Corruption information and vote share: A meta-analysis and lessons for survey experiments

Trevor Incerti

6 September 2019

Prepared for the Yale ISPS Experiments Workshop

Introduction

Research Question

Do voters hold politicians accountable for corruption?

Research Question

Do voters hold politicians accountable for corruption?

- Key question of electoral accountability.

Research Question

Do voters hold politicians accountable for corruption?

- Key question of electoral accountability.
- ARPS review: "Empirical evidence to date is mixed, and it often suggests that the electoral punishment of corruption is rather mild." (De Vries & Solaz 2017)

Research Question

Do voters hold politicians accountable for corruption?

- Key question of electoral accountability.
- ARPS review: "Empirical evidence to date is mixed, and it often suggests that the electoral punishment of corruption is rather mild." (De Vries & Solaz 2017)
- Recent explosion of experimental research on this subject.

Research Question

Do voters hold politicians accountable for corruption?

- Key question of electoral accountability.
- ARPS review: "Empirical evidence to date is mixed, and it often suggests that the electoral punishment of corruption is rather mild." (De Vries & Solaz 2017)
- Recent explosion of experimental research on this subject.
- What have we learned from this research? Is evidence actually mixed?

Methods

Meta-Analysis

Meta-analysis of all experimental studies conducted to date.

Meta-Analysis

Meta-analysis of all experimental studies conducted to date.

- Treatment: corruption information provision.

Meta-Analysis

Meta-analysis of all experimental studies conducted to date.

- Treatment: corruption information provision.
- Outcome: vote-share of corrupt politician.

Meta-Analysis

Meta-analysis of all experimental studies conducted to date.

- Treatment: corruption information provision.
- Outcome: vote-share of corrupt politician.
- Random assignment of corruption information $\rightarrow$ measurement of voting outcomes.

Meta-Analysis

Meta-analysis of all experimental studies conducted to date.

- Treatment: corruption information provision.
- Outcome: vote-share of corrupt politician.
- Random assignment of corruption information $\rightarrow$ measurement of voting outcomes.
- As treatments are not always assigned identically, I take steps to standardize where possible. Details

Meta-Analysis

Meta-analysis of all experimental studies conducted to date.

- Treatment: corruption information provision.
- Outcome: vote-share of corrupt politician.
- Random assignment of corruption information $\rightarrow$ measurement of voting outcomes.
- As treatments are not always assigned identically, I take steps to standardize where possible. Details
- Includes both published articles and working papers.

Meta-Analysis

I use fixed effects and random effects estimation.

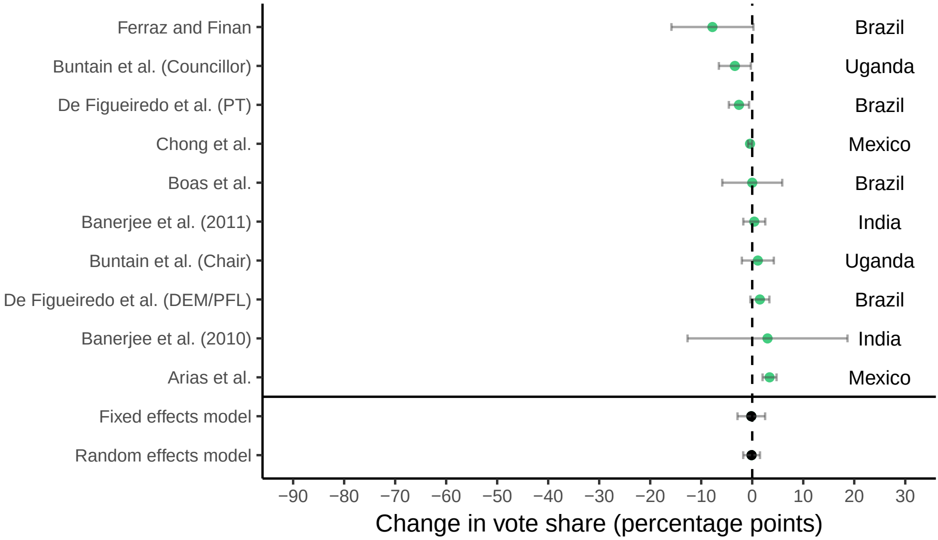
Meta-Analysis

I use fixed effects and random effects estimation.

