The Ground Game

GOV 1347 Lab: Week VII

Matthew E. Dardet

Harvard University

October 16, 2024

Check-In

- Questions? Prognostications?

Check-In

- Questions? Prognostications?

```
election.day <- as.Date("2024-11-05")
current.date <- Sys.Date()
cat(paste0("There are ", election.day-current.date, " days until election d
```

```
## There are 20 days until election day!
```

Laboratory Session Recap

- We learned how to model elections using linear models.

Laboratory Session Recap

- We learned how to model elections using linear models.
- Then, we learned how to *evaluate* models using in-sample (e.g., R^2) and out-of-sample (k-fold cross-validated MSE) goodness-of-fit and predictive accuracy measures.

Laboratory Session Recap

- We learned how to model elections using linear models.
- Then, we learned how to *evaluate* models using in-sample (e.g., R^2) and out-of-sample (k-fold cross-validated MSE) goodness-of-fit and predictive accuracy measures.
- Next, we learned how to incorporate ML techniques like regularization to select features and minimize MSE, particularly in high-dimensional settings.

Laboratory Session Recap

- We learned how to model elections using linear models.
- Then, we learned how to *evaluate* models using in-sample (e.g., R^2) and out-of-sample (k-fold cross-validated MSE) goodness-of-fit and predictive accuracy measures.
- Next, we learned how to incorporate ML techniques like regularization to select features and minimize MSE, particularly in high-dimensional settings.
- Last week, we overviewed Bayesian models and the theory of using the Linzer Model and MCMC to simulate possible changes in the election as we approach the election date.

Laboratory Session Recap

- We learned how to model elections using linear models.
- Then, we learned how to *evaluate* models using in-sample (e.g., R^2) and out-of-sample (k-fold cross-validated MSE) goodness-of-fit and predictive accuracy measures.
- Next, we learned how to incorporate ML techniques like regularization to select features and minimize MSE, particularly in high-dimensional settings.
- Last week, we overviewed Bayesian models and the theory of using the Linzer Model and MCMC to simulate possible changes in the election as we approach the election date.
- This week, I'm going to go over a simpler electoral simulation model that uses the Binomial distribution to simulate the election outcome as well as some data about field offices and ground game campaign strategies.

This Week

1. **Binomial Simulations and Probabilistic Models**
2. **Field Offices and the Ground Game**
3. **Pooled and Unpooled Models**

Section 1

Binomial Simulation and Probabilistic Models

Returning to Simulations and Binomial Models

- We know that sometimes, with linear regression, the bounds of Y can exceed those of the true values of the construct or variable being predicted (i.e., having a Y that exceeds 100).

Returning to Simulations and Binomial Models

- We know that sometimes, with linear regression, the bounds of Y can exceed those of the true values of the construct or variable being predicted (i.e., having a Y that exceeds 100).
- We know that we can bound our outcome variable probabilistically using logistic regression.

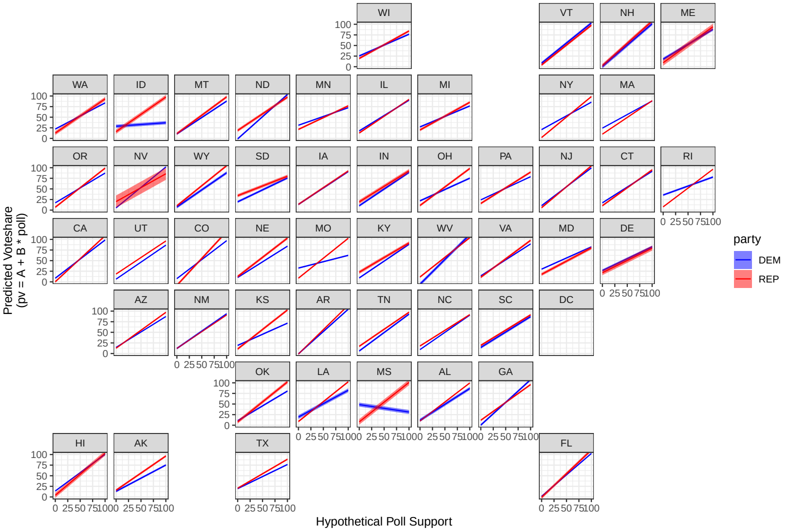
Returning to Simulations and Binomial Models

- We know that sometimes, with linear regression, the bounds of Y can exceed those of the true values of the construct or variable being predicted (i.e., having a Y that exceeds 100).
- We know that we can bound our outcome variable probabilistically using logistic regression.
- Let's see some of the differences between the two methods, particularly in the case of state-level election prediction using polling averages.

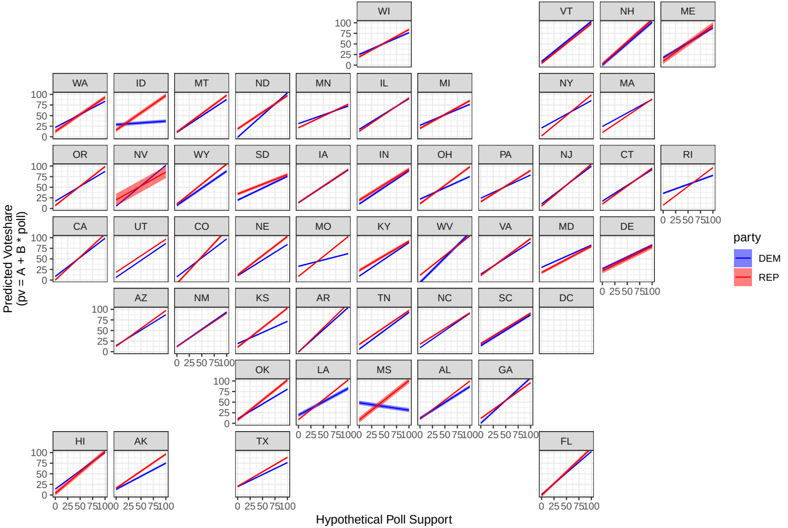
Returning to Simulations and Binomial Models

- We know that sometimes, with linear regression, the bounds of Y can exceed those of the true values of the construct or variable being predicted (i.e., having a Y that exceeds 100).
- We know that we can bound our outcome variable probabilistically using logistic regression.
- Let's see some of the differences between the two methods, particularly in the case of state-level election prediction using polling averages.

Poll-Only State-Level Linear Regression Predictions

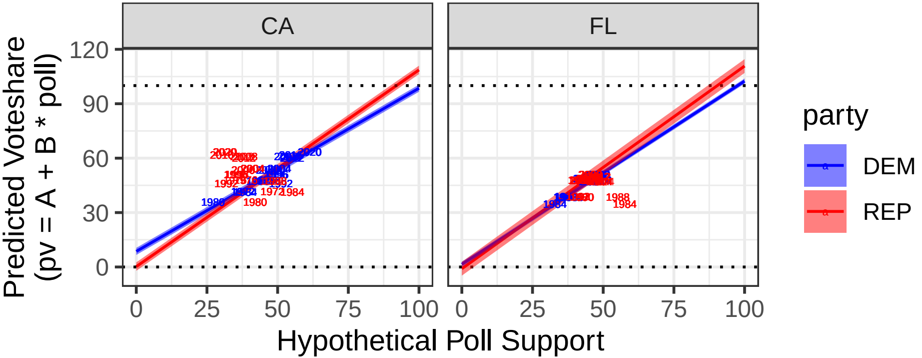


Poll-Only State-Level Linear Regression Predictions



Question: What's wrong with this map?

