Experimental Methods: Lecture 6

Social Media and Experimental Design

Raymond Duch

June 2, 2021

University of Oxford

Road Map

- Digital Trace
- Behavioral Outcomes
- Embedded experiments
- Treatment Delivery
- Twitter data collection

Digital Trace and Machine Learning with Post-Stratification to Measure Public Opinion [Cerina and Duch, 2020]

Digital Vote Overview

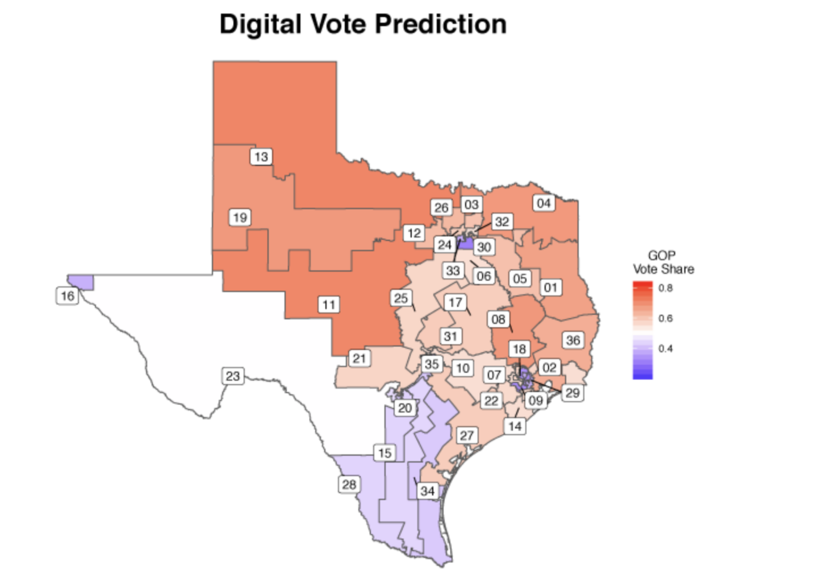
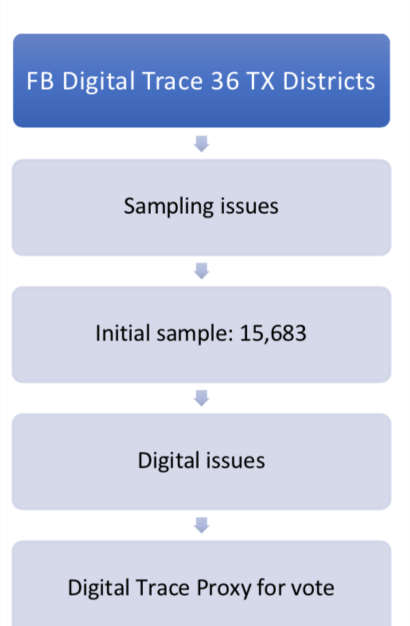
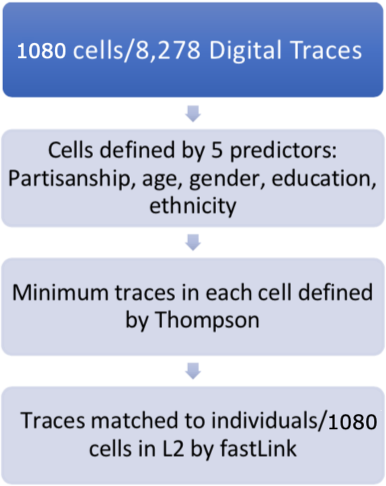


Figure 1: Digital Vote predicted GOP Percent Vote Share.

