EI Considerations

Gabe and Amy
11/18/2020

EI Overview

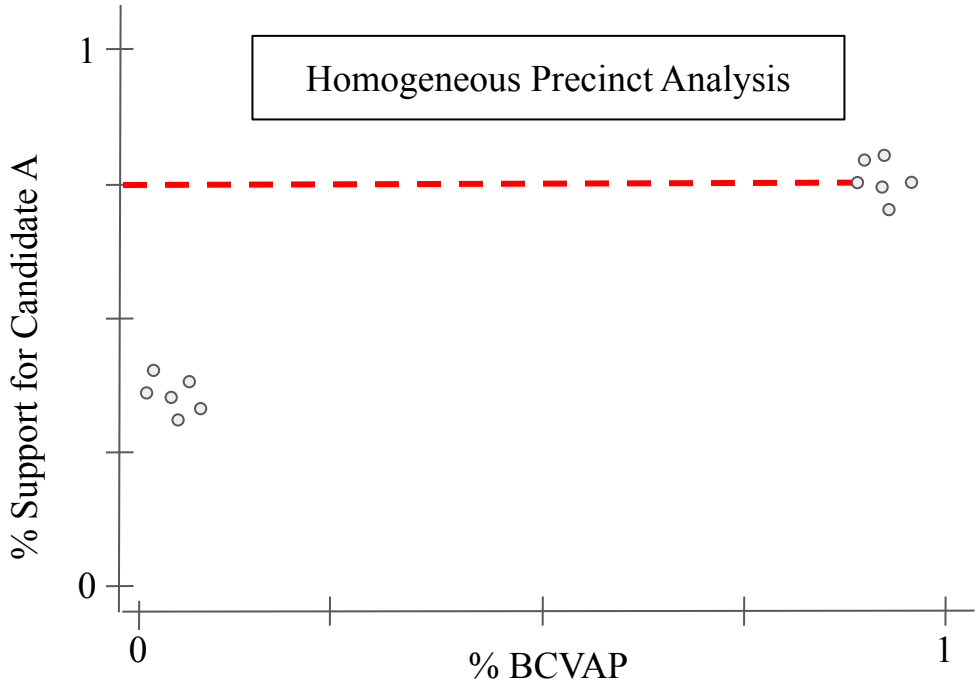
What % of Black voters chose Candidate A?
(e.g. for VRA)

For each precinct we have:

	Candidate A	Candidate B	Totals
BCVAP	?	?	2000
non-BCVAP	?	?	3000
Totals	1000	800	

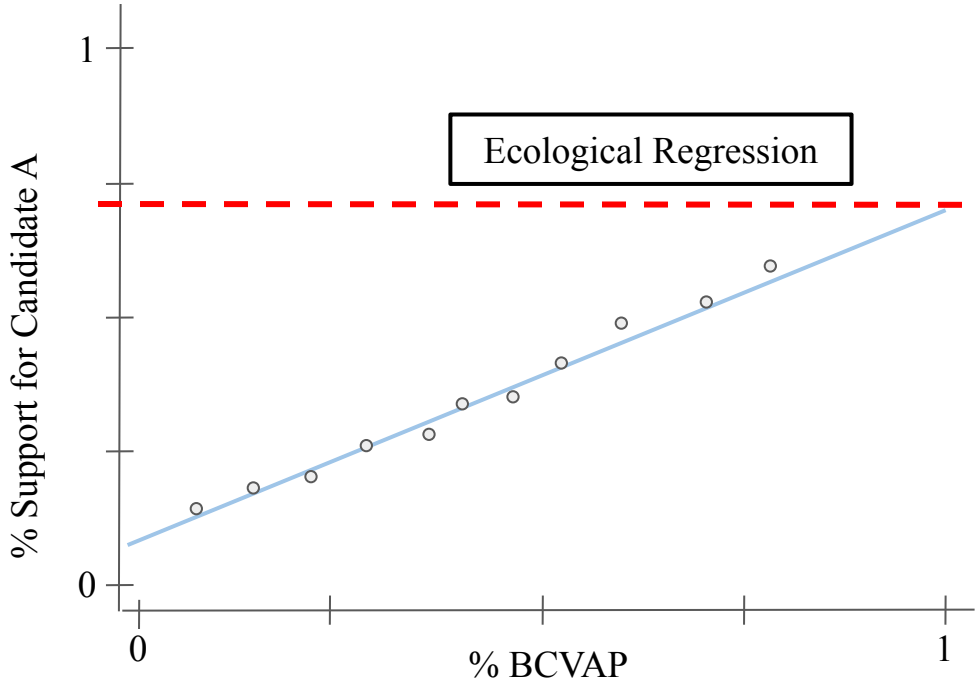
EI Overview

(Look for patterns across all the precincts)



