Assessing Balance in Natural Experiments with ML

Tyler Reny
Bryan Wilcox

UCLA Political Science

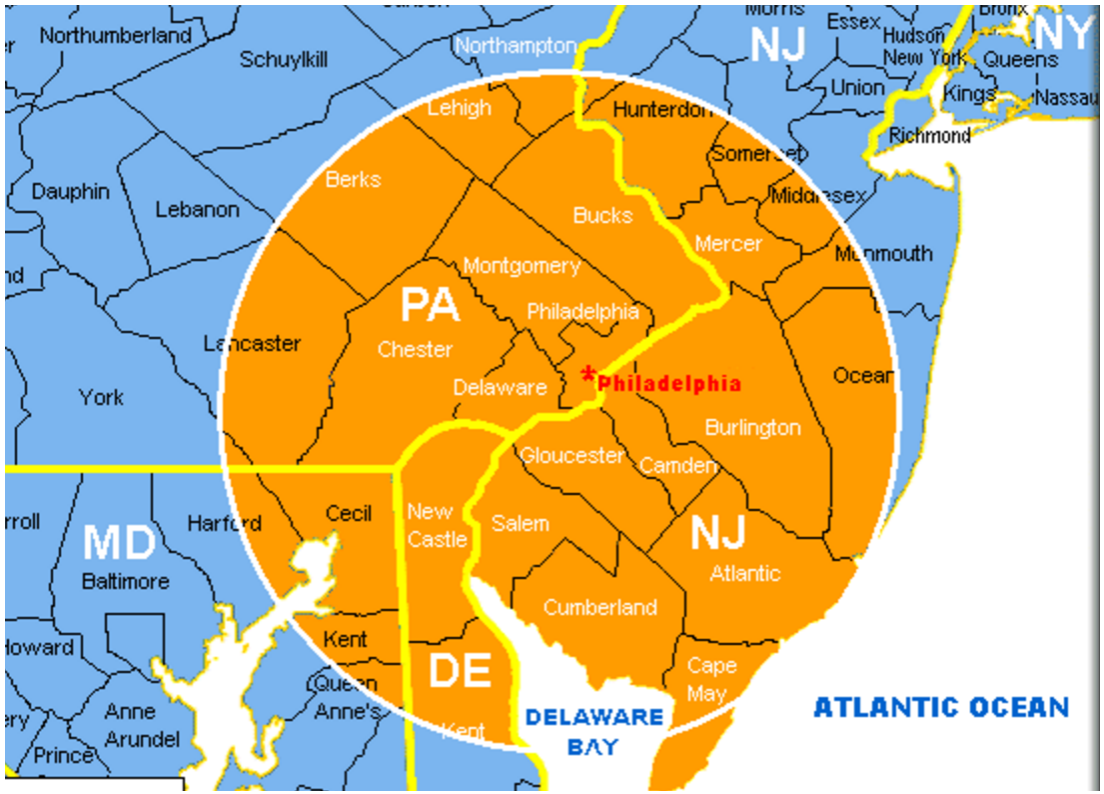
March 10, 2016

Balance in Natural Experiments

- Natural experiments increasingly popular in political science
- Often very cursory attempts to assess balance across 'control' and 'treatment' groups
- Mean balance on observables suggests as-if random assignment of treatment
- But on what functions of the covariates should we be checking balance? What about issues of non-linearity/non-additivity?