Poll-Only State-Level Logistic Regression Predictions



Solution: Probabilistic Models

Linear regression: outcome can be any value in a continuous range (-∞, $+\infty$)

Solution: Probabilistic Models

Linear regression: outcome can be any value in a continuous range $(-\infty, +\infty)$

$$\%DemPV_{state} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k \quad \text{or}$$

and modeled as

$$DemPV_{state} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k,$$

Solution: Probabilistic Models

Linear regression: outcome can be any value in a continuous range (-∞, +∞)

$$\%DemPV_{state} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k \quad \text{or}$$

and modeled as

$$DemPV_{state} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k,$$

but the true outcome is bounded to (0, 100) or (0, VEP_{state}) ...

Solution: Probabilistic Models

Linear regression: outcome can be any value in a continuous range (-∞, +∞)

$$\%DemPV_{state} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k \quad \text{or}$$

and modeled as

$$DemPV_{state} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k,$$

but the true outcome is bounded to (0, 100) or (0, VEP_{state}) ...

Binomial logistic regression: election outcome for Democrats is finite draw of voters from the voting-eligible population (VEP_{state}) turning out to vote Democrat (a binomial process) modeled as

$$Pr(\underbrace{\text{Vote for Dem}_{state,i}}_{voter\ \ i\ \ in\ \ state}) = f(\alpha + \beta_1 x_1 + \ldots + \beta_k x_k)$$

$$= \frac{exp(\alpha + \beta_1 x_1 + \ldots + \beta_k x_k)}{1 + exp(\alpha + \beta_1 x_1 + \ldots + \beta_k x_k)} \text{ (for i = 1, ..., } VEP_{state}\text{)}$$

Solution: Probabilistic Models

Linear regression: outcome can be any value in a continuous range (-∞, +∞)

$$\%DemPV_{state} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k \quad \text{or}$$

and modeled as

$$DemPV_{state} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k,$$

but the true outcome is bounded to (0, 100) or (0, VEP_{state}) ...

Binomial logistic regression: election outcome for Democrats is finite draw of voters from the voting-eligible population (VEP_{state}) turning out to vote Democrat (a binomial process) modeled as

$$Pr(\underbrace{\text{Vote for Dem}_{state,i}}_{voter\ \ i\ \ in\ \ state}) = f(\alpha + \beta_1 x_1 + \ldots + \beta_k x_k)$$

$$= \frac{exp(\alpha + \beta_1 x_1 + \ldots + \beta_k x_k)}{1 + exp(\alpha + \beta_1 x_1 + \ldots + \beta_k x_k)} \text{ (for i = 1, ..., } VEP_{state}\text{)}$$

where link function f (inverse logistic function) bounds (-∞, +∞) to (0, 1)

Recap: Binomial Logistic Regression

Supposing we have `x` (a single IV), `y` (a DV) as Dem's popular vote share (%):

Recap: Binomial Logistic Regression

Supposing we have `x` (a single IV), `y` (a DV) as Dem's popular vote share (%):

	Linear regression	**Binomial logistic regression (binomial logit)**
link function	$f(\alpha + \beta x) = \alpha + \beta x$	$f(\alpha + \beta x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$

Recap: Binomial Logistic Regression

Supposing we have `x` (a single IV), `y` (a DV) as Dem's popular vote share (%):

	Linear regression	**Binomial logistic regression (binomial logit)**
link function	$f(\alpha + \beta x) = \alpha + \beta x$	$f(\alpha + \beta x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$
link function name	identity	inverse logistic function

Recap: Binomial Logistic Regression

Supposing we have `x` (a single IV), `y` (a DV) as Dem's popular vote share (%):

	Linear regression	**Binomial logistic regression (binomial logit)**
link function	$f(\alpha + \beta x) = \alpha + \beta x$	$f(\alpha + \beta x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$
link function name	identity	inverse logistic function
link function output	predicted outcome	predicted probability of one draw

Recap: Binomial Logistic Regression

Supposing we have `x` (a single IV), `y` (a DV) as Dem's popular vote share (%):

	Linear regression	**Binomial logistic regression (binomial logit)**
link function	$f(\alpha + \beta x) = \alpha + \beta x$	$f(\alpha + \beta x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$
link function name	identity	inverse logistic function
link function output	predicted outcome	predicted probability of one draw
`R` code	`lm(y∼x)`	`glm(cbind(draws, vep-draws)∼x, family=binomial)`

Recap: Binomial Logistic Regression

Supposing we have `x` (a single IV), `y` (a DV) as Dem's popular vote share (%):

	Linear regression	**Binomial logistic regression (binomial logit)**
link function	$f(\alpha + \beta x) = \alpha + \beta x$	$f(\alpha + \beta x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$
link function name	identity	inverse logistic function
link function output	predicted outcome	predicted probability of one draw
`R` code	`lm(y∼x)`	`glm(cbind(draws, vep-draws)∼x, family=binomial)`
fitting intuition	"do OLS to find coefficients that minimize $\sum (y - \hat{y})^2$"	"find coefficients where fitted draw probabilities $f(\hat{\alpha} + \hat{\beta}x)$ best predict observed `draws`"

Recap: Binomial Logistic Regression

Supposing we have `x` (a single IV), `y` (a DV) as Dem's popular vote share (%):

	Linear regression	**Binomial logistic regression (binomial logit)**
link function	$f(\alpha + \beta x) = \alpha + \beta x$	$f(\alpha + \beta x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$
link function name	identity	inverse logistic function
link function output	predicted outcome	predicted probability of one draw
`R` code	`lm(y~x)`	`glm(cbind(draws, vep-draws)~x, family=binomial)`
fitting intuition	"do OLS to find coefficients that minimize $\sum (y - \hat{y})^2$"	"find coefficients where fitted draw probabilities $f(\hat{\alpha} + \hat{\beta} x)$ best predict observed `draws`"
prediction intuition	"plug in x_{new} and get (i) predicted outcome $\hat{y}_{new} = \hat{\alpha} + \hat{\beta} x_{new}$ and (ii) prediction interval $\hat{y}_{new} \pm 1.96 \times \text{se}(\hat{y}_{new})$"	"plug in x_{new} and get (i) predicted probability of one draw, $f(\hat{\alpha} + \hat{\beta} x_{new})$; also plug in `vep` to get (ii) predicted <u>expected number</u> of draws, $\widehat{\texttt{draws}}$, $\rightsquigarrow \frac{\widehat{\texttt{draws}}}{\texttt{vep}}$ and (iii) predicted <u>distribution</u> of draws from repeated binomial process simulations"

Poll-Only State-Level Binomial Logit

State-level dataset merged with turnout (VEP) data:

Poll-Only State-Level Binomial Logit

State-level dataset merged with turnout (VEP) data:

```
d <- d |>
  left_join(d_turnout, by = c("state", "year")) |>
  filter(year >= 1980) # Filter to when turnout dataset begins.

FL_D <- d %>%
  filter(state=="Florida", party=="DEM")
```