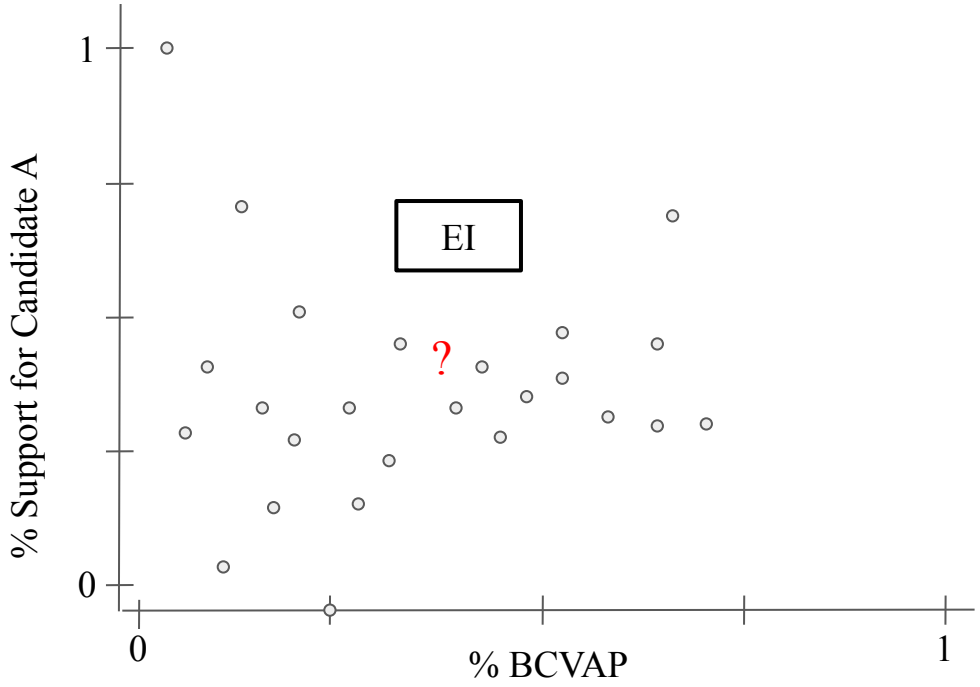
EI Overview

(Look for patterns across all the precincts)



EI Overview

(Look for patterns across all the precincts)



EI Overview

- “EI” is not a single well-defined technique! Lots of flavors:
  - King’s EI, Rosen, Greiner-Quinn, …
  - Underlying distributions (truncated bivariate normal, univariate gamma, multinomial dirichlet)
  - How to address turnout issues (e.g. Double equation vs Single equation, ‘abstain’ columns, etc)
  - Different implementations (e.g. eiPack, pyEI)
  - Lots of parameter knobs to adjust (tuning, mcmc specs, etc)
  - 2x2 vs RxC tables
  - Counts vs Proportions (each have issues); CVAP/VAP/voter file/exit polls

- EI techniques are not always reliable
  - depend on underlying assumptions that do not hold well in many scenarios

- Most practitioners likely treat EI as a “black box” without needing to understand inner-workings

- Reasonable and innocuous-seeming choices can have significant impacts!

Our EI Method

precinct p_i	Candidate 1	Candidate 2	Neither
BCVAP			
HCVAP			
WCVAP			
OCVAP			

Use 4 demographic groups

Our EI Method

precinct p_i	Candidate 1	Candidate 2	Neither
BCVAP			
HCVAP			
WCVAP			
OCVAP			

Use 4 demographic groups	Aggregate precincts statewide

Our EI Method

precinct p_i	Candidate 1	Candidate 2	Neither
BCVAP			
HCVAP			
WCVAP			
OCVAP			

Use 4 demographic groups	Aggregate precincts statewide	Immediately infer turnout

Our EI Method

precinct p_i	Candidate 1	Candidate 2	Neither
BCVAP			
HCVAP			
WCVAP			
OCVAP			

Toggles

Use 4 demographic groups	Aggregate precincts statewide	Immediately infer turnout
Combine White + Other	Aggregate precincts countywide	Make a turnout phase

