economics* 114 (2): 533–75.

Angrist, Joshua, and Guido Imbens. 1995. "Identification and Estimation of Local Average Treatment Effects." National Bureau of Economic Research Cambridge, Mass., USA.

Bellemare, Marc F, and Casey J Wichman. 2020. "Elasticities and the Inverse Hyperbolic Sine Transformation." *Oxford Bulletin of Economics and Statistics* 82 (1): 50–61.

Brown, Adam R. 2011. "Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage." *PS: Political Science & Politics* 44 (2): 339–43.

Changyong, FENG, WANG Hongyue, LU Naiji, CHEN Tian, HE Hua, LU Ying, et al. 2014. "Log-Transformation and Its Implications for Data Analysis." *Shanghai Archives of Psychiatry* 26 (2): 105.

Cook, Thomas D. 2008. "'Waiting for Life to Arrive': A History of the Regression-Discontinuity Design in Psychology, Statistics and Economics." *Journal of Econometrics* 142 (2): 636–54.

Eggers, Andrew C, Anthony Fowler, Jens Hainmueller, Andrew B Hall, and James M Snyder Jr. 2015. "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from over 40,000 Close Races." *American Journal of Political Science* 59 (1): 259–74.

Eggers, Andrew C, and Jens Hainmueller. 2009. "MPs for Sale? Returns to Office in Postwar British Politics." *American Political Science Review* 103 (4): 513–33.

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.

Göbel, Sascha, and Simon Munzert. 2018. "Political Advertising on the Wikipedia Marketplace of Information." *Social Science Computer Review* 36 (2): 157–75.

———. 2021. "The Comparative Legislators Database." *British Journal of Political Science*, 1–11.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1): 201–9.

Hainmueller, Jens, and Holger Lutz Kern. 2008. "Incumbency as a Source of Spillover Effects in Mixed Electoral Systems: Evidence from a Regression-Discontinuity Design." *Electoral Studies* 27 (2): 213–27.

Herrmann, Michael, and Holger Döring. 2021. "Party Positions from Wikipedia Classifications of Party Ideology." *Political Analysis*, 1–20.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (3): 933–59.

Lee, David S. 2008. "Randomized Experiments from Non-Random Selection in US House Elections." *Journal of Econometrics* 142 (2): 675–97.

Marcinkowski, Frank. 2014. "Mediatisation of Politics: Reflections on the State of the Concept." *Javnost-the Public* 21 (2): 5–22.

Nohlen, Dieter. 2013. *Wahlrecht Und Parteiensystem*. Springer-Verlag.

Rauh, Christian, and Jan Schwalbach. 2020. "The ParlSpeech V2 Data Set: Full-Text Corpora of 6.3 Million Parliamentary Speeches in the Key Legislative Chambers of Nine Representative Democracies."

Sältzer, Marius. 2020. "Finding the Bird's Wings: Dimensions of Factional Conflict on Twitter." *Party Politics*, 1354068820957960.

Steiner, Peter M, Yongnam Kim, Courtney E Hall, and Dan Su. 2017. "Graphical Models for Quasi-Experimental Designs." *Sociological Methods & Research* 46 (2): 155–88.

Yasseri, Taha, and Jonathan Bright. 2016. "Wikipedia Traffic Data and Electoral Prediction: Towards Theoretically Informed Models." *EPJ Data Science* 5 (1): 1–15.