Motivating Example

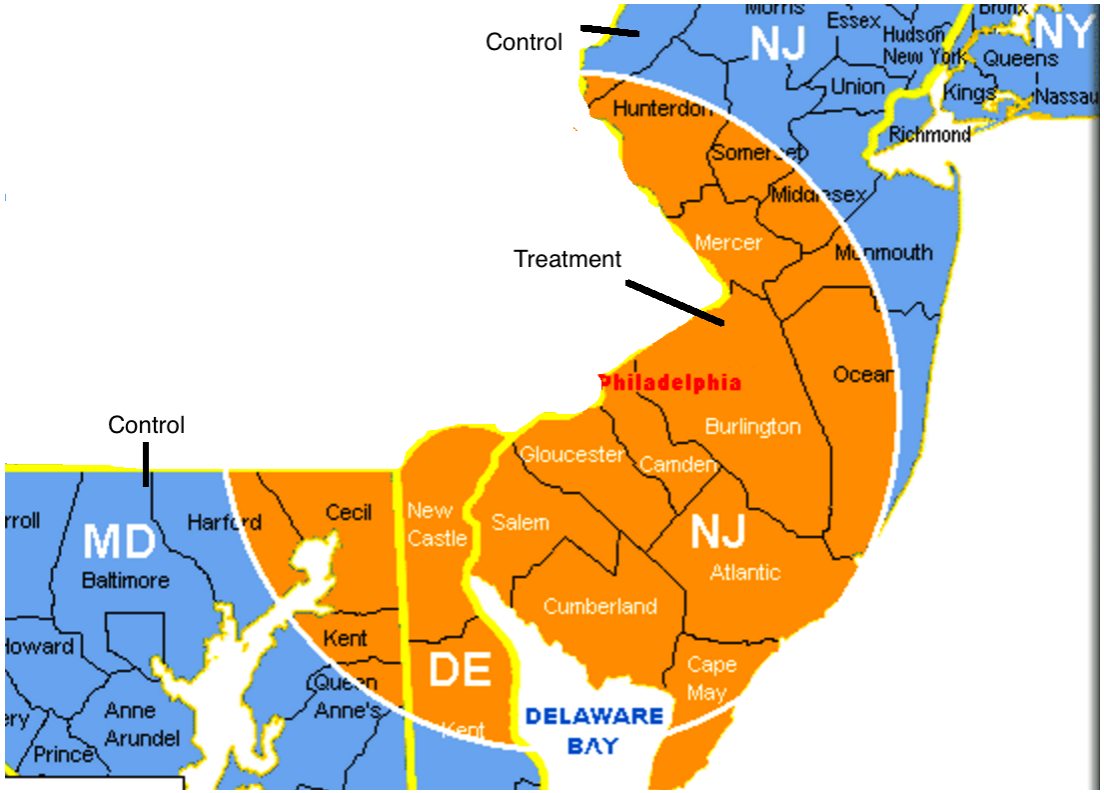
Huber and Arceneaux (2007) "Identifying Persuasive Effects of Presidential Advertising" *AJPS*



No attempt to assess balance

Presidential Advertising

Are treatment and control groups really identical?



Question

- Can machine learning provide tools to better assess balance for natural experiments?

Classification: Treatment V. Control

- Specifically, can we predict where someone lived geographically based on their covariates? (ideology, partisan ID, church attendance, union membership, income, missing income, employed, education, race and ethnicity, gender, age, media consumption, health insurance coverage)
- Data: 2000 NAES Cross Section (8,320 observations in 20 non-battleground states)
- Discriminant and Generative Methods: GLM, QDA, LDA, Naive Bayes, SVM, KRLS, KR Logit, Lasso Logit
- Train (5253); Test (1750)