Chen's EI Method

precinct p_i	Any candidate	Neither
BCVAP		
HCVAP		
WOCVAP		

Phase 1
Turnout

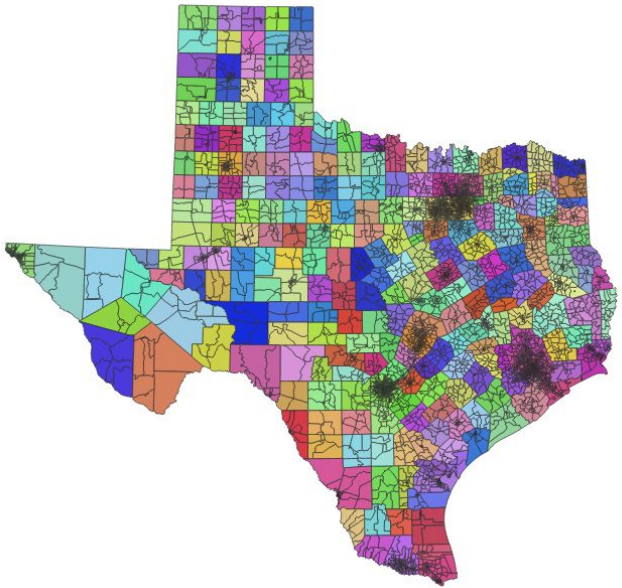
precinct p_i	Candidate 1	Candidate 2
BCVAP		
HCVAP		
WOCVAP		

Phase 2
Candidate Preference

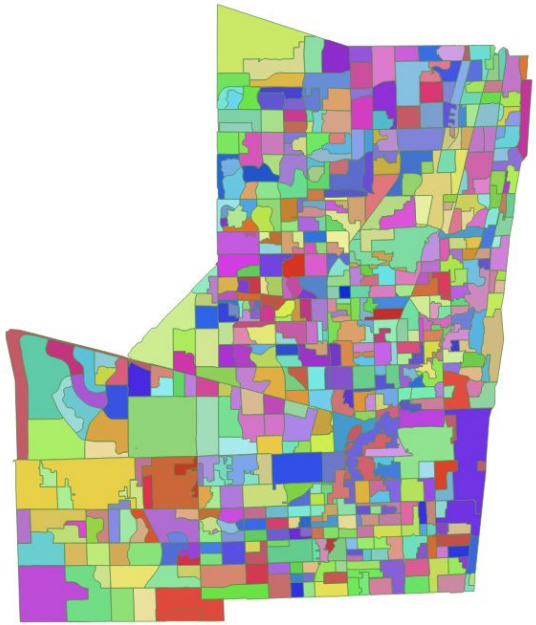
Toggles

Use 4 demographic groups	Aggregate precincts statewide	Immediately infer turnout
Combine White + Other	Aggregate precincts countywide	Make a turnout phase

Data



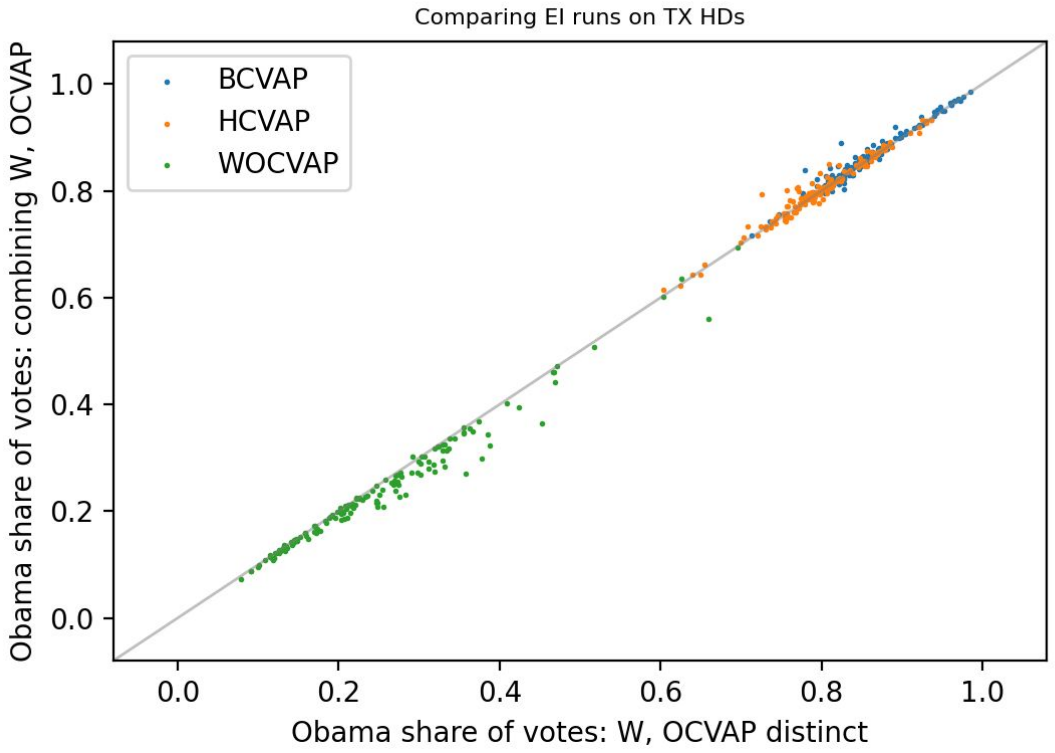
Texas
9083 precincts
~8-14 hours/election