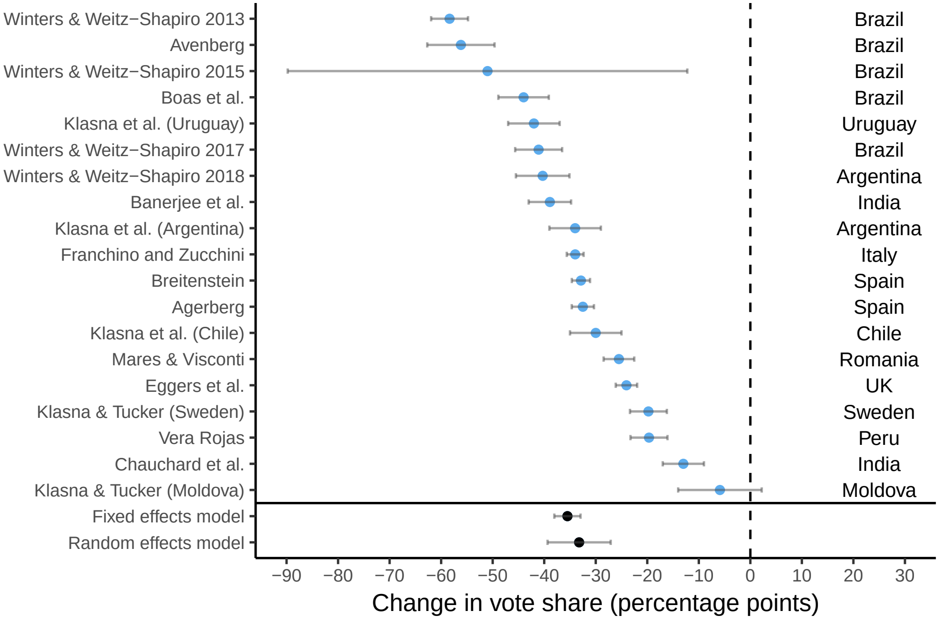
- Random effects likely more appropriate in this case.
    - Fixed effects assumes one true effect size across all studies, with differences in effects due to sampling error. Details
    - Random effects assumes effect sizes vary due to population heterogeneity, differences in treatment, etc.
    - In random effects, effect sizes are therefore assumed to represent a random sample of a distribution of effect sizes.
- In this case, differences in estimated effect size between the two methods are minor.

Results

Results: Field Experiments



Results: Survey Experiments



Results

- Survey experiments estimate much larger negative treatment effects relative to field experiments. Lab experiments similar to survey

Results

- Survey experiments estimate much larger negative treatment effects relative to field experiments. Lab experiments similar to survey
- Corrupt candidates punished by approximately zero percentage points in field experiments.

Results

- Survey experiments estimate much larger negative treatment effects relative to field experiments. Lab experiments similar to survey
- Corrupt candidates punished by approximately zero percentage points in field experiments.
- Punished by between 33 percentage points (random effects) and 35 percentage points (fixed effects) in survey experiments.

Assessing within study heterogeneity

How much heterogeneity across studies is accounted for by type of experiment?

- First estimate total amount of heterogeneity across all studies with random effects model.

Assessing within study heterogeneity

How much heterogeneity across studies is accounted for by type of experiment?

- First estimate total amount of heterogeneity across all studies with random effects model.
- Next, estimate residual heterogeneity after including a categorical indicator for experiment type in the model.

Assessing within study heterogeneity

How much heterogeneity across studies is accounted for by type of experiment?

- First estimate total amount of heterogeneity across all studies with random effects model.
- Next, estimate residual heterogeneity after including a categorical indicator for experiment type in the model.
  - Point estimate of this dummy ($0 =$ field, $1 =$ survey) is equal to -0.33 (significant at 1% level), the same as the overall estimate across all studies (with no moderator). Results

Assessing within study heterogeneity

How much heterogeneity across studies is accounted for by type of experiment?