Appendix

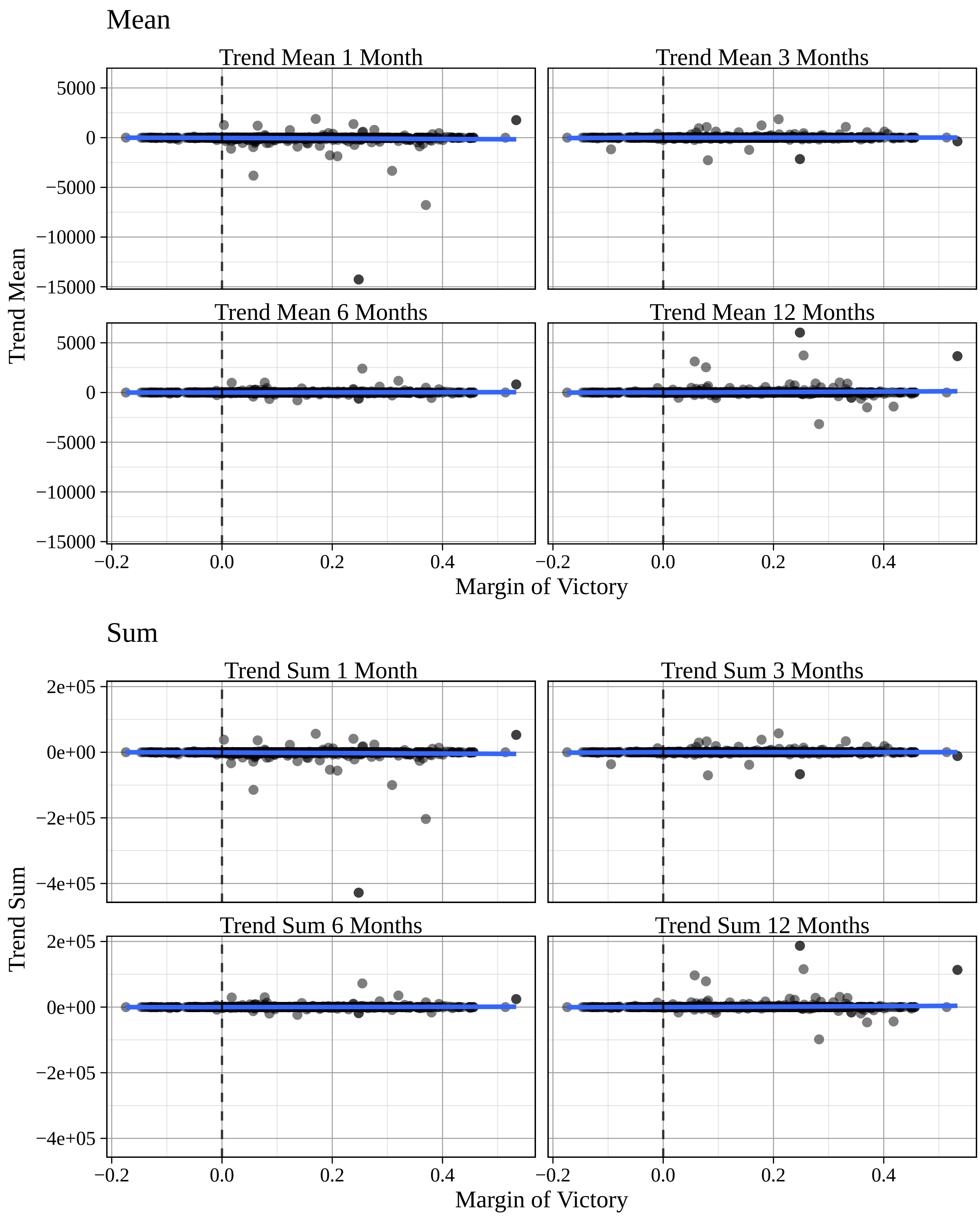
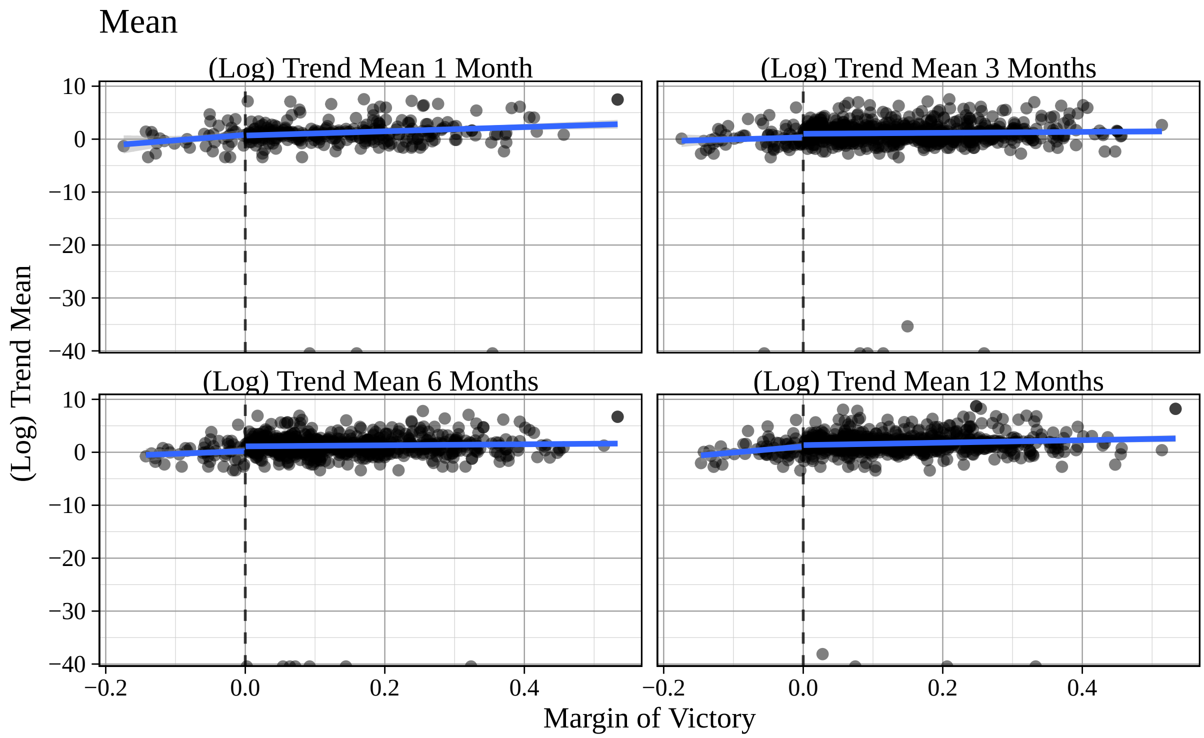