Broward Co., FL
576 precincts
~15 min/election

Results: Consolidating Demographic Groups

- Election: PRES12 General
- Aggregate precinct results to Texas state house districts
- Consolidating tends to slightly underestimate WCVAP support for Obama, *relative to not consolidating*

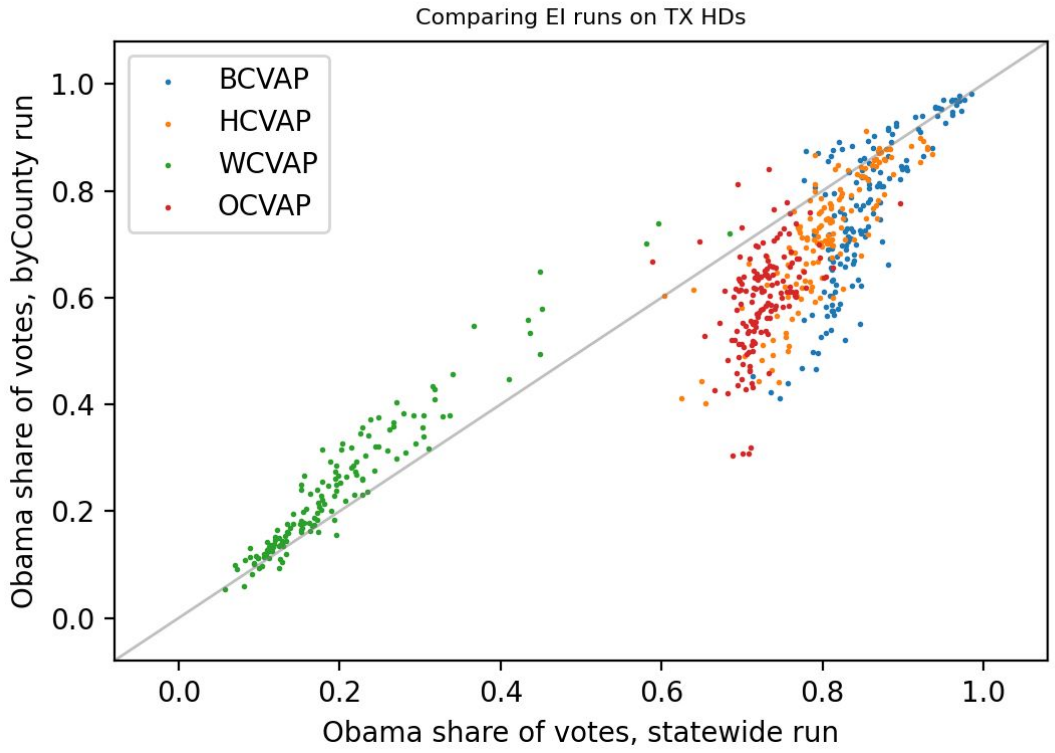


Use 4 demographic groups
Combine White + Other

Results: Statewide vs. County-by-County

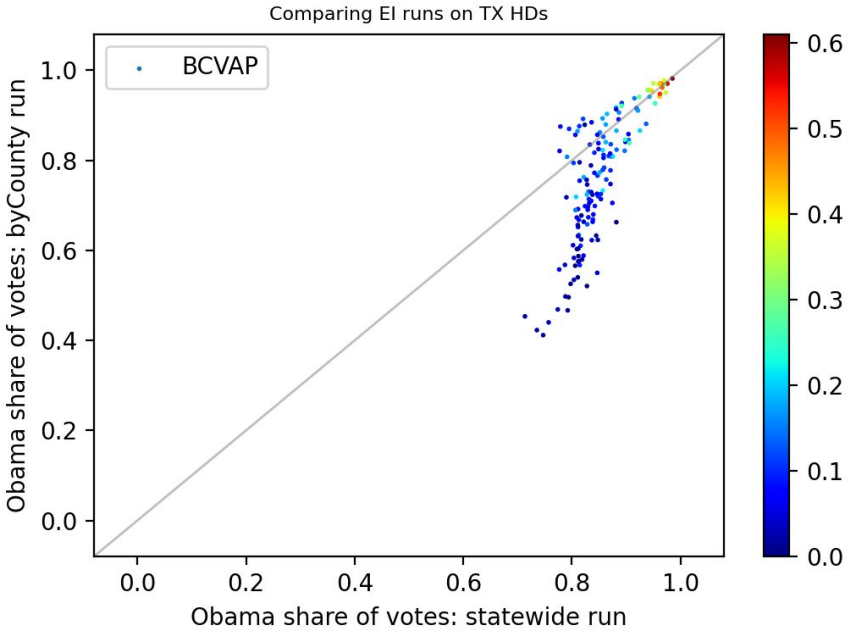
- Election: PRES12 General
- Aggregate precinct results to Texas state house districts
- byCounty tends to overestimate WCVAP support for Obama, *relative to statewide run*
- Underestimates for BCVAP, HCVAP, OCVAP