Poll-Only State-Level Binomial Logit

State-level dataset merged with turnout (VEP) data:

```
d <- d |>
  left_join(d_turnout, by = c("state", "year")) |>
  filter(year >= 1980) # Filter to when turnout dataset begins.

FL_D <- d %>%
  filter(state=="Florida", party=="DEM")
```

Linear regression model for Florida race (Dem side):

```
FL_D_lm <- lm(D_pv ~ poll_support,
              FL_D)
```

Poll-Only State-Level Binomial Logit

State-level dataset merged with turnout (VEP) data:

```
d <- d |>
  left_join(d_turnout, by = c("state", "year")) |>
  filter(year >= 1980) # Filter to when turnout dataset begins.

FL_D <- d %>%
  filter(state=="Florida", party=="DEM")
```

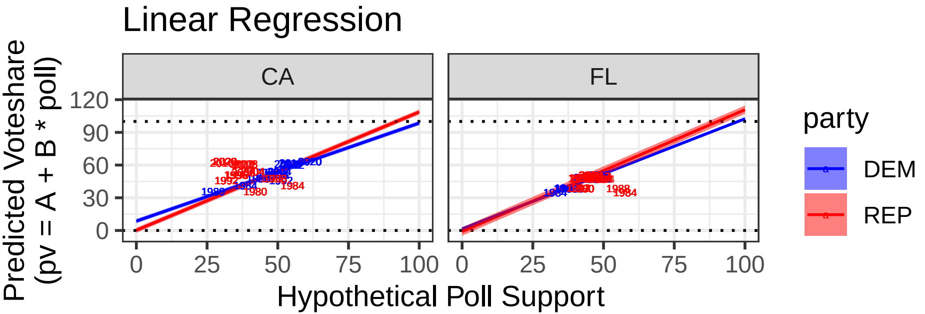
Linear regression model for Florida race (Dem side):

```
FL_D_lm <- lm(D_pv ~ poll_support,
              FL_D)
```

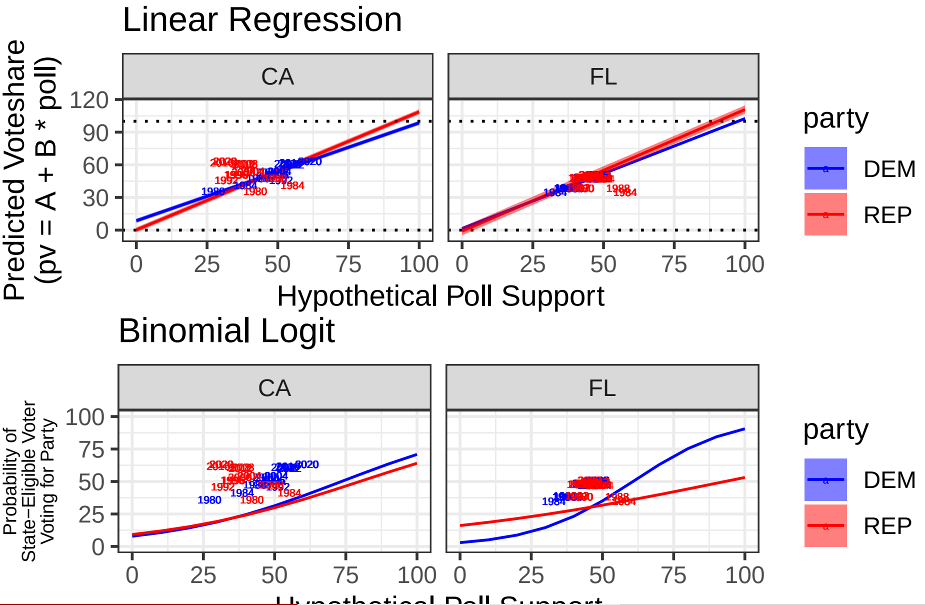
Binomial logit model for Florida race (Dem side):

```
FL_D_glm <- glm(cbind(votes_D, vep-votes_D) ~ poll_support,
                data = FL_D,
                family = binomial(link="logit"))
```

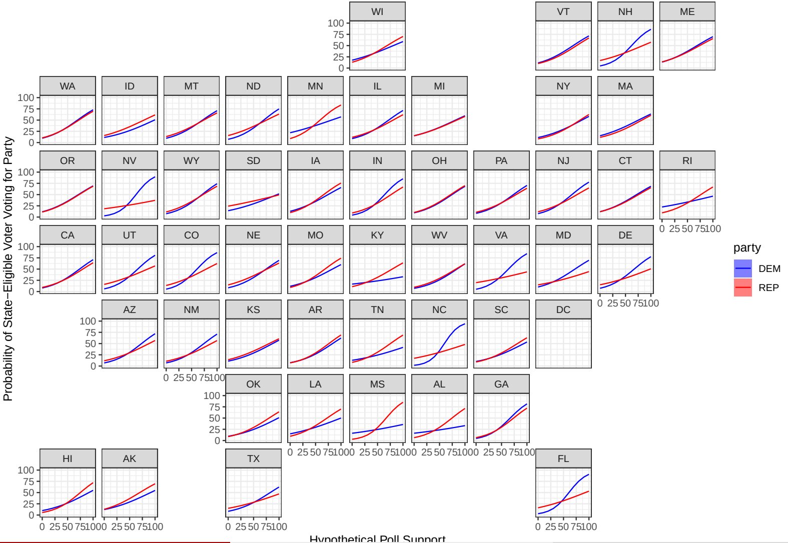
Compoarison of Poll-Only State-Level Predictions



Compoarison of Poll-Only State-Level Predictions



Poll-Only State-Level Binomial Logit Predictions (All States)



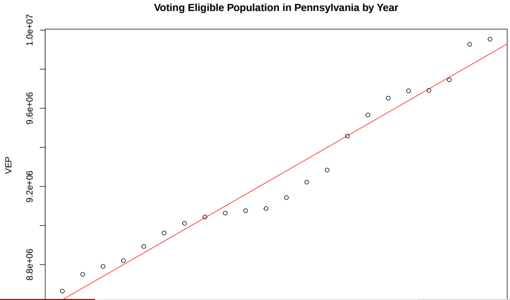
Simulating a Distribution of Election Results

Instead of (i) a probability for a single D voter or (ii) single expected number of D voters from `vep`, $\widehat{\texttt{draws}}$, we can predict a (iii) distribution of draws from binomial process on that `vep`.