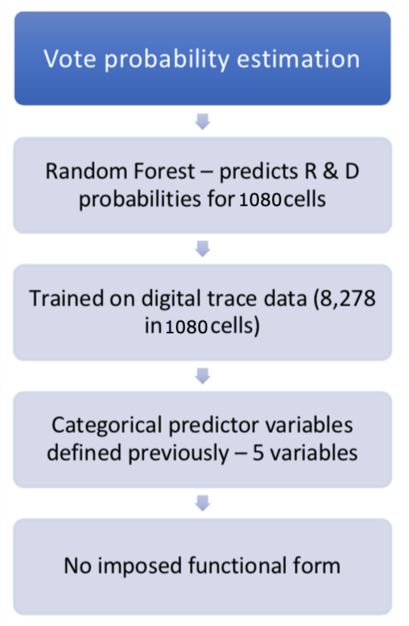
Digital Vote Overview



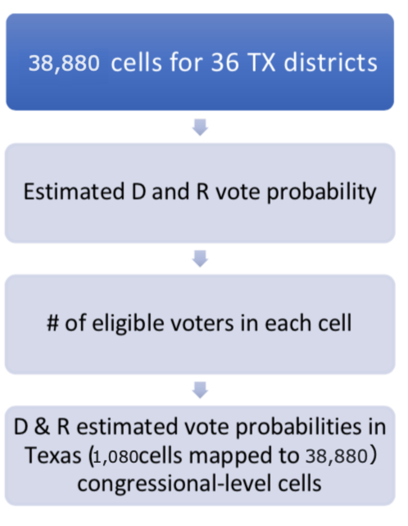
Digital Vote Overview



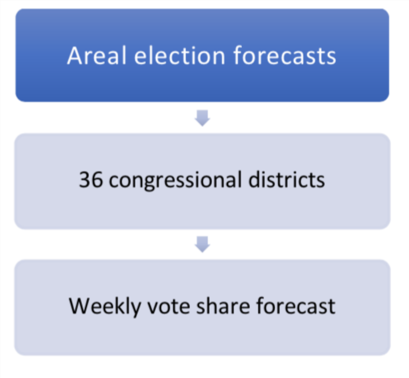
Digital Vote Overview



Digital Vote Overview



Digital Vote Overview



Sampling

Sampling

- Tune sample-size to capture differences in prevalence of voting-groups;
- theoretical sampling distribution:

$$\boldsymbol{n} \sim \text{Multinomial}(p_{1,r=1}, \ldots, p_{G,r=1}, p_{1,r=0}, \ldots, p_{G,r=0}, N); \quad (1)$$

- *worst-case-scenario* sampling for a multinomial distribution following the recommendations of Thompson[Thompson, 1987];
- digital sample is non-probability, but we use theoretical probability distribution as benchmark to guide our power calculation

Sampling

- find N such that we have probability of at least 0.9 that all estimates of the multinational parameters are within 0.025 of the population proportions;
- ensures sample will contain representatives from groups which are 2.5% of voting population or higher;
- N = 1610 per week[Thompson, 1987]

Facebook Sampling

Sampling

- monitor 7 *weeks-to-election-day*;
- assume each multinomial parameter is time-independent;
- sample N for every week within our monitoring frame, for a total of $N = 11270$;
- effective weekly sample size must be further inflated to account for sample loss by poor matching with the voter registry;
- parameter independence is likely true only for swing-voters who are affected by the election-cycle - too harsh for other categories.

Matching

Matching

- get rich voter profiles by matching our social media sample to the state-wide Texas L2 voter registration file;
- individual-level data for over 13 million registered Texas residents, ranging from voting history to socio-economic and demographic characteristics;
- use `R` package `fastLink`[Enamorado et al., 2018] to match our virtual sample to the voter registration list, based on names, *sex* and *city of residency*.

Matching

Threshold	0.75	0.85	0.9	0.95	0.99
Match Count	8739	8394	8384	8283	7322
Match Rate (%)	54.59	52.89	52.84	52.24	46.31
FDR (%)	2.03	1.18	1.16	1.08	0.81

Table 1: `fastLink` output summary table.

- From $15,683$ digital traces in the sample, we successfully match $8,278$;
- success rate of 52.8%;
- these come from $4,475$ distinct registered users;
- `fastlink` reveals $1,417$ borderline cases.

Matching

Conditional on being a registered partisan:

- $\text{P(Dig. Rep.}|\text{Reg. Rep.)} = 0.82$
- $\text{P(Dig. Dem.}|\text{Reg. Dem.)} = 0.82$

Conditional on having attended a partisan primary:

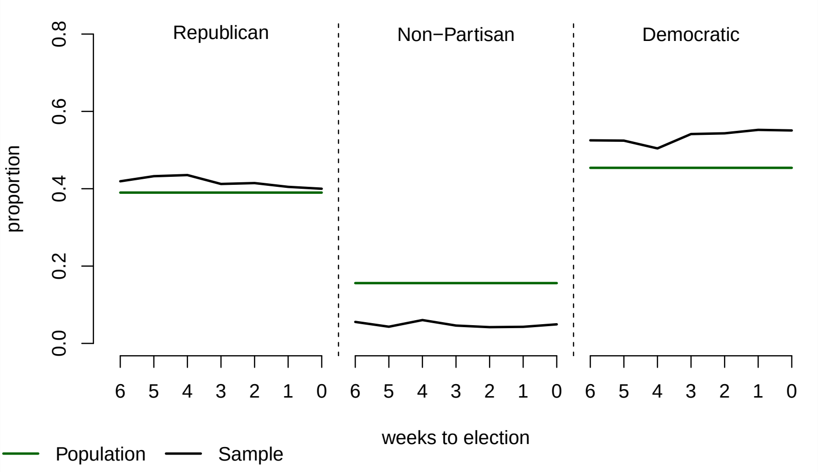
- $\text{P(Dig. Rep.}|\text{Rep. Primary)} = 0.86$
- $\text{P(Dig. Dem.}|\text{Dem. Primary)} = 0.91$

We are comfortable concluding that digital traces represent revealed preferences.

Matching

- final sample size is $8,278$ digital traces;
- slightly below the Thompson number;
- holding probability constant at 0.9, powered to estimate voter categories larger than 2.9% percent of the total population.

Sample v. Population



Vote-Share Estimation

Vote-Share Estimation

Three steps to obtaining area-estimates of vote-share:

- identification of voter characteristics from L2 file to use in vote choice prediction model (voter-categories);
- estimation, at the voter-category level, of vote choice conditional on the probability of vote turnout;
- weighting predictions by cell counts and summing over the area of interest to recover estimates of support.

Vote-Share Estimation

- select just under 50 variables from VR file for turnout prediction at the individual level;
- Missing values in $\boldsymbol{X}$ are imputed with a random-forest multiple-imputation strategy implemented via the packages `ranger`[Wright and Ziegler, 2015] and `missForest`[Stekhoven and Bühlmann, 2011];
- further select subset of variables for vote-choice model Z - we are not powered to include all the variables as we do for turnout;
- define the voter category C_g, for categories $g = 1, ..., G$ as a unique realization of the set of variables which compose $\boldsymbol{Z}$, i.e. $C_g = \{Z_1 = z_1, ..., Z_m = z_m\}$.

Vote-Share Estimation

- Estimate the joint probability of a voter-category supporting the Republican candidate and turning out on election-day;
- decompose the problem similarly to Lauderdale et al.[Lauderdale et al., 2017]:

$$P_g\left(R=1, T=1|C\right) = P_g\left(R=1|T=1, C\right) \times P_g\left(T=1|C\right) \tag{2}$$

.

Vote-Share Estimation

- train a random forest on the expanded (long) turnout dataset $\boldsymbol{X}^{+}$;
- output a prediction probability of turning out for each member of the voting population:

$$\hat{\mathrm{P}}_h\left(T=1|\boldsymbol{x}\right)=\varphi^T\left(\boldsymbol{x}_h\right); \tag{3}$$

- 1) extract turnout probabilities from matched sample $\hat{\mathrm{P}}_s\left(T=1|\boldsymbol{x}\right)$;
- 2) average within voter categories identified by $\boldsymbol{C}$ to obtain category-level estimates of turnout probabilities:

$$\mathrm{P}_g\left(T=1|C\right). \tag{4}$$

Vote-Share Estimation

- calculate global empirical error distribution using the MSPE1 procedure[Lu, 2017];
- Normal distribution does not characterize the empirical distribution perfectly; it is still useful to obtain reasonable prediction intervals:

$$\mathrm{P}_h\left(T = 1|\boldsymbol{x}\right) \sim \mathrm{N}\left(\varphi^T\left(\boldsymbol{x}_h^{\mathsf{I}}\right), (\hat{\sigma}^T_{\text{RMSE1}})^2\right). \tag{5}$$

Likelihood Estimation

Likelihood Estimation

- train a probability machine as above to estimate the probability of voting Republican, conditional on the individual turning out and the set of their voter characteristics:
$$\hat{\mathrm{P}}_s\left(R = 1 | T = 1, \boldsymbol{z}\right) = \varphi^R\left(\boldsymbol{z}_s | \hat{\mathrm{P}}_s\left(T = 1 | \boldsymbol{x}\right)\right); \qquad (6)$$

- output category-level predictions; for category g such that $s \in g$ if $\boldsymbol{z}_s = C_g$:
$$\hat{\mathrm{P}}_g\left(R = 1 | T = 1, C\right) = \varphi^R\left(C_g\right). \qquad (7)$$

- estimate the global error via MSPE1 and approximate distribution with Normal density:

Area Estimation

Area Estimation

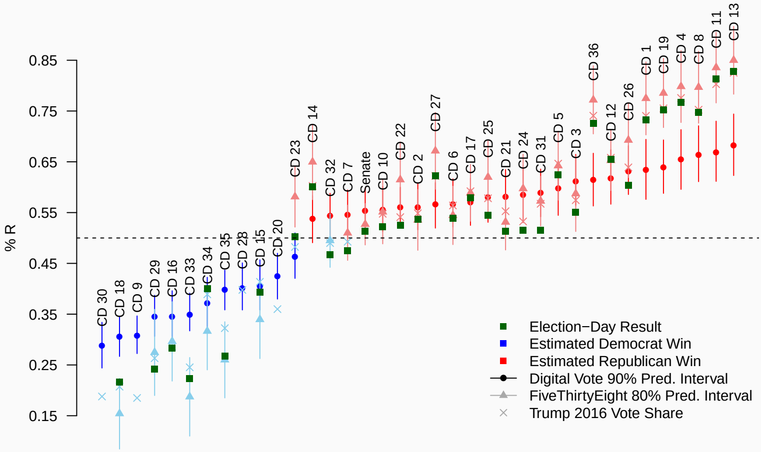
Senate Election (similar for Congress):

- State cell counts: $Q_g = \sum_h 1(\boldsymbol{z}_h = C_g)$;
- Republican support Senate estimate:

$$V^R_{gw} = \frac{\sum_g \mathrm{P}_g \left(R = 1, T = 1 | C, L = 1, W = w\right) \times Q_g}{\sum_g \mathrm{P}_g \left(T = 1 | C, L = 1, W = w\right) \times Q_g}; \tag{9}$$

Results

Results

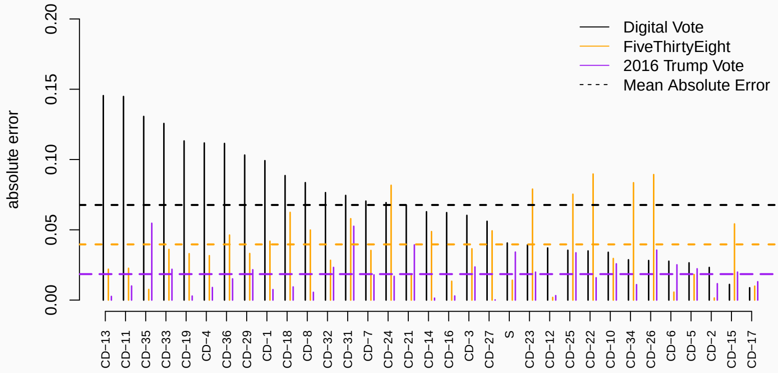


% R
0.85
0.75
0.65
0.55
0.45
0.35
0.25
0.15
CD 30
CD 18
CD 9
CD 29
CD 16
CD 33
CD 34
CD 35
CD 28
CD 15
CD 20
CD 23
CD 14
CD 32
CD 7
Senate
CD 10
CD 22
CD 2
CD 27
CD 6
CD 17
CD 25
CD 21
CD 24
CD 31
CD 5
CD 3
CD 36
CD 12
CD 26
CD 1
CD 19
CD 4
CD 8
CD 11
CD 13
Election−Day Result
Estimated Democrat Win
Estimated Republican Win
Digital Vote 90% Pred. Interval
FiveThirtyEight 80% Pred. Interval
Trump 2016 Vote Share

Results

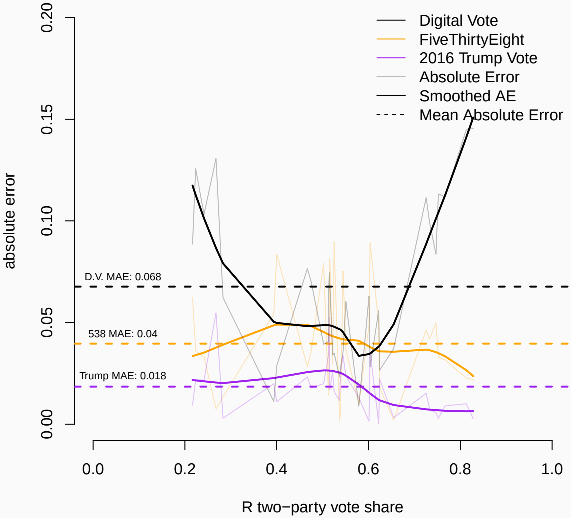
- we produce 'hypothetical' estimates for all districts;
- no race-specific effects due to systematic non-response;
- there is evident attenuation bias;
- directionally, our estimates are correct for all but 3 districts (7 and 32 and 23), and correct for the senate;
- on absolute error `FiveThirtyEight` outperforms.