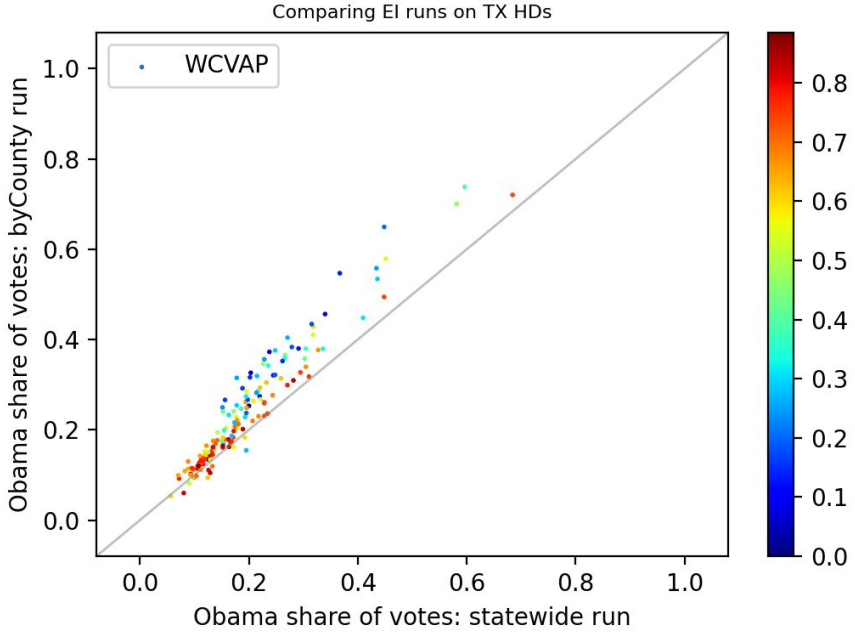
Aggregate precincts statewide
Run EI county-by-county



Results: Statewide vs. County-by-County

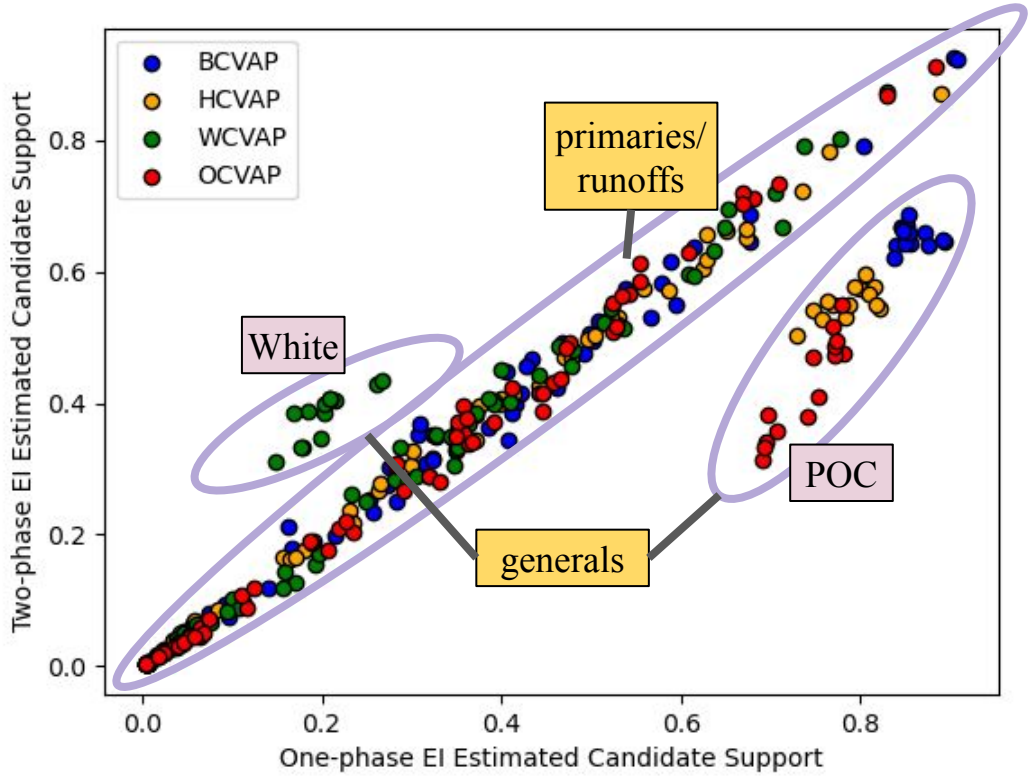
- Which HDs are most affected by this toggle?