Simulating a Distribution of Election Results

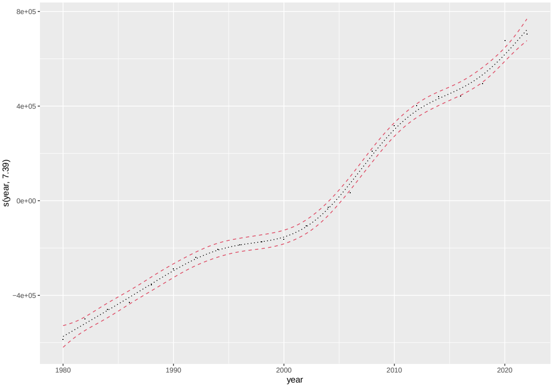
Instead of (i) a probability for a single D voter or (ii) single expected number of D voters from `vep`, $\widehat{\texttt{draws}}$, we can predict a (iii) distribution of draws from binomial process on that `vep`.

```r
# Get historical eligible voting population in Pennsylvania.
vep_PA_2020 <- as.integer(d_turnout$vep[d_turnout$state == "Pennsylvania" & d_turnout$year == 2020])
vep_PA <- d_turnout |> filter(state == "Pennsylvania") |> select(vep, year)
# Fit regression for 2024 VEP prediction.
lm_vep_PA <- lm(vep ~ year, vep_PA)
vep_PA_2024_ols <- predict(lm_vep_PA, newdata = data.frame(year = 2024)) |> as.numeric()
plot(x = vep_PA$year, y = vep_PA$vep, xlab = "Year", ylab = "VEP", main = "Voting Eligible Population in Pennsylvania by Year")
abline(lm_vep_PA, col = "red")
```



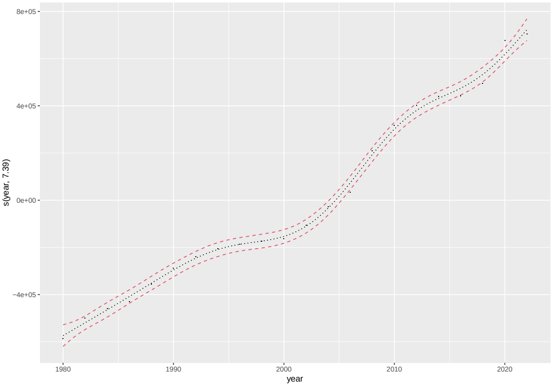
Simulating a Distribution of Election Results

```
gam_vep_PA <- mgcv::gam(vep ~ s(year), data = vep_PA)
print(plot(getViz(gam_vep_PA)) + l_points() + l_fitLine(linetype = 3) + l_ciLine(colour = 2) + theme_get())
```



Simulating a Distribution of Election Results

```r
gam_vep_PA <- mgcv::gam(vep ~ s(year), data = vep_PA)
print(plot(getViz(gam_vep_PA)) + l_points() + l_fitLine(linetype = 3) + l_ciLine(colour = 2) + theme_get())
```




```r
# Use generalized additive model (GAM) to predict 2024 VEP in Pennsylvania.
vep_PA_2024_gam <- predict(gam_vep_PA, newdata = data.frame(year = 2024)) |> as.numeric()
# Take weighted average of linear and GAM predictions for final prediction.
vep_PA_2024 <- as.integer(0.75*vep_PA_2024_gam + 0.25*vep_PA_2024_ols)
```

Simulating a Distribution of Election Results

```r
# Split datasets by party.
PA_D <- d |> filter(state == "Pennsylvania" & party == "DEM")
PA_R <- d |> filter(state == "Pennsylvania" & party == "REP")

# Fit Democrat and Republican models.
PA_D_glm <- glm(cbind(votes_D, vep - votes_D) ~ poll_support, data = PA_D, family = binomial(link = "logit"))
PA_R_glm <- glm(cbind(votes_R, vep - votes_R) ~ poll_support, data = PA_R, family = binomial(link = "logit"))
```

Simulating a Distribution of Election Results

```r
# Split datasets by party.
PA_D <- d |> filter(state == "Pennsylvania" & party == "DEM")
PA_R <- d |> filter(state == "Pennsylvania" & party == "REP")

# Fit Democrat and Republican models.
PA_D_glm <- glm(cbind(votes_D, vep - votes_D) ~ poll_support, data = PA_D, family = binomial(link = "logit"))
PA_R_glm <- glm(cbind(votes_R, vep - votes_R) ~ poll_support, data = PA_R, family = binomial(link = "logit"))
```

```r
# Get predicted draw probabilities for D and R.
PA_pollav_D <- d_state_polls$poll_support[d_state_polls$state == "Pennsylvania" & d_state_polls$weeks_left == 3 & d_state_polls$party
PA_pollav_R <- d_state_polls$poll_support[d_state_polls$state == "Pennsylvania" & d_state_polls$weeks_left == 3 & d_state_polls$party

prob_D_vote_PA_2024 <- predict(PA_D_glm, newdata = data.frame(poll_support = PA_pollav_D), se = TRUE, type = "response")[[1]] |> as.nu
prob_R_vote_PA_2024 <- predict(PA_R_glm, newdata = data.frame(poll_support = PA_pollav_R), se = TRUE, type = "response")[[1]] |> as.nu
```

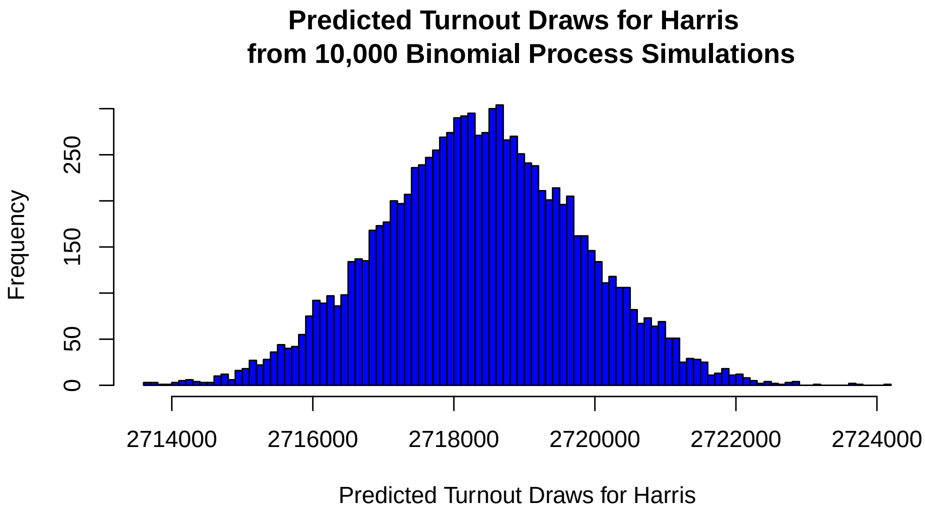
Simulating a Distribution of Election Results

```
# Split datasets by party.
PA_D <- d |> filter(state == "Pennsylvania" & party == "DEM")
PA_R <- d |> filter(state == "Pennsylvania" & party == "REP")

# Fit Democrat and Republican models.
PA_D_glm <- glm(cbind(votes_D, vep - votes_D) ~ poll_support, data = PA_D, family = binomial(link = "logit"))
PA_R_glm <- glm(cbind(votes_R, vep - votes_R) ~ poll_support, data = PA_R, family = binomial(link = "logit"))
```