Figure 4: Scatterplot with Separate Linear Regression (No Transformation)



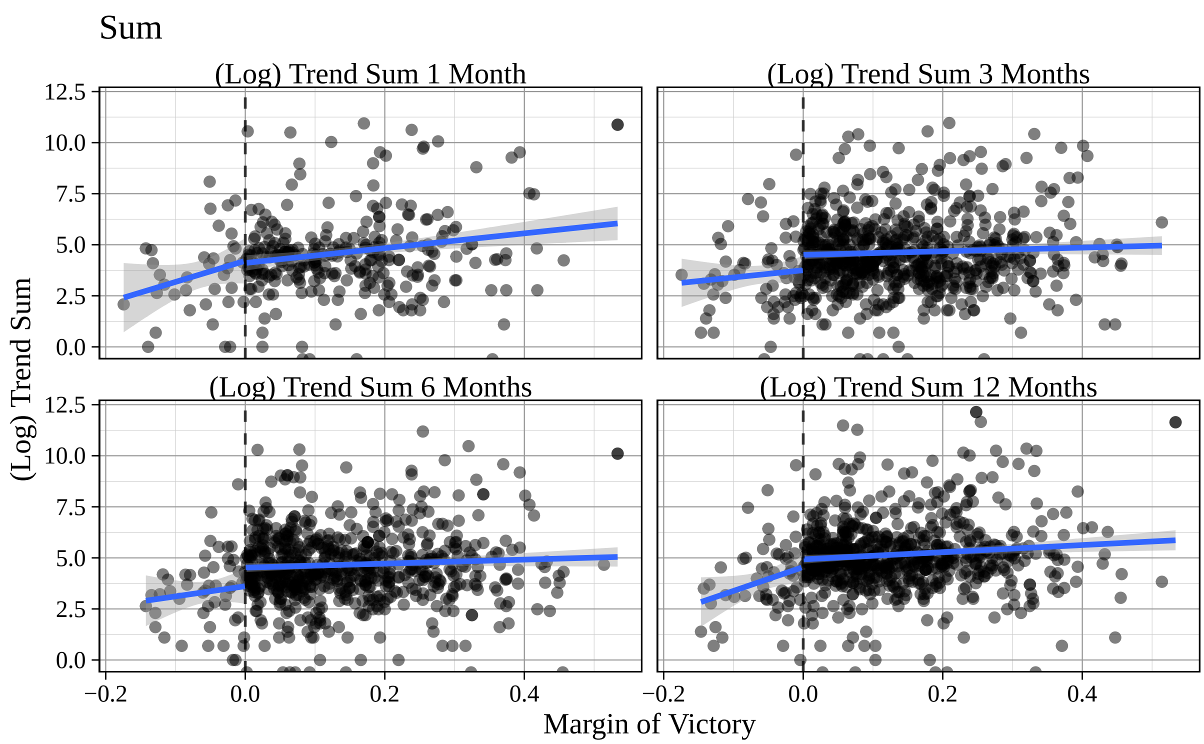


Figure 5: Scatterplot with Separate Linear Regression (Log Transformation)

Table 6: RDD Mean Trends (Same Slope)