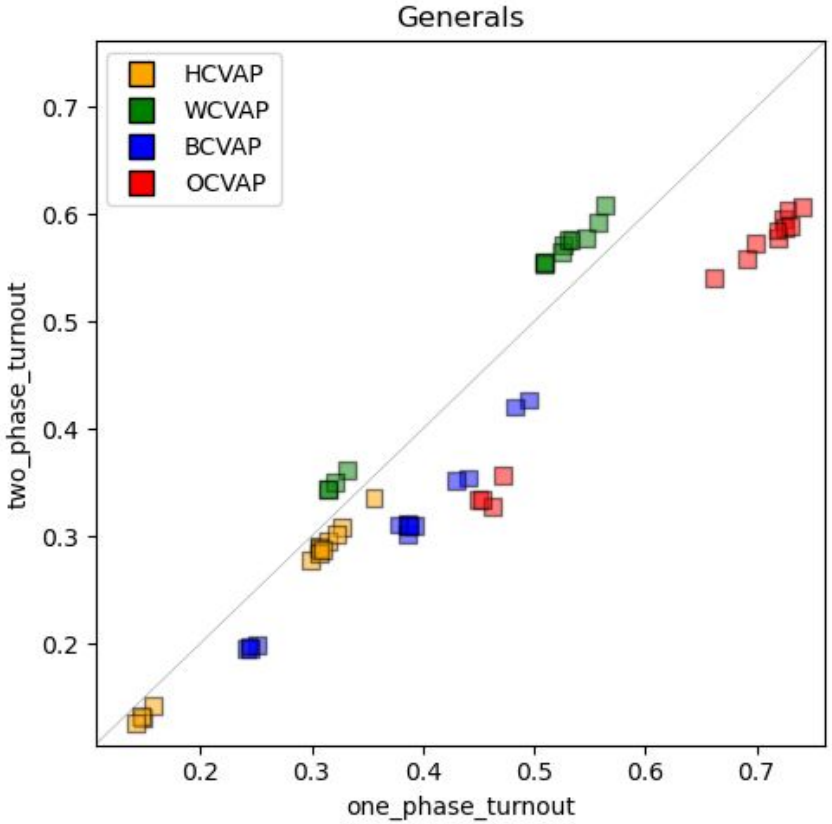


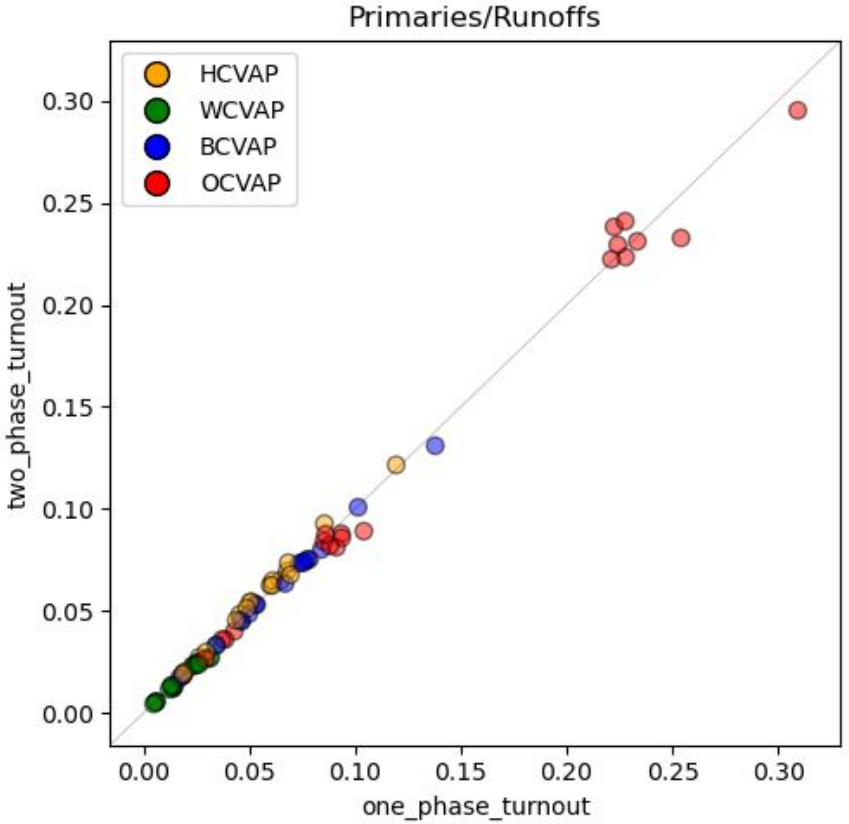
1 Phase vs 2 Phase: Texas Results

One Phase: Estimate candidate support and turnout in single run, using ‘abstain’/‘neither’ column (i.e. Rx(C+1))