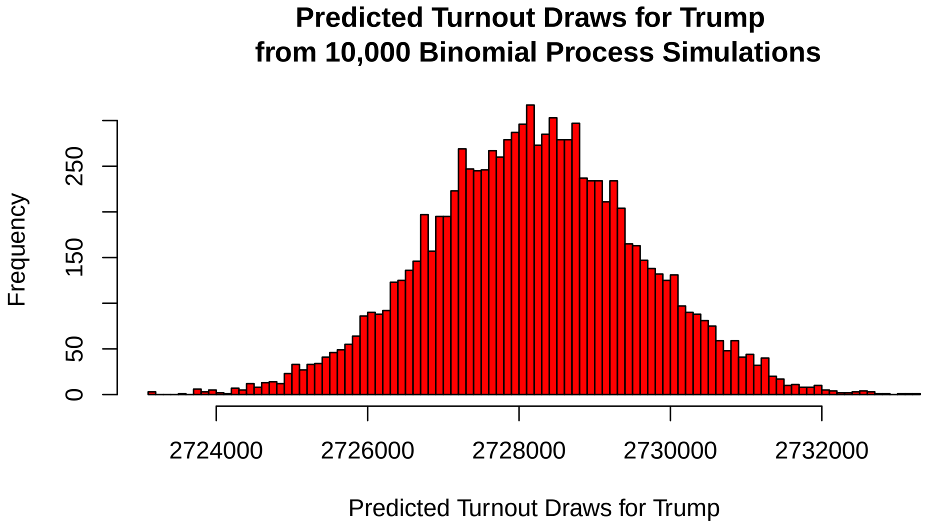
```
# Get predicted draw probabilities for D and R.
PA_pollav_D <- d_state_polls$poll_support[d_state_polls$state == "Pennsylvania" & d_state_polls$weeks_left == 3 & d_state_polls$party
PA_pollav_R <- d_state_polls$poll_support[d_state_polls$state == "Pennsylvania" & d_state_polls$weeks_left == 3 & d_state_polls$party

prob_D_vote_PA_2024 <- predict(PA_D_glm, newdata = data.frame(poll_support = PA_pollav_D), se = TRUE, type = "response")[[1]] |> as.nu
prob_R_vote_PA_2024 <- predict(PA_R_glm, newdata = data.frame(poll_support = PA_pollav_R), se = TRUE, type = "response")[[1]] |> as.nu
```

```
# Get predicted distribution of draws from the population.
sim_D_votes_PA_2024 <- rbinom(n = 10000, size = vep_PA_2024, prob = prob_D_vote_PA_2024)
sim_R_votes_PA_2024 <- rbinom(n = 10000, size = vep_PA_2024, prob = prob_R_vote_PA_2024)
```

Simulating a Distribution of Election Results: Harris PA PV

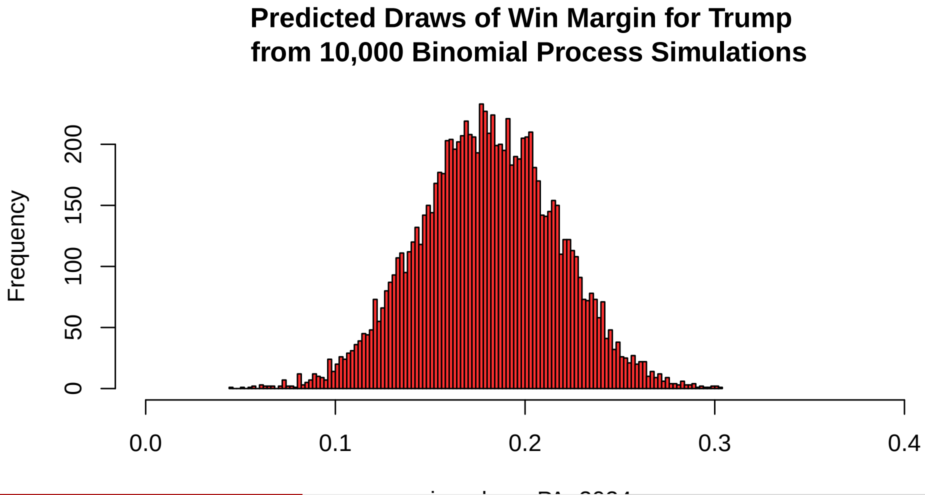


Simulating a Distribution of Election Results: Trump PA PV



Simulating a Distribution of Election Results: Trump Win Margin

```r
# Simulating a distribution of election results: Trump win margin.
sim_elxns_PA_2024 <- ((sim_R_votes_PA_2024-sim_D_votes_PA_2024)/(sim_D_votes_PA_2024 + sim_R_votes_PA_2024))*100
```



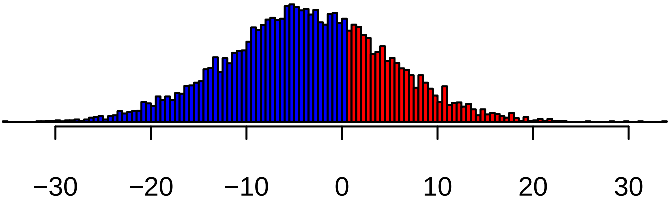
Making the Simulation Realistic by Incorporating Priors

We can simulate fluctuations in the probability of a voter from the VEP voting for a party by using an informed guess (a prior) at its standard deviation, e.g., a standard deviation of polls:

Making the Simulation Realistic by Incorporating Priors

We can simulate fluctuations in the probability of a voter from the VEP voting for a party by using an informed guess (a prior) at its standard deviation, e.g., a standard deviation of polls:

```
PA_sdpoll_D <- sd(d_state_polls$poll_support[d_state_polls$state == "Pennsylvania" & d_state_pol
PA_sdpoll_R <- sd(d_state_polls$poll_support[d_state_polls$state == "Pennsylvania" & d_state_pol
sim_D_votes_PA_2024_2 <- rbinom(n = 10000, size = vep_PA_2024, prob = rnorm(10000, PA_pollav_D/
sim_R_votes_PA_2024_2 <- rbinom(n = 10000, size = vep_PA_2024, prob = rnorm(10000, PA_pollav_R/
sim_elxns_PA_2024_2 <- ((sim_R_votes_PA_2024_2-sim_D_votes_PA_2024_2)/(sim_D_votes_PA_2024_2 +
```



Summary of Probabilistic Models

- Explicitly capture a random or probabilistic process of the world
    - ex: some draw of voters from VEP turning out
- Models like binomial logit (**generalized linear models**) use a link function to bound the outcome to a probability value
    - link functions like the inverse logistic function allow us to **non-linearly** predict DV from IVs (solving another problem of linear regression)
- Workflow: estimate the parameters of a probabilistic model $\rightsquigarrow$ obtain distributions from repeated simulations of probabilistic process
    - ex: in binomial logit, we repeatedly draw voters from a binomial process based on predicted probability of one voter turning out Dem
    - $\sim$ last week: how The Economist simulates elections
- Diagnostics: can still use out-of-sample evaluation tools; see glossary and `had.co.nz/notes/modelling/logistic-regression.html` for other diagnostics.

Section 2

Field Offices and the Ground Game

Field Offices

joebiden · Follow

joebiden 27w
We've opened over 100 campaign headquarters and field offices, hiring staff all across the country before Trump and his MAGA Republicans have announced opening one single office.

Let's keep it going.

42,593 likes
April 7

Log in to like or comment.

Where Should Campaigns Build Field Offices?

Strategic placement of field offices

- Strategic decisions at the **state** level:

Where Should Campaigns Build Field Offices?

Strategic placement of field offices

- Strategic decisions at the **state** level: 1. build in core states or battleground states?

Where Should Campaigns Build Field Offices?

Strategic placement of field offices

- Strategic decisions at the **state** level: 1. build in core states or battleground states? 2. build in opponents' core states?

Where Should Campaigns Build Field Offices?

Strategic placement of field offices

- Strategic decisions at the **state** level: 1. build in core states or battleground states? 2. build in opponents' core states?
- Strategic decisions at the **county** level:

Where Should Campaigns Build Field Offices?

Strategic placement of field offices

- Strategic decisions at the **state** level: 1. build in core states or battleground states? 2. build in opponents' core states?
- Strategic decisions at the **county** level: 1. build in core counties (deep investment) or swing counties (broad outreach) within a given battleground state?