Results



Digital Vote
FiveThirtyEight
2016 Trump Vote
Mean Absolute Error
absolute error
0.00
0.05
0.10
0.15
0.20
CD−13
CD−11
CD−35
CD−33
CD−19
CD−4
CD−36
CD−29
CD−1
CD−18
CD−8
CD−32
CD−31
CD−7
CD−24
CD−21
CD−14
CD−16
CD−3
CD−27
S
CD−23
CD−12
CD−25
CD−22
CD−10
CD−34
CD−26
CD−6
CD−5
CD−2
CD−15
CD−17

Results



Digital Vote
FiveThirtyEight
2016 Trump Vote
Absolute Error
Smoothed AE
Mean Absolute Error
D.V. MAE: 0.068
538 MAE: 0.04
Trump MAE: 0.018
absolute error
0.00
0.05
0.10
0.15
0.20
0.0
0.2
0.4
0.6
0.8
1.0
R two−party vote share

Sub-Category Predictions

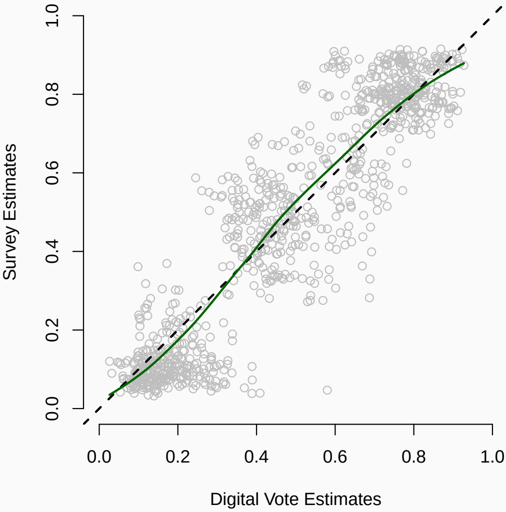


Figure 1: Voter sub-categories comparisons: Surveys v. Digital

Conclusions

Conclusions

- our method is a novel variation on MRP, which exploits Random Forests to gain generalizability (via capturing non-linear relationships) and computing speed;
- our sampling strategy suggests careful sampling on social media with respect to politics can be quite representative of the population of interest;
- digital traces are shown to be highly correlated with partisanship, and hence valid as measures of political support.

Social Media Treatment + Behavioral Outcomes

Online Political Information [Enrique et al., 2019]

- Outcome: Vote for the incumbent party in Mexico municipal elections
- Treatment: Online malfeasance video campaign
- Design: Saturation

Information Treatment via Social Media

- 61 million person experiment [Bond et al., 2012]
- Twitter India Vaccines: [Alatas et al., 2019]
- Information Architecture Malfeasance: [Duch and Torres, 2021]
- Vaccine Compliance and Incentives: (Duch et al 2021)

[Enrique et al., 2019] Design: Saturation

- Randomly assign municipalities to 20% % 80% saturation
- Block random assignment:
  - 128 Municipalities
  - 42 Blocks – 3 municipalities in each block have similar incumbent governing parties
- Within each block the municipalities are randomly assigned to three treatments:
  - Control: No FB ad
  - Low saturation: FB information ad
  - High saturation: FB information ad
- Within each municipality:
  - 5 equally-populated segments
  - 20% saturation - 1 in 5 assigned FB info
  - 80% saturation - 4 in 5 assigned FB info

[Enrique et al., 2019] Design: Saturation

$$Y_{psm} = \alpha Y_{psm}^{lag} + \beta Facebook\ ads_{sm} + \gamma Spillover_{sm} + \mu_b + \varepsilon_{psm},$$

[]

$$\begin{aligned} Y_{psm} \quad = \quad & \alpha Y_{psm}^{lag} + \beta_1 Facebook\ ads\ in\ Low\ Saturation_{sm} + \beta_2 Facebook\ ads\ in\ High\ Saturation_{sm} \\ & + \gamma_1 Spillover\ in\ Low\ Saturation_{sm} + \gamma_2 Spillover\ in\ High\ Saturation_{sm} + \mu_b + \varepsilon_{psm},\ (2) \end{aligned}$$

[]