Confusion Matrices - Generative

True Treated: 54.5%

Logit

	0	1
0	0.042	0.036
1	0.417	0.504
		54.6%

QDA

	0	1
0	0.162	0.298
1	0.155	0.385
		54.7%

LDA

	0	1
0	0.042	0.418
1	0.035	0.505
		54.7%

NBC

	0	1
0	0.089	0.371
1	0.079	0.461
		55%

Confusion Matrices - Discriminant

	SVM	
	0	1
0	0.000	0.000
1	0.46	0.54
		54%

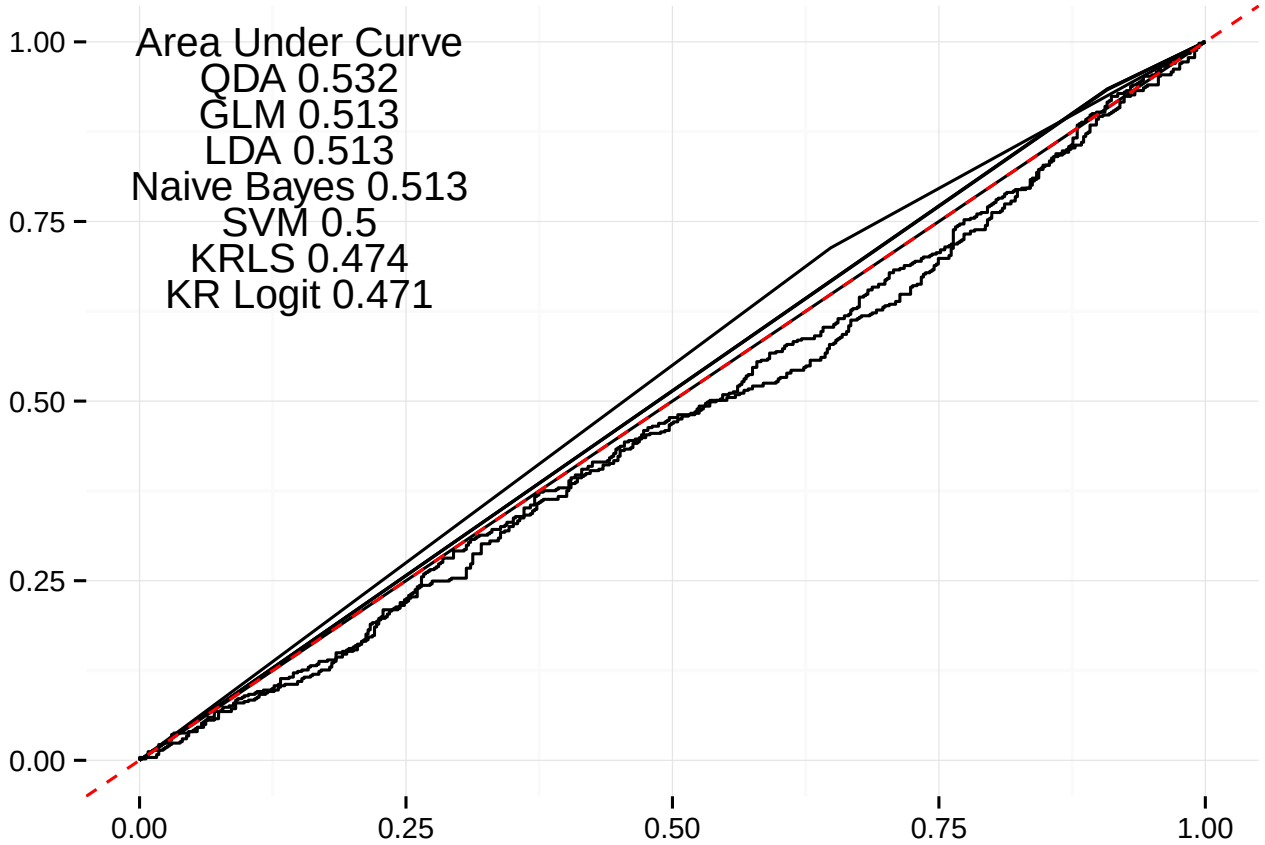
	KRLS	
	0	1
0	0.000	0.499
1	0.000	0.501
		50.1%

	KR Logit	
	0	1
0	0.000	0.499
1	0.000	0.501
		50.1%

Lasso Logit

```
18 x 1 sparse Matrix of class "dgCMatrix"
                            1
(Intercept)      0.1867199
ideology         .
pid5             .
attend_church    .
union            .
income           .
income_dkna      .
employed         .
education        .
hispanic         .
white            .
female           .
age              .
watch_tv         .
newspaper        .
talk_radio       .
married          .
health_insurance .
```

ROC Plot


Area Under Curve
QDA 0.532
GLM 0.513
LDA 0.513
Naive Bayes 0.513
SVM 0.5
KRLS 0.474
KR Logit 0.471
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00

Do we see balance at the state level?