Where Should Campaigns Build Field Offices?

Strategic placement of field offices

- Strategic decisions at the **state** level: 1. build in core states or battleground states? 2. build in opponents' core states?
- Strategic decisions at the **county** level: 1. build in core counties (deep investment) or swing counties (broad outreach) within a given battleground state? 2. build field offices to match opponents' field offices?

Where Should Campaigns Build Field Offices?

`fieldoffice_2012_bycounty.csv`: field office locations in 2012 presidential election.

Where Should Campaigns Build Field Offices?

`fieldoffice_2012_bycounty.csv`: field office locations in 2012 presidential election.

`state`	state that county is in
`fips`	4-digit county code (like a ZIP code)
`battle`	battleground state according to New York Times
`normal`	average of 2000-2008 Republican voteshare
`swing`	counties with '45 $<$ normal $<=$ 55'
`core_dem`	counties with 'normal $<$ 45'
`core_rep`	counties with 'normal $>=$ 55'
`medage08`	median age in county in 2008
`pop2008`	median population in county in 2008
`black`	% black in county in 2012
`hispanic`	% hispanic in county in 2012
`pc_less_hs00`	% without high school diploma in 2000
`pc_degree00`	% with college degree in 2000

Where Should Campaigns Build Field Offices?

```
fo_2012 <- read_csv("fieldoffice_2012_bycounty.csv")
lm_obama <- lm(obama12fo ~ romney12fo + swing +
                 core_rep + swing:romney12fo +
                 core_rep:romney12fo +  battle +
                 medage08 +  pop2008 +
                 pop2008^2 + medinc08 +
                 black + hispanic +
                 pc_less_hs00 + pc_degree00 +
                 as.factor(state),
               fo_2012)
lm_romney <- lm(romney12fo ~ obama12fo + swing +
                  core_dem + swing:obama12fo +
                  core_dem:obama12fo +  battle +
                  medage08 + pop2008 +
                  pop2008^2 +  medinc08 +
                  black + hispanic +
                  pc_less_hs00 + pc_degree00 +
                  as.factor(state),
                  fo_2012)
```

Where Should Campaigns Build Field Offices?

Table 1: Placement of Field Offices (2012)

	Dependent variable:	
	obama12fo	romney12fo
	(1)	(2)
romney12fo	2.546*** (0.114)	
obama12fo		0.374*** (0.020)
swing	0.001 (0.055)	−0.012 (0.011)
core_rep	0.007 (0.061)	
core_dem		0.004 (0.027)
battle	0.541*** (0.096)	0.014 (0.042)
medage08		
romney12fo:swing	−0.765*** (0.116)	
romney12fo:core_rep	−1.875*** (0.131)	
obama12fo:swing		−0.081*** (0.020)
obama12fo:core_dem		−0.164*** (0.023)
Constant	−0.340* (0.196)	0.001 (0.079)

Where Should Campaigns Build Field Offices?

Table 1: Placement of Field Offices (2012)

	Dependent variable:	
	obama12fo	romney12fo
	(1)	(2)
romney12fo	2.546^{***} (0.114)	
obama12fo		0.374^{***} (0.020)
swing	0.001 (0.055)	−0.012 (0.011)
core_rep	0.007 (0.061)	
core_dem		0.004 (0.027)
battle	0.541^{***} (0.096)	0.014 (0.042)
medage08		
romney12fo:swing	-0.765^{***} (0.116)	
romney12fo:core_rep	-1.875^{***} (0.131)	
obama12fo:swing		-0.081^{***} (0.020)
obama12fo:core_dem		-0.164^{***} (0.023)
Constant	-0.340^{*} (0.196)	0.001 (0.079)

Q1: Core state vs Battleground state vs Opponent's core state?

Where Should Campaigns Build Field Offices?

Table 1: Placement of Field Offices (2012)

	Dependent variable:	
	obama12fo	romney12fo
	(1)	(2)
romney12fo	2.546*** (0.114)	
obama12fo		0.374*** (0.020)
swing	0.001 (0.055)	−0.012 (0.011)
core_rep	0.007 (0.061)	
core_dem		0.004 (0.027)
battle	0.541*** (0.096)	0.014 (0.042)
medage08		
romney12fo:swing	−0.765*** (0.116)	
romney12fo:core_rep	−1.875*** (0.131)	
obama12fo:swing		−0.081*** (0.020)
obama12fo:core_dem		−0.164*** (0.023)
Constant	−0.340* (0.196)	0.001 (0.079)

Q1: Core state vs Battleground state vs Opponent's core state?

Q2: Core county vs Swing county? Matching opponents?

Do Field Offices Work?

Effects of field offices: mobilization and persuasion[1]

```
fo_dem <- read_csv("fieldoffice_2004-2012_dems.csv")
```

[1] Data only available for Dems.

Do Field Offices Work?

Effects of field offices: mobilization and persuasion[1]

```
fo_dem <- read_csv("fieldoffice_2004-2012_dems.csv")
```

mobilization $\rightsquigarrow$ boost turnout?

[1]Data only available for Dems.

Do Field Offices Work?

Effects of field offices: mobilization and persuasion[1]

```
fo_dem <- read_csv("fieldoffice_2004-2012_dems.csv")
```

mobilization $\rightsquigarrow$ boost turnout?

```
ef_t <- lm(turnout_change ~ dummy_fo_change +
               battle + dummy_fo_change:battle +
               as.factor(state) + as.factor(year), fo_dem)
```

[1] Data only available for Dems.

Do Field Offices Work?

Effects of field offices: mobilization and persuasion[1]

```r
fo_dem <- read_csv("fieldoffice_2004-2012_dems.csv")
```

mobilization $\rightsquigarrow$ boost turnout?

```r
ef_t <- lm(turnout_change ~ dummy_fo_change +
               battle + dummy_fo_change:battle +
               as.factor(state) + as.factor(year), fo_dem)
```

persuasion + mobilization $\rightsquigarrow$ boost Dem vote?

[1] Data only available for Dems.

Do Field Offices Work?

Effects of field offices: mobilization and persuasion[1]

```
fo_dem <- read_csv("fieldoffice_2004-2012_dems.csv")
```

mobilization $\rightsquigarrow$ boost turnout?

```
ef_t <- lm(turnout_change ~ dummy_fo_change +
             battle + dummy_fo_change:battle +
             as.factor(state) + as.factor(year), fo_dem)
```

persuasion + mobilization $\rightsquigarrow$ boost Dem vote?

```
ef_d <- lm(dempct_change ~ dummy_fo_change +
             battle + dummy_fo_change:battle +
             as.factor(state) + as.factor(year), fo_dem)
```

[1] Data only available for Dems.

Effects of Field Offices on Turnout and Vote Share

Table 2: Effect of DEM Field Offices on Turnout and DEM Vote Share (2004-2012)

	Dependent variable:	
	turnout_change	dempct_change
	(1)	(2)
dummy_fo_change	0.004*** (0.001)	0.009*** (0.002)
battle	0.024*** (0.002)	0.043*** (0.003)
as.factor(state)Arizona		
dummy_fo_change:battle	−0.002 (0.002)	0.007** (0.003)
Constant	0.029*** (0.002)	0.022*** (0.003)
Observations	6,224	6,224
Adjusted R^2	0.419	0.469
Note:		*p<0.1; **p<0.05; ***p<0.01

Section 3

Field Strategy Exercise

Field Strategy Exercise

With three weeks to go until the election, your campaign strategist boss, without any notice, is asking you to prepare some notes about field strategy in 2016 presidential elections. Luckily, you still have your all-star team of former GOV 1347 election analysts to help you out.