- First estimate total amount of heterogeneity across all studies with random effects model.
- Next, estimate residual heterogeneity after including a categorical indicator for experiment type in the model.
  - Point estimate of this dummy (0 = field, 1 = survey) is equal to -0.33 (significant at 1% level), the same as the overall estimate across all studies (with no moderator). Results
- Subtract residual heterogeneity from total heterogeneity and divide by total heterogeneity.

Assessing within study heterogeneity

How much heterogeneity across studies is accounted for by type of experiment?

- First estimate total amount of heterogeneity across all studies with random effects model.
- Next, estimate residual heterogeneity after including a categorical indicator for experiment type in the model.
    - Point estimate of this dummy (0 = field, 1 = survey) is equal to -0.33 (significant at 1% level), the same as the overall estimate across all studies (with no moderator). Results
- Subtract residual heterogeneity from total heterogeneity and divide by total heterogeneity.
- 68% of the total heterogeneity across studies accounted for by including a dummy variable for type of experiment.

Discussion

What might account for the discrepancy between field and survey experiments?

What might account for the discrepancy between field and survey experiments?

- Publication bias and/or p-hacking

What might account for the discrepancy between field and survey experiments?

- Publication bias and/or p-hacking
- Social desirability bias

What might account for the discrepancy between field and survey experiments?

- Publication bias and/or p-hacking
- Social desirability bias
- Survey context does not mirror real-world settings:
    - Non-compliance
    - Differences in outcome choices
    - Costliness/decision complexity

Publication bias and p-hacking

No clear evidence of publication bias within survey experiments:

Publication bias and p-hacking

No clear evidence of publication bias within survey experiments:

- P-curve - virtually all results significant at 1% level (not clustered around 0.05 or 0.01).
- Tests for funnel plot asymmetry. Figures
- P-value not a significant predictor of publication status. Table

Publication bias and p-hacking

No clear evidence of publication bias within survey experiments:

- P-curve - virtually all results significant at 1% level (not clustered around 0.05 or 0.01).
- Tests for funnel plot asymmetry. Figures
- P-value not a significant predictor of publication status. Table

Unclear with field experiments

- Five of eight papers published. Three unpublished papers all have null findings. Figure

Publication bias and p-hacking

No clear evidence of publication bias within survey experiments:

- P-curve - virtually all results significant at 1% level (not clustered around 0.05 or 0.01).
- Tests for funnel plot asymmetry. Figures
- P-value not a significant predictor of publication status. Table

Unclear with field experiments

- Five of eight papers published. Three unpublished papers all have null findings. Figure
- Not enough data for formal tests.

Social desirability bias

Social desirability bias

- Anti-corruption norms exist in most countries.

Social desirability bias

- Anti-corruption norms exist in most countries.
- No costs to selecting the socially desirable option in hypothetical vignette.

Social desirability bias

- Anti-corruption norms exist in most countries.
- No costs to selecting the socially desirable option in hypothetical vignette.
- Voting against corruption in the abstract may therefore reflect the respondents' actual preferences.

Social desirability bias

- Anti-corruption norms exist in most countries.
- No costs to selecting the socially desirable option in hypothetical vignette.
- Voting against corruption in the abstract may therefore reflect the respondents' actual preferences.
- In actual election voters may discount information, or have strong material/ideological incentives to stick with candidate.

Differences in experimental context: non compliance

Treatments are weak and easily missed in field experiments.

- In survey experiments ITT = ATE = CACE (LATE)
- Field experiments measure ITT as they do not know the non compliance rate. Non compliance necessarily reduces the ITT.
    - $ITT = CACE \times \pi_C$

Differences in experimental context: outcome choice

Choice set offered to voters is not necessarily identical across experiments. Example:

- Field experiment: Candidate A is randomly revealed to be corrupt, and voters can cast vote for corrupt candidate A, or candidate B, who may be clean or corrupt.

- Survey experiment: Candidate A is randomly revealed to be corrupt, and voters can cast vote for corrupt candidate A, or counterfactual Candidate A who *is not* corrupt.