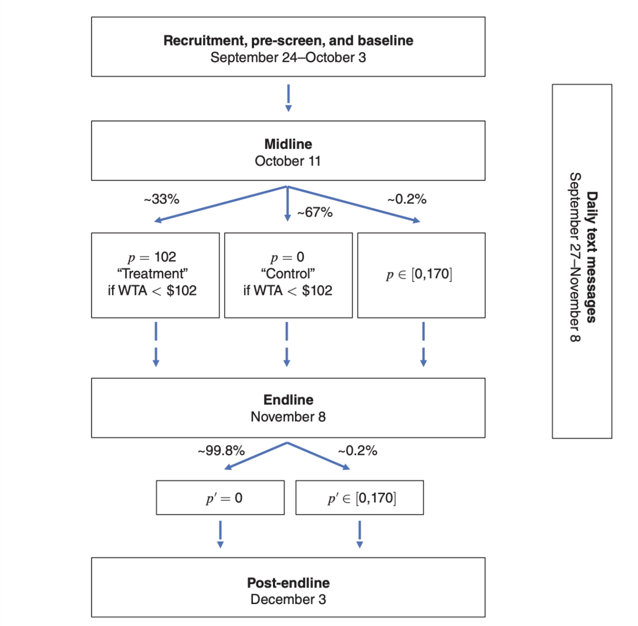
Design: Spillover

- Estimation issues: [Baird et al., 2018]
- Estimation issues: [Aronow et al., 2021]
- Cash Transfer Kenya: [Haushofer and Shapiro, 2016]
- School instruction interventions [Paluck et al., 2016]

Welfare Effects of FB [Allcott et al., 2020]

- Outcome: WTA
- Treatment: FB Deactivation
- Design: Online recruitment and random assignment to T and C

Welfare Effects of FB [Allcott et al., 2020]



Welfare Effects of FB [Allcott et al., 2020]

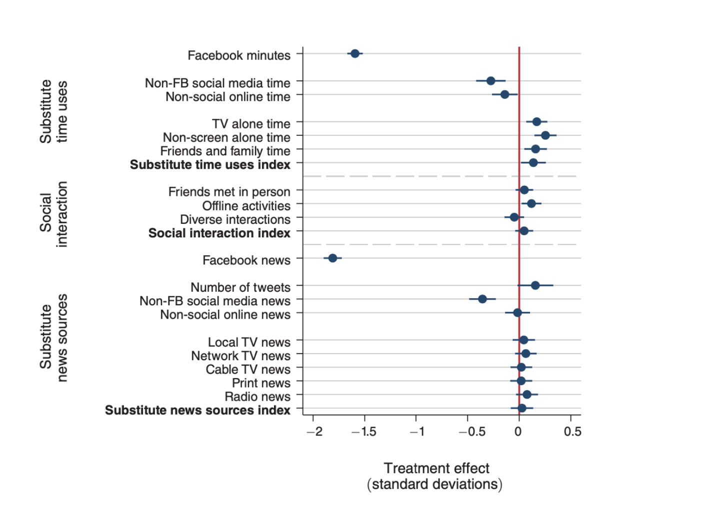
TABLE 1—SAMPLE SIZES

Phase	Sample size
Recruitment and baseline	N = 1,892,191 were shown ads N = 32,201 clicked on ads N = 22,324 completed pre-screen survey N = 20,959 were from United States and born between 1900 and 2000 N = 17,335 had $15 <$ daily Facebook minutes ≤ 600 N = 7,455 consented to participate N = 3,910 finished baseline N = 2,897 had valid baseline and were randomized, of which:
Midline	N = 2,897 began midline N = 2,743 received a price offer, of which: N = 1,661 were in impact evaluation sample
Endline	N = 2,710 began endline N = 2,684 finished endline, of which: N = 1,637 were in impact evaluation sample
Post-endline	N = 2,067 reported Facebook mobile app use, of which: N = 1,219 were in impact evaluation sample

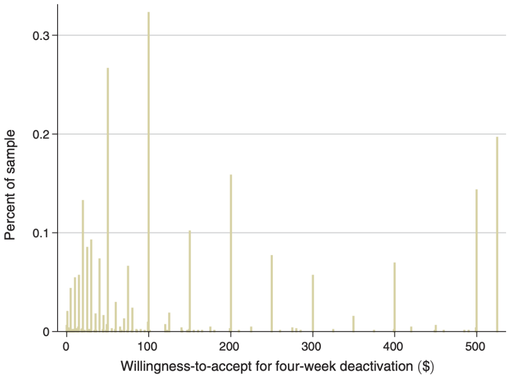
LATE [Allcott et al., 2020]

$$Y_i = \tau D_i + \phi Y_{bi} + \sigma_s + \mu_i \tag{10}$$

Welfare Effects of FB [Allcott et al., 2020]