Two Phase: First estimate turnout-by-race (Rx2), then feed output into second run to estimate candidate support (RxC)



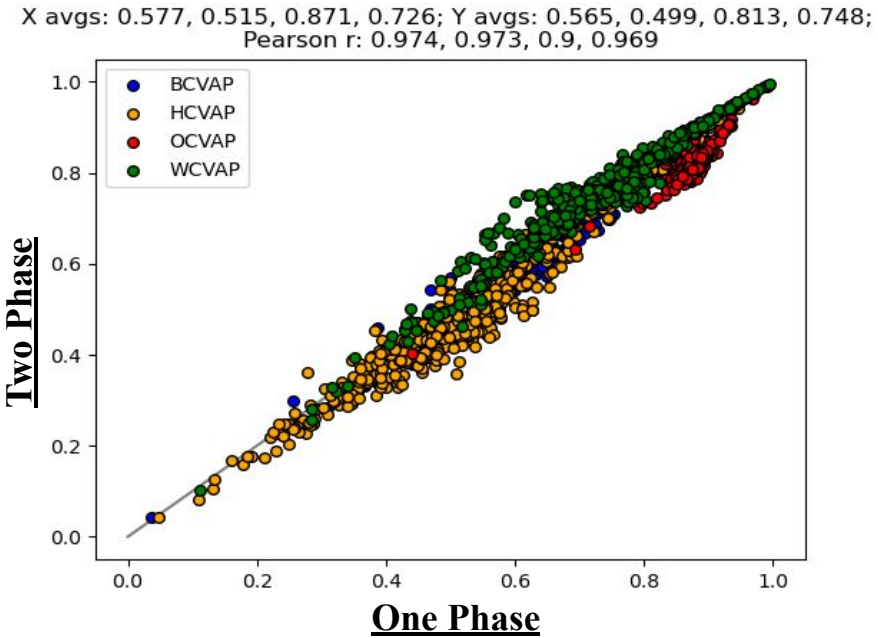
1 Phase vs 2 Phase: Texas Results





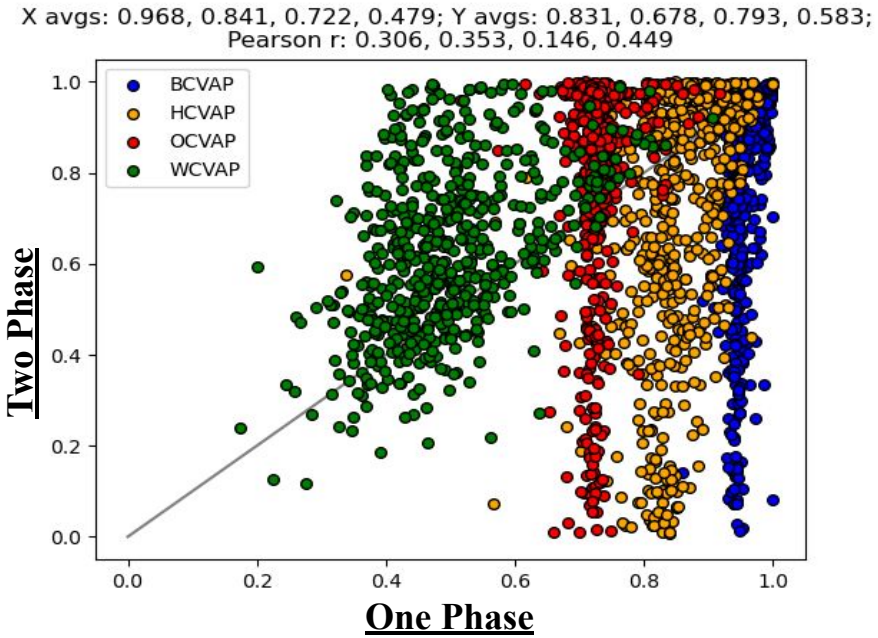
1 Phase vs 2 Phase: Broward 2016 Presidential Results