Differences in experimental context: complexity, costliness and conjoint experiments

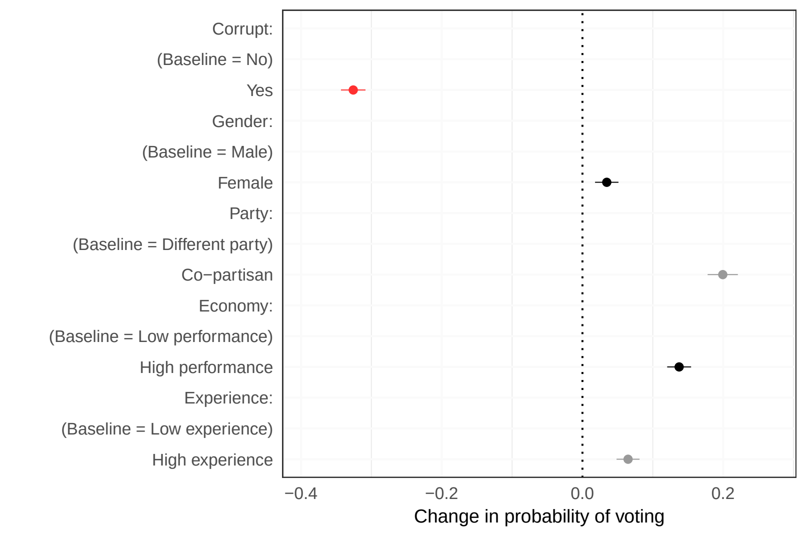
Conjoint experiments: Randomizing more candidate characteristics may capture variety of moderating factors and reduce social desirability bias.

Differences in experimental context: complexity, costliness and conjoint experiments

Conjoint experiments: Randomizing more candidate characteristics may capture variety of moderating factors and reduce social desirability bias.

- But, traditional method of analysis (comparing magnitudes of individual average marginal component effects) may be misleading.

Differences in experimental context: complexity, costliness and conjoint experiments



Differences in experimental context: complexity, costliness and conjoint experiments

Proposal: When researchers have strong theories about the conditions that shape voter decision-making, a more appropriate method may be to calculate average marginal effects in order to present predicted probabilities of voting for a candidate under these conditions.

- E.g. Compare the probability of voting for a realistic candidate with outlier characteristics such as corruption to the probability of voting for a realistic candidate without this characteristic. Example 1 Example 2 Example 3

Differences in experimental context: complexity, costliness and conjoint experiments

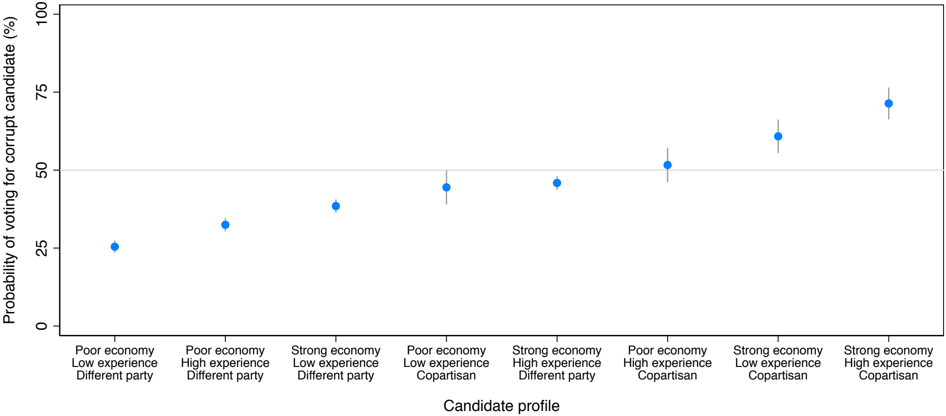


Figure 2: Breitenstein (2019) conjoint: can the right candidate overcome corruption?

Differences in experimental context: complexity, costliness and conjoint experiments

Proposal: When we do not have strong theories about the conditions that shape voter decision-making, we can use regression trees to illuminate them.