- While there is evidence of balance in the larger sample (national level), do we see balance among individual state subsamples?
- We may be interested in analyzing sub-groups where we don't see balance
  - Check balance among subgroup
  - Use matching to process on observables
  - Re-check balance after matching
- Here is an example of California

Confusion Matrices - California

Logit

	0	1
0	0.00	0.00
1	0.27	0.73
		73%

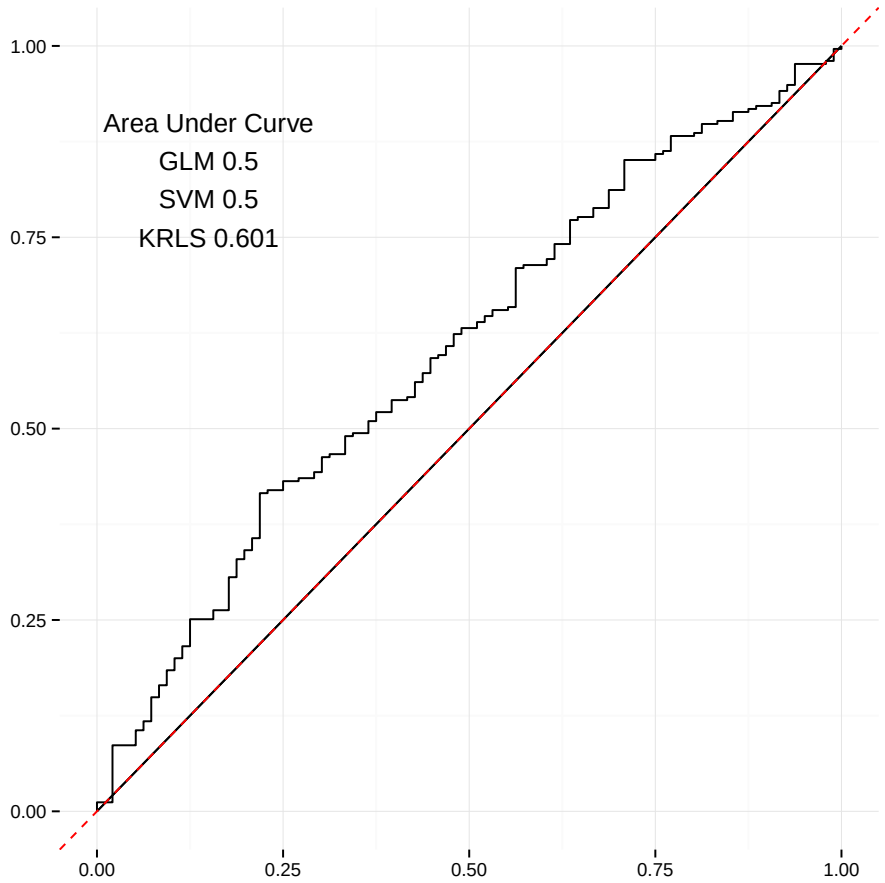
KRLS

	0	1
0	0.000	0.27
1	0.000	0.73
		73%

SVM

	0	1
0	0.000	0.000
1	0.27	0.73
		73%

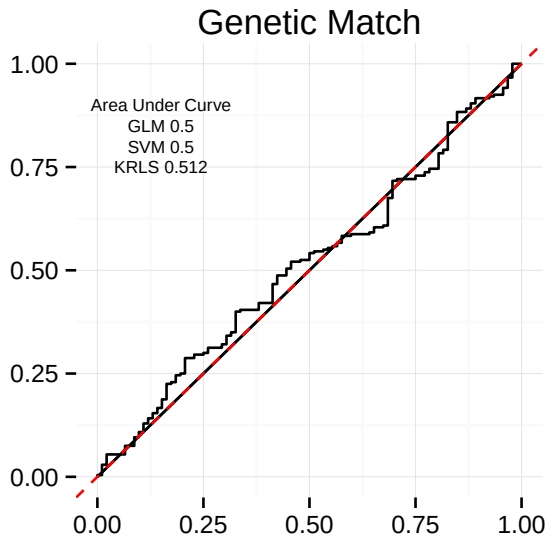
California

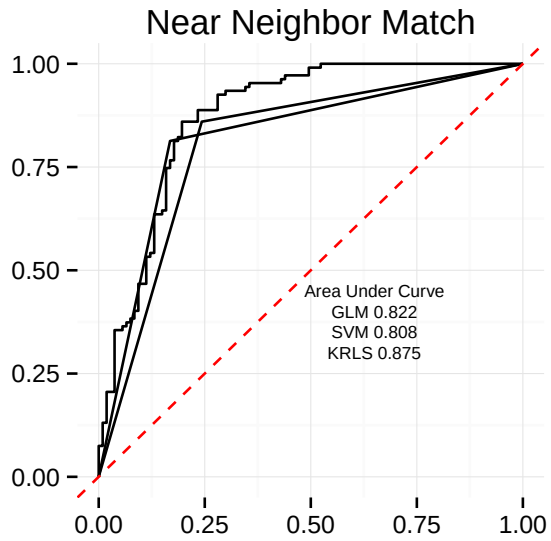

Area Under Curve
GLM 0.5
SVM 0.5
KRLS 0.601
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00

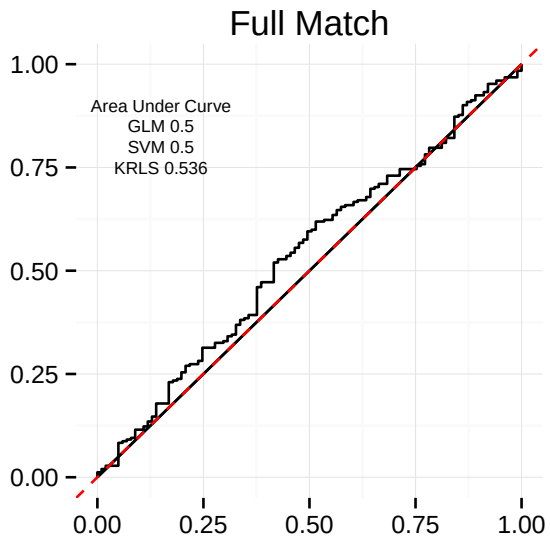
Matching to balance on observables