	Trend Mean 1 Month	Trend Mean 3 Months	Trend Mean 6 Months	Trend Mean 12 Months
Intercept	−19.661	4.644	−6.669	26.648
	(25.432)	(18.148)	(8.561)	(23.975)
LATE	−13.276	−2.457	18.324	−32.298
	(37.901)	(27.605)	(12.221)	(37.025)
Margin of Victory	−123.540	−2.667	−81.519	583.520
	(320.517)	(286.292)	(69.669)	(421.641)
Num.Obs.	513	428	691	376
R2	0.002	0.000	0.003	0.006
R2 Adj.	−0.002	−0.005	0.000	0.000
AIC	6937.6	5434.0	8021.9	4938.2
BIC	6954.6	5450.2	8040.0	4953.9
Log.Lik.	−3464.798	−2713.001	−4006.945	−2465.098
F	0.596	0.012	1.130	1.035
RMSE	208.09	137.46	80.00	170.92

Note: Standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01, **** p<0.001.

Table 7: RDD Sum Trends (Same Slope)

	Trend Sum 1 Month	Trend Sum 3 Months	Trend Sum 6 Months	Trend Sum 12 Months
Intercept	−714.936	140.465	−200.072	826.203
	(780.645)	(562.520)	(256.818)	(743.220)
LATE	−128.853	−64.992	549.740	−1001.400
	(1167.596)	(855.675)	(366.639)	(1147.759)
Margin of Victory	−6761.744	−186.070	−2445.608	18 090.137
	(10 179.563)	(8874.172)	(2090.077)	(13 070.875)
Num.Obs.	499	428	691	376
R2	0.003	0.000	0.003	0.006
R2 Adj.	−0.001	−0.005	0.000	0.000
AIC	10 153.8	8373.4	12 722.3	7520.6
BIC	10 170.6	8389.6	12 740.5	7536.3
Log.Lik.	−5072.884	−4182.705	−6357.172	−3756.277
F	0.716	0.011	1.130	1.035
RMSE	6311.81	4260.97	2399.94	5298.43

Note: Standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01, **** p<0.001.

Table 8: RDD Mean Trends (Separate Slopes)

	Trend Mean 1 Month	Trend Mean 3 Months	Trend Mean 6 Months	Trend Mean 12 Months
Intercept	−14.552	9.708	−1.522	15.875
	(38.542)	(28.332)	(12.832)	(38.221)
LATE	−17.089	−6.271	14.277	−24.184
	(43.655)	(32.122)	(14.351)	(43.312)
Margin of Victory (Control)	1.719	147.589	12.425	244.966
	(778.647)	(705.761)	(187.813)	(1025.622)
Margin of Victory (Treatment)	−150.872	−179.930	−108.952	407.604
	(854.555)	(772.313)	(202.259)	(1125.363)
Num.Obs.	513	428	691	376
R2	0.002	0.000	0.004	0.006
R2 Adj.	−0.003	−0.007	−0.001	−0.002
AIC	6939.6	5435.9	8023.6	4940.1
BIC	6960.8	5456.2	8046.3	4959.7
Log.Lik.	−3464.782	−2712.974	−4006.799	−2465.032
F	0.407	0.026	0.849	0.732
RMSE	208.29	137.62	80.04	171.12

Note: Standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01, **** p<0.001.

Table 9: RDD Sum Trends (Separate Slopes)

	Trend Sum 1 Month	Trend Sum 3 Months	Trend Sum 6 Months	Trend Sum 12 Months
Intercept	−425.113	300.961	−45.649	492.384
	(1179.095)	(878.195)	(384.969)	(1184.864)
LATE	−342.809	−185.862	428.325	−749.967
	(1338.167)	(995.668)	(430.539)	(1342.676)
Margin of Victory (Control)	463.075	4575.253	372.761	7599.920
	(24 256.329)	(21 876.318)	(5634.378)	(31 794.286)
Margin of Victory (Treatment)	−8772.639	−5701.634	−3268.609	12 629.749
	(26 728.630)	(23 939.224)	(6067.762)	(34 886.243)
Num.Obs.	499	428	691	376
R2	0.003	0.000	0.004	0.006
R2 Adj.	−0.003	−0.007	−0.001	−0.002
AIC	10 155.7	8375.4	12 724.1	7522.4
BIC	10 176.7	8395.6	12 746.7	7542.1
Log.Lik.	−5072.830	−4182.676	−6357.027	−3756.211
F	0.512	0.027	0.849	0.732
RMSE	6317.49	4265.71	2401.17	5304.61

Note: Standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01, **** p<0.001.