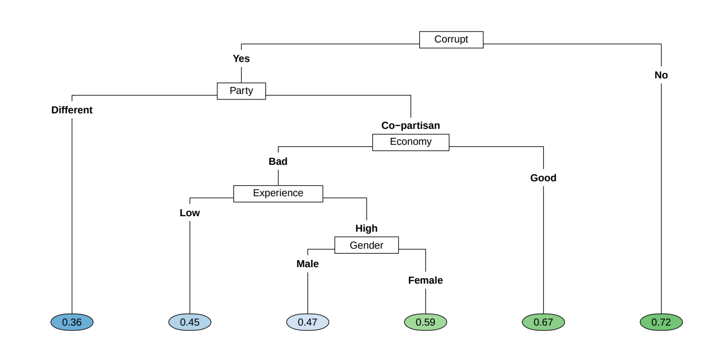


Figure 3: Breitenstein (2019) conjoint decision tree: predicted probabilities of voting for candidate

Conclusion

Conclusion

Conclusion

- Effect of corruption information on vote-choice differs drastically between field and survey experiments.

Conclusion

- Effect of corruption information on vote-choice differs drastically between field and survey experiments.
  - Zero in field experiments.
  - -33 to -35 percentage points in survey experiments.

Conclusion

- Effect of corruption information on vote-choice differs drastically between field and survey experiments.
  - Zero in field experiments.
  - -33 to -35 percentage points in survey experiments.
- Discrepancy does not seem to be driven by publication bias/p-hacking.

Conclusion

- Effect of corruption information on vote-choice differs drastically between field and survey experiments.
  - Zero in field experiments.
  - -33 to -35 percentage points in survey experiments.
- Discrepancy does not seem to be driven by publication bias/p-hacking.
- May arise from:

Conclusion

- Effect of corruption information on vote-choice differs drastically between field and survey experiments.
  - Zero in field experiments.
  - -33 to -35 percentage points in survey experiments.
- Discrepancy does not seem to be driven by publication bias/p-hacking.
- May arise from:
  - Social desirability bias

Conclusion

- Effect of corruption information on vote-choice differs drastically between field and survey experiments.
  - Zero in field experiments.
  - -33 to -35 percentage points in survey experiments.
- Discrepancy does not seem to be driven by publication bias/p-hacking.
- May arise from:
  - Social desirability bias
  - Survey context failing to mirror real-world settings:
    - Non-compliance
    - Differences in outcome choices
    - Costliness/decision complexity

Conclusion

- Vote-choice survey experiments may provide information on the directionality of informational treatments in hypothetical scenarios, but point estimates they provide may not be representative of real-world voting behavior.

Conclusion

- Vote-choice survey experiments may provide information on the directionality of informational treatments in hypothetical scenarios, but point estimates they provide may not be representative of real-world voting behavior.
- Researchers should exercise caution when interpreting actions taken in hypothetical vignettes as indicative of real world behavior such as voting.

Feedback

- Other possible reasons behind the discrepancy?

Feedback

- Other possible reasons behind the discrepancy?
- Are empirical tests of the reasons I have already identified necessary?

Feedback

- Other possible reasons behind the discrepancy?
- Are empirical tests of the reasons I have already identified necessary?
- If so, what tests would you find convincing?

Supplemental material

Analytical details Back

- Where there are multiple corruption treatments (e.g. varying source of information), I replicate the studies and code corruption as a binary treatment ($0 =$ clean, $1 =$ corrupt).

Analytical details Back

- Where there are multiple corruption treatments (e.g. varying source of information), I replicate the studies and code corruption as a binary treatment ($0 =$ clean, $1 =$ corrupt).
- Studies that use non-binary vote choices are rescaled into a binary vote choice.

Analytical details Back

- Where there are multiple corruption treatments (e.g. varying source of information), I replicate the studies and code corruption as a binary treatment ($0 =$ clean, $1 =$ corrupt).
- Studies that use non-binary vote choices are rescaled into a binary vote choice.
- Point estimates, standard errors and/or confidence intervals are not always explicitly reported (4 cases). In these cases standard errors are estimated by digitally measuring coefficient plots.

Analytical details Back

- Where there are multiple corruption treatments (e.g. varying source of information), I replicate the studies and code corruption as a binary treatment (0 = clean, 1 = corrupt).
- Studies that use non-binary vote choices are rescaled into a binary vote choice.
- Point estimates, standard errors and/or confidence intervals are not always explicitly reported (4 cases). In these cases standard errors are estimated by digitally measuring coefficient plots.
- Two field experiments include general anti-corruption treatments not specific to candidates. Robustness check excludes these studies.