Turnout



	BCVAP	HCVAP	OCVAP	WCVAP
One Phase	57%	52%	87%	73%
Two Phase	57%	50%	81%	75%

Support Clinton

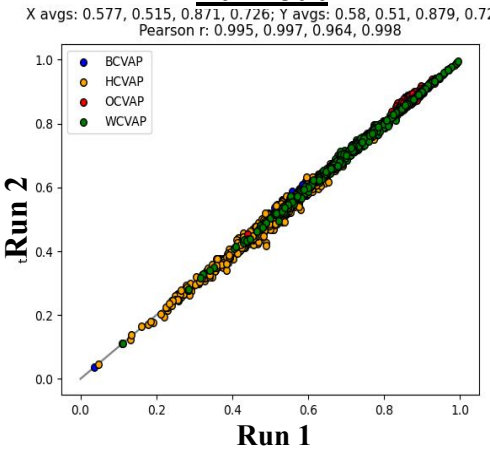


	BCVAP	HCVAP	OCVAP	WCVAP
One Phase	97%	84%	72%	48%
Two Phase	83%	68%	79%	58%

1 Phase vs 2 Phase: Replicability

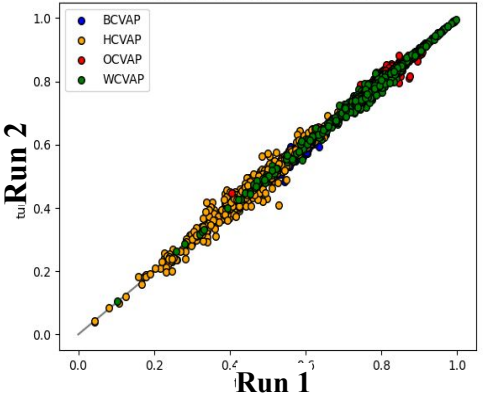
Turnout

One Phase



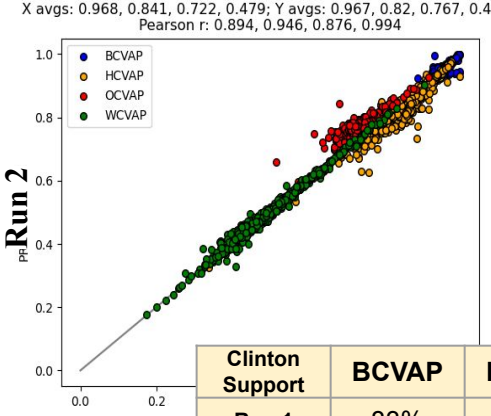
	Pearson r
BCVAP	.995
HCVAP	.997
OCVAP	.964
WCVAP	.998

Two Phase



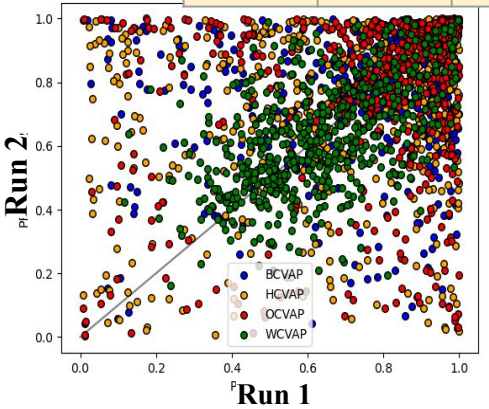
	Pearson r
BCVAP	.993
HCVAP	.992
OCVAP	.918
WCVAP	.996

Support Clinton



	Pearson r
BCVAP	.894
HCVAP	.946
OCVAP	.876
WCVAP	.994

Clinton Support	BCVAP	HCVAP	OCVAP	WCVAP
Run 1	83%	68%	79%	58%
Run 2	84%	69%	78%	58%



	Pearson r
BCVAP	.270
HCVAP	.268
OCVAP	.255
WCVAP	.555

Next Steps

- Understand replicability issue with Two Phase
- Use FL voter file to see how ground-truth turnout-by-race compares to EI estimates
- Repeat analysis for other elections (including primaries)
- Try replicating using PyEI