Using `fieldoffice_2012-2016.csv_byaddress.csv`, evaluate the **field strategies of Clinton and Trump in 2016.**

1. How many field offices did Clinton and Trump open? How does that compare to Obama and Romney's numbers?
2. Which states did Clinton and Trump focus on? Did they win the ones they built many field offices?
3. After Clinton's loss, pundits said Clinton lost Wisconsin because her ground game was not strong there in the weeks leading up to election day. Do you agree with this accusation? Where would you have built more field offices in Wisconsin if you were the campaign manager? (Hint: Check what Obama did.)

Field Strategies

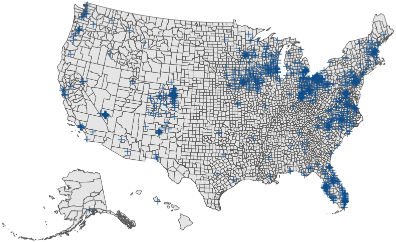
Total number of field offices:

- Obama: **965**, Romney: **283**
- Clinton: **538**, Trump: **165**

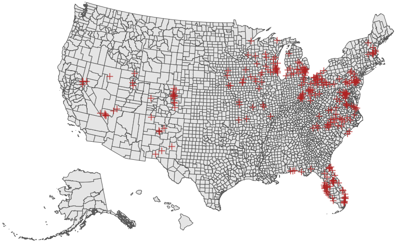
Field strategies of Clinton and Trump in 2016

|

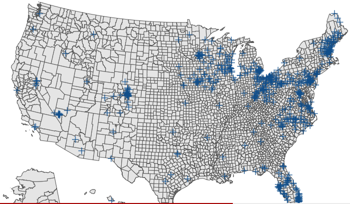
Obama 2012 Field Offices



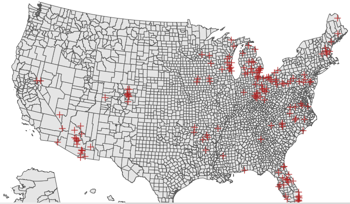
Romney 2012 Field Offices



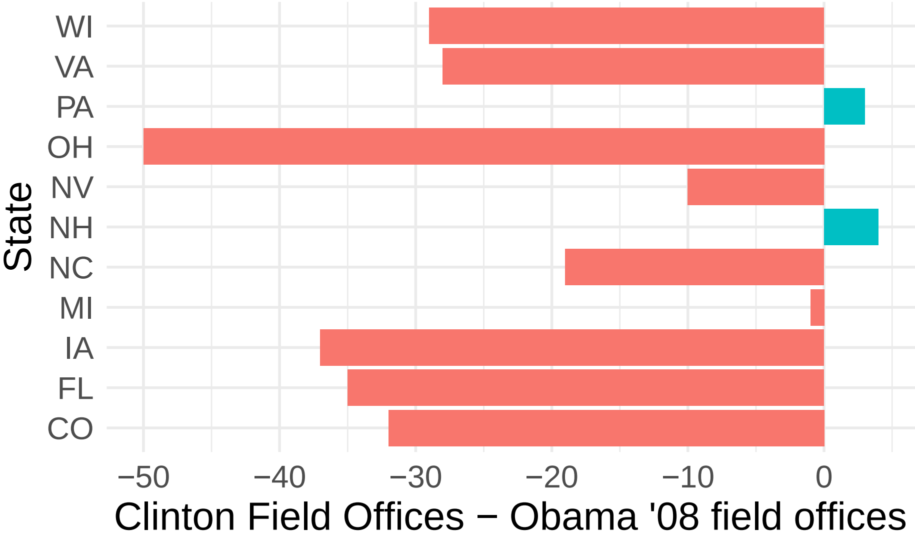
Clinton 2016 Field Offices



Trump 2016 Field Offices

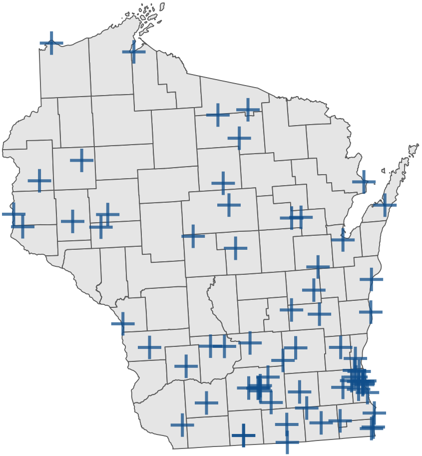


Field Strategies in 2016 (Darr 2019)

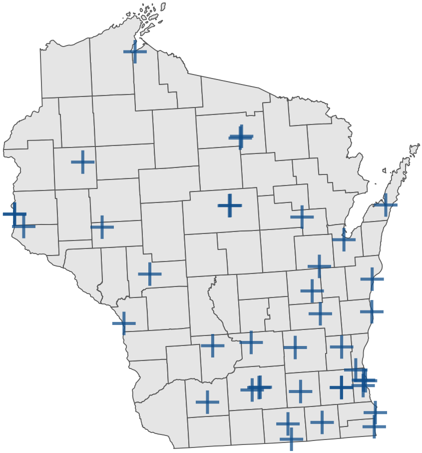


Field Strategies in 2016 (Darr 2019)

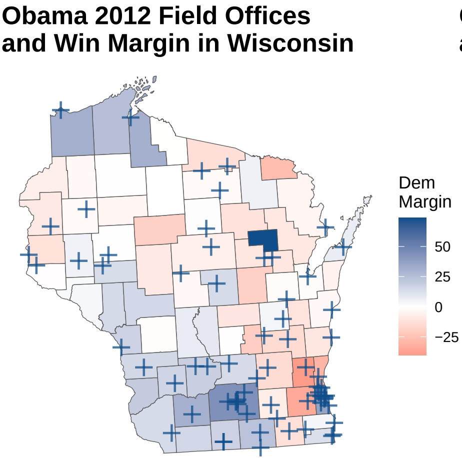
Obama 2012 Field Offices in Wisconsin

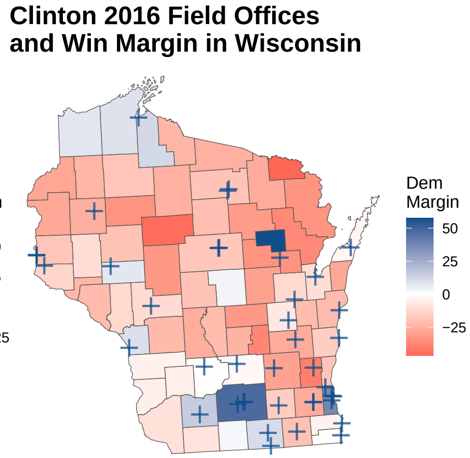


Clinton 2016 Field Offices in Wiscons



Field Strategies in 2016 (Darr 2019)