Fixed versus random effects estimation Back

Fixed effects:

$\hat{\theta} = \frac{\sum w_i \theta_i}{\sum w_i}$ where $w_i = \frac{1}{var_i}$

Random effects:

$\hat{\theta} = u + u_i$ where $u_i \sim N(0, \tau^2)$ and:

u is equal to the average "true effect", and τ^2 is the heterogeneity amongst true effects.

Lab experiments Back

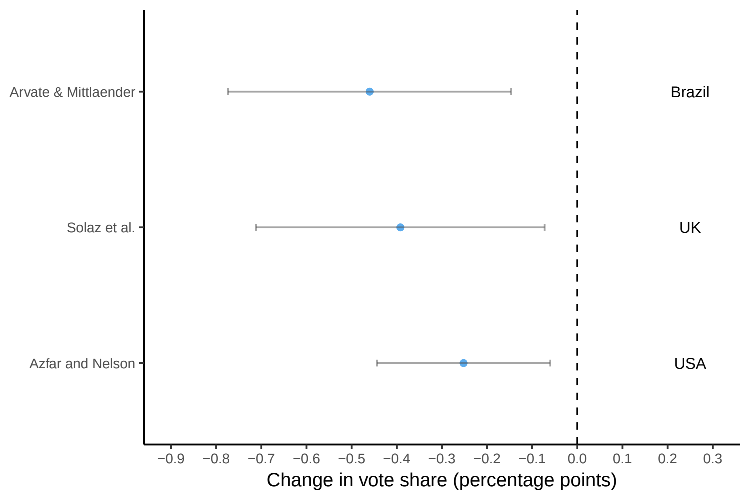


Figure 4: Lab experiments: Average treatment effect of corruption information on vote share

Robustness checks

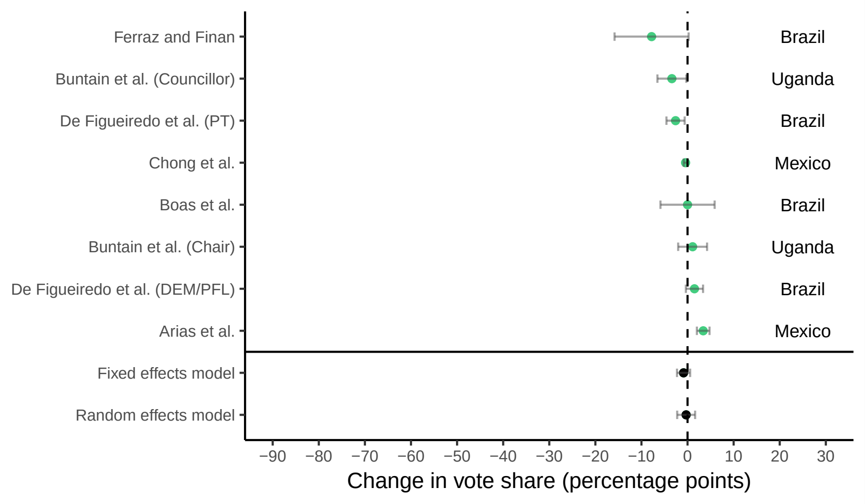


Figure 5: Field experiments: Average treatment effect of corruption information on incumbent vote share (excluding Banerjee et al. (2010) and Banerjee et al. (2011))

Mixed effects meta-analysis with survey experiment moderator

Back

Table 1: Mixed effects meta-analysis with survey experiment moderator

Value	Estimate
Constant	-0.005
	(0.035)
Survey experiment moderator	-0.326
	(0.043)

Note: Standard errors in parenthesis.

Funnel plot asymmetry

Back

Table 2: Regression tests for funnel plot asymmetry

Studies included	p value
All	0.0004
All with moderator	0.765
Field	0.840
Survey	0.630