Welfare Effects of FB [Allcott et al., 2020]



Welfare Effects of FB [Allcott et al., 2020]

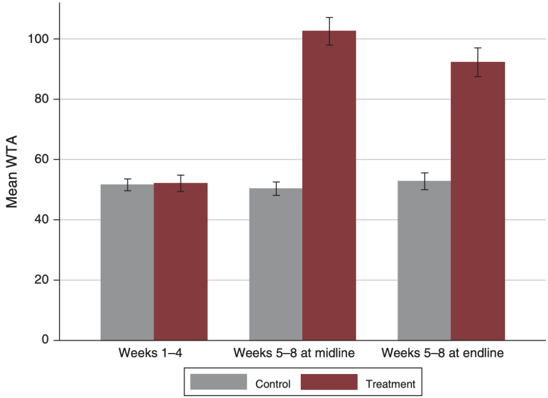


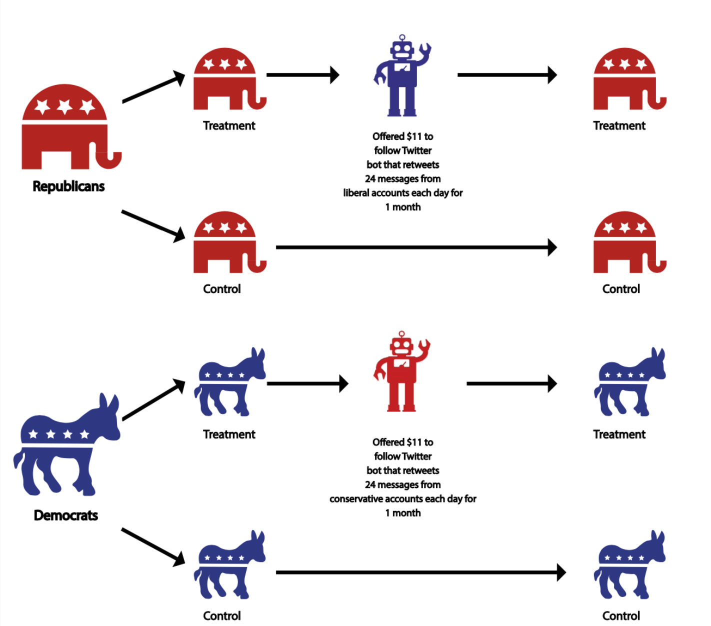
FIGURE 12. AVERAGE VALUATION OF FACEBOOK IN TREATMENT AND CONTROL

Embedded Experiments

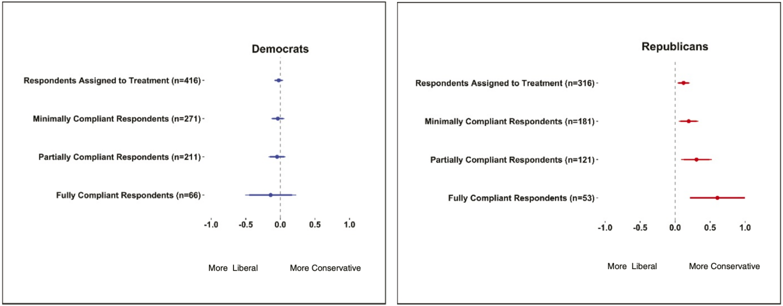
Embedded Experiments

- Outcome Unobtrusive
    - Digital Trace
    - Digital Choice
- Outcome Obtrusive
    - Attitudes/Preferences
    - Hypothetical choices
- Treatments
    - Tasks
    - Vignettes/Conjoints
    - Video

Polarization Twitter [Bail et al., 2018]



Polarization Twitter [Bail et al., 2018]



Embedded Experiments: Examples

- [Coppock et al., 2016]; When Treatments are Tweets
- Duch et al 2021 Vaccine Incentives – Video + Digital Choice

Duch et al Vaccine Incentives

```
https://www.dropbox.com/s/ybxiyowhason2qu/
VID-20210601-WA0011.mp4?dl=0
```