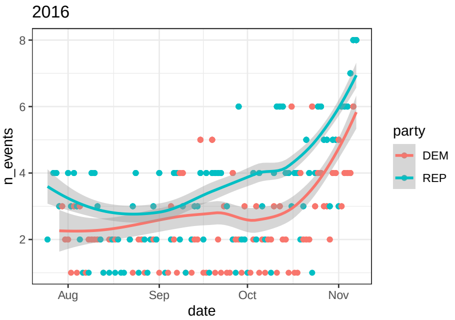


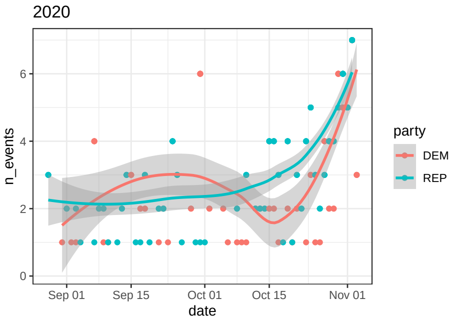
Campaign Events

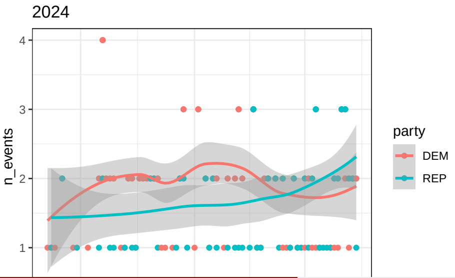
Campaign Events

Campaign Events

Visualizing Campaign Events Over Time



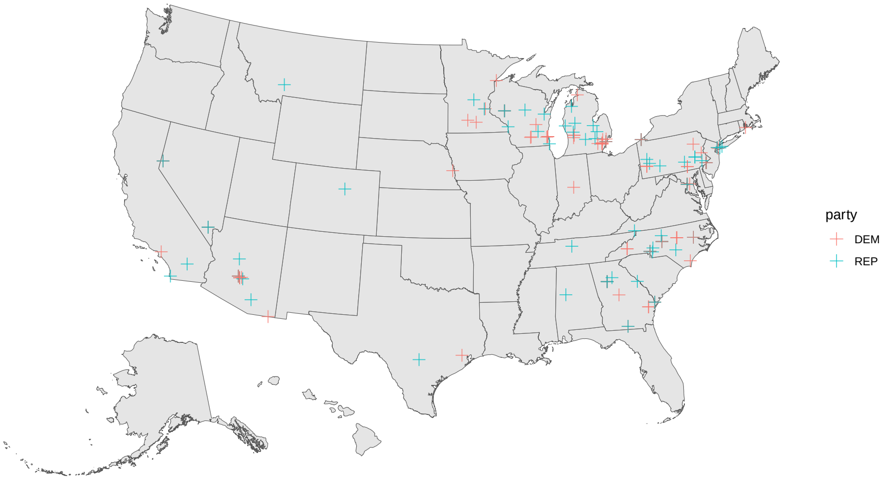




Visualizing Campaign Events in 2024

|

2024 Campaign Events



Section 4

Can Campaign Events Predict State-Level Election Outcomes?

Can Campaign Events Predict State-Level Election Outcomes?

Table 3: Association Between Campaign Events and Voting Outcomes of Interest

	Dependent variable:	
	D_pv2p	R_pv2p
	(1)	(2)
n_ev_D	0.126*** (0.034)	
ev_diff_D_R	0.105 (0.067)	
n_ev_R		−0.126*** (0.034)
ev_diff_R_D		0.230*** (0.078)
Constant	48.189*** (0.369)	51.810*** (0.369)
Observations	714	714
Adjusted R^2	0.019	0.019
Note:		*p<0.1; **p<0.05; ***p<0.01

Section 5

Pooled and Unpooled Models

Pooled and Unpooled Models

Generally, we fit a separate poll-based regression for each state $\rightsquigarrow$ *assumes separate parameters for each state*. (unpooled)

$$DemPV_{state} = f(\alpha_{state} + \beta_{1,state}x_1 + \ldots + \beta_{k,state}x_{k,state})$$

Pooled and Unpooled Models

Generally, we fit a separate poll-based regression for <u>each</u> state $\rightsquigarrow$ *assumes separate parameters for each state*. (unpooled)

$$DemPV_{state} = f(\alpha_{state} + \beta_{1,state}x_1 + \ldots + \beta_{k,state}x_{k,state})$$

We can also fit one regression <u>across</u> states $\rightsquigarrow$ *assumes one set of parameters for all states*. (pooled)

$$DemPV_{state} = f(\alpha + \beta_1 x_1 + \ldots + \beta_k x_k)$$

Pooled and Unpooled Models

Generally, we fit a separate poll-based regression for each state $\rightsquigarrow$ *assumes separate parameters for each state*. (unpooled)

$$DemPV_{state} = f(\alpha_{state} + \beta_{1,state}x_1 + \ldots + \beta_{k,state}x_{k,state})$$

We can also fit one regression across states $\rightsquigarrow$ *assumes one set of parameters for all states*. (pooled)

$$DemPV_{state} = f(\alpha + \beta_1 x_1 + \ldots + \beta_k x_k)$$

Q: Why are some advantages of the latter?

Across-State Correlations in Pooled Models

- Pooled models can capture correlations across states by using the same coefficients to produce each state's outcome (there are even more advanced ways of doing this using clustering).

Across-State Correlations in Pooled Models

- Pooled models can capture correlations across states by using the same coefficients to produce each state's outcome (there are even more advanced ways of doing this using clustering).
- A pooled model relies on less data from each state, thus "drawing strength" from less data-sparse states.

Across-State Correlations in Pooled Models

- Pooled models can capture correlations across states by using the same coefficients to produce each state's outcome (there are even more advanced ways of doing this using clustering).
- A pooled model relies on less data from each state, thus "drawing strength" from less data-sparse states.
- Professional election forecasters use these correlations to update state A's prediction if state B's outcome is hypothetically known.

Across-State Correlations in Pooled Models

- Pooled models can capture correlations across states by using the same coefficients to produce each state's outcome (there are even more advanced ways of doing this using clustering).
- A pooled model relies on less data from each state, thus "drawing strength" from less data-sparse states.
- Professional election forecasters use these correlations to update state A's prediction if state B's outcome is hypothetically known.
- Can combine pooled and unpooled models through the power of ensembling. And with appropriate checks on the model perfomance.

Blog Extensions (Optional)

1. **Demographic Model.** Incorporate demographic surge data into your state-level popular vote predictions and see if it changes your predictions even though there is a good deal of evidence that these variables are unpredictive of the outcome.
2. **Air War vs. Ground Game** Assume that a field office costs $21,000 and 1 GRP buy is $300. If you are a campaign manager for Harris or Trump with a fixed budget, where would you build or air and how much? What should the optimal ground game strategy for Trump and Harris in 2024?
3. **2024 Turnout (Un)predictability.** Visualize trends in turnout using our data. In 2024, it's hard to predict turnout due to the lasting shock of COVID-19 and mail-in voting:
    - How do the *FiveThirtyEight* and *Economist* forecasts deal with this unpredictability? Explain their approaches and critically evaluate them.
    - Build probabilistic models to simulate fluctatoins in turnout. (Rather than fixing each state's maximum number of Binomial draws to `VEP`, draw a random number from a distribution constructed using `VEP` and your guess on the effect of vote-by-mail and other factors.)