Funnel plot asymmetry

Back

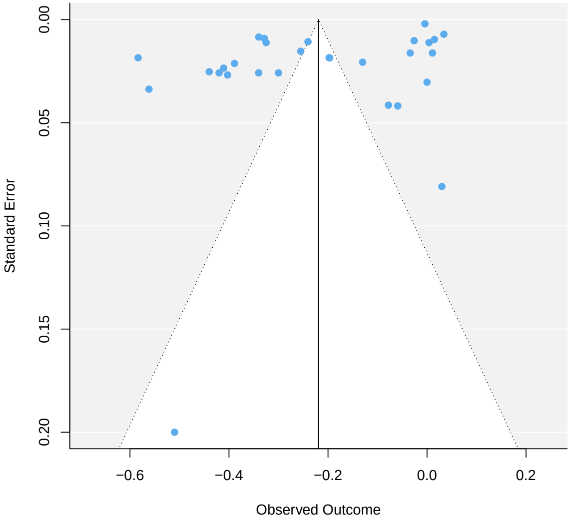


Figure 6: Funnel plot: All experiments

Funnel plot asymmetry

Back

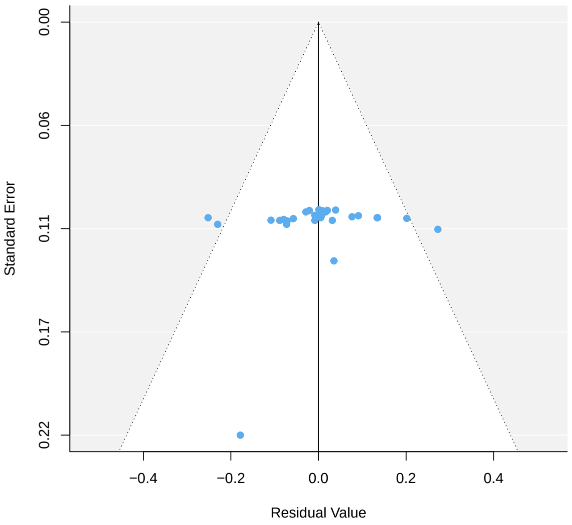


Figure 7: Funnel plot: All experiments with field experiment moderator

Funnel plot asymmetry Back

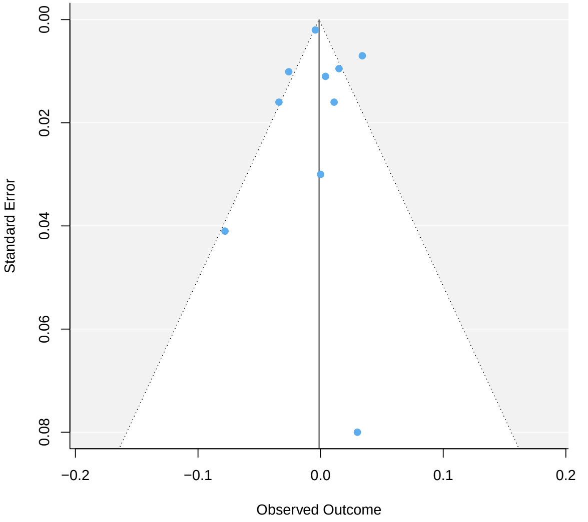


Figure 8: Funnel plot: Field experiments

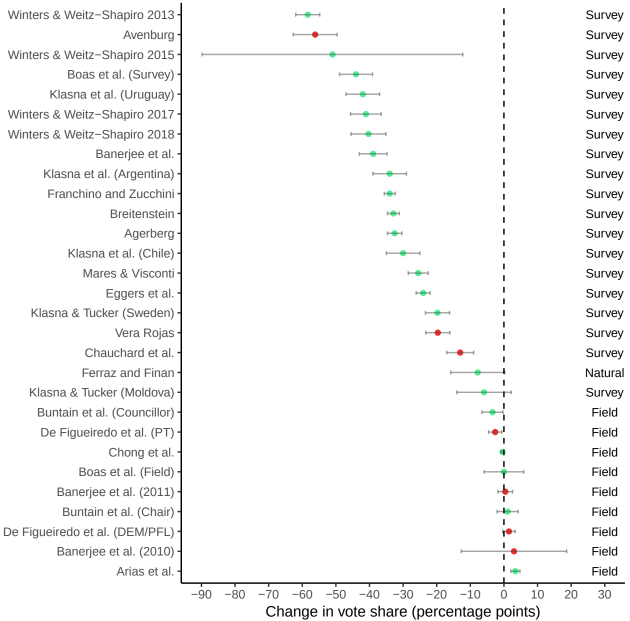
Does p-value predict publication status? Back

Table 3: Do p-values predict publication status?