Treatment Delivery: WhatsApp

Figure 3: Screen shoots - Negative Standard Temporal treatment

Figure 4: Screen shoots - Positive Severity Temporal treatment

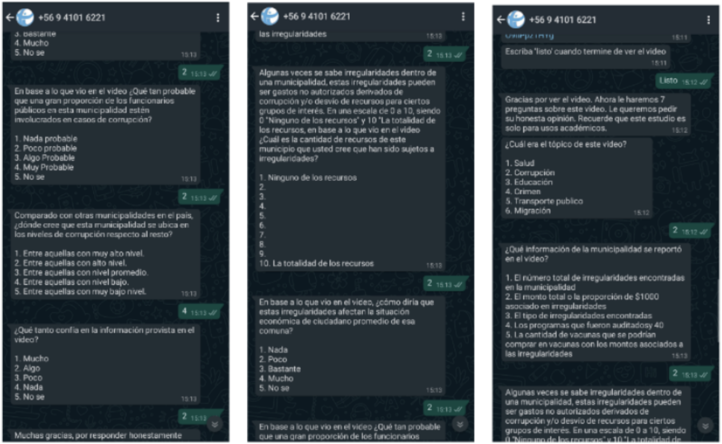
Figure 5: Screen shoots - Positive Resources Temporal treatment

Figure 6: Screen shoots - Positive Standard Spatial treatment

Figure 7: Screen shoots - Positive Resources Foregone treatment

Figure 8: Screen shoots - Negative Program Temporal treatment

Figure 9: Screen shoots - Chatbot survey



Twitter Data Collection

References i

- Alatas, V., Chandrasekhar, A. G., Mobius, M., Olken, B. A., and Paladines, C. (2019).
  When celebrities speak: A nationwide twitter experiment promoting vaccination in indonesia.
  Working Paper 25589, National Bureau of Economic Research.
- Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. (2020).
  The welfare effects of social media.
  American Economic Review, 110(3):629–76.

References ii

Aronow, P., Eckles, D., Samii, C., and Zonszein, S. (2021).
Spillover Effects in Experimental Data, page 526–543.
Cambridge University Press.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., and Volfovsky, A. (2018).
Exposure to opposing views on social media can increase political polarization.

Proceedings of the National Academy of Sciences, 115(37):9216–9221.

Baird, S., Bohren, J. A., McIntosh, C., and Ozler, B. (2018).
Optimal design of experiments in the presence of interference.
The Review of Economics and Statistics, 100(5):844–860.

References iv

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., and Fowler, J. H. (2012).
A 61-million-person experiment in social influence and political mobilization.
Nature, 489:295 EP –.

Cerina, R. and Duch, R. (2020).
Measuring public opinion via digital footprints.
International Journal of Forecasting.

- Coppock, A., Guess, A., and Ternovski, J. (2016).
When treatments are tweets: A network mobilization experiment over twitter.
Political Behavior, 38(1):105–128.
- Duch, R. and Torres, F. (2021).
Information architecture for corruption messaging.
Working Paper.
- Enamorado, T., Fifield, B., and Imai, K. (2018).
Using a probabilistic model to assist merging of large-scale administrative records.

References vi

Enrique, J. R., Larreguy, H., Marshall, J., and Simpser, A. (2019).
Online political information: Facebook ad saturation and electoral accountability in mexico.
Working paper.

Haushofer, J. and Shapiro, J. (2016).
The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya*.
The Quarterly Journal of Economics, 131(4):1973–2042.

References vii

Lauderdale, B. E., Bailey, D., Blumenau, Y. J., and Rivers, D. (2017).
Model-based pre-election polling for national and sub-national outcomes in the us and uk.
Technical report, Working paper.

Lu, B. (2017).
Constructing Prediction Intervals for Random Forests.
PhD thesis, Pomona College.

Paluck, E. L., Shepherd, H., and Aronow, P. M. (2016). **Changing climates of conflict: A social network experiment in 56 schools.** *Proceedings of the National Academy of Sciences*, 113(3):566–571.

Stekhoven, D. J. and Bühlmann, P. (2011). **Missforest—non-parametric missing value imputation for mixed-type data.** *Bioinformatics*, 28(1):112–118.

References ix

Thompson, S. K. (1987).
Sample size for estimating multinomial proportions.
The American Statistician, 41(1):42–46.

Wright, M. N. and Ziegler, A. (2015).
Ranger: a fast implementation of random forests for high dimensional data in c++ and r.
arXiv preprint arXiv:1508.04409.