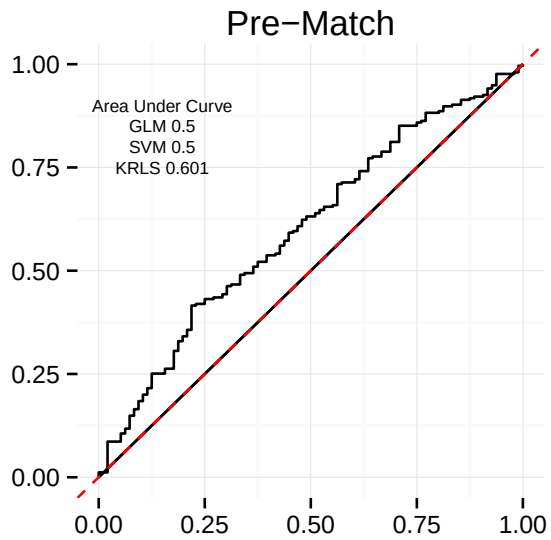
- **Nearest neighbor matching** (1:1)
- **Genetic matching** uses a genetic search algorithm to choose weights that optimize a loss function
- **Full matching** minimizes a weighted average of the estimated distance measure between each treatment and each control within a subclass

ROC in post-matched samples









Next steps & concluding thoughts

- Machine learning techniques are a more powerful diagnostic tool to assess balance across treatment and control groups in observational data
- Balance in means across covariates not sufficient, may not pick up other forms of imbalance in samples (i.e., $\frac{X_1}{X_2}$)
- Use a matching algorithm that will assess balance across other functions of covariates (`KBAL`, `GenMatch`), could help alleviate problem (at least on observables) when/if present