	Dependent variable:	
	Published	
	OLS	Logit
Reference: P less than 0.01	0.84***	1.67***
	(0.10)	(0.63)
P less than 0.05	−0.18	−0.98
	(0.27)	(1.38)
P less than 0.1	0.16	14.89
	(0.44)	(2, 399.54)
P greater than 0.1	−0.34	−1.67
	(0.20)	(1.03)
Observations	29	29
Note:	*p<0.1; **p<0.05; ***p<0.01	

All experiments by publication status

Back



Additional conjoint replications: Franchino and Zucchini (2015)

Back

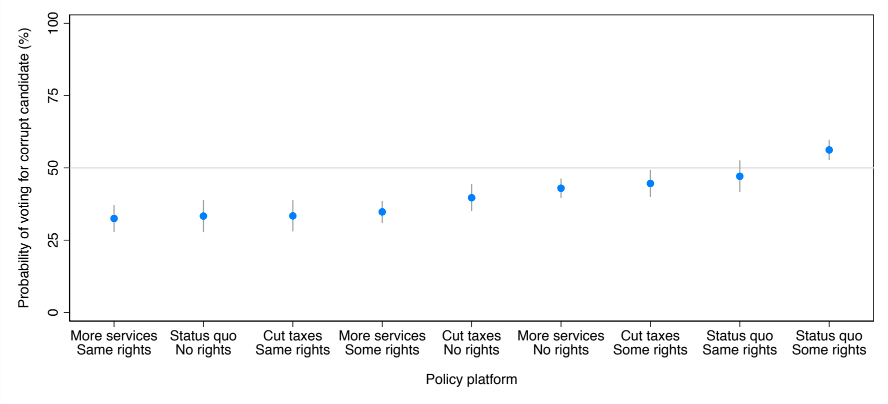


Figure 10: Franchino and Zucchini (2015) conjoint: can policy positions overcome corruption (conservative respondents)?

Additional conjoint replications: Franchino and Zucchini (2015)

Back

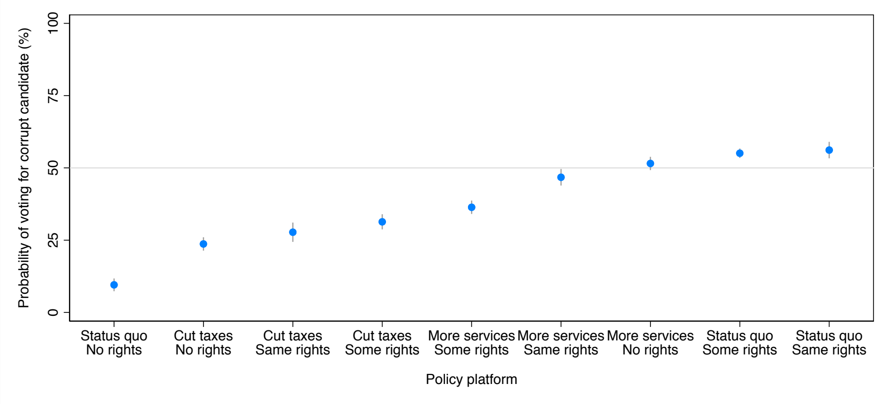


Figure 11: Franchino and Zucchini (2015) conjoint: can policy positions overcome corruption (liberal respondents)?

References

Banerjee, A., Green, D., Green, J., & Pande, R. (2010). Can voters be primed to choose better legislators? experimental evidence from rural india. In *Presented and the political economics seminar, stanford university.*

Banerjee, A., Kumar, S., Pande, R., & Su, F. (2011). Do informed voters make better choices? experimental evidence from urban india. *Unpublished manuscript*.

Breitenstein, S. (2019). Choosing the crook: A conjoint experiment on voting for corrupt politicians. *Research & Politics*, *6*(1), 2053168019832230.

De Vries, C. E., & Solaz, H. (2017). The electoral consequences of corruption. *Annual Review of Political Science*, *20*, 391–408.

Franchino, F., & Zucchini, F. (2015). Voting in a multi-dimensional space: a conjoint analysis employing valence and ideology attributes of candidates. *Political Science Research and Methods*, *3*(2), 221–241.