Reasoning about Interference: Randomization Based Statistical Inference about Causal Models of Interference Between Units

Jake Bowers[1,2] Mark Fredrickson[3] Nahomi Ichino[4]

Zubizarreta Causal inference Class, Columbia University

[1]Political Science & Statistics & NCSA @ University of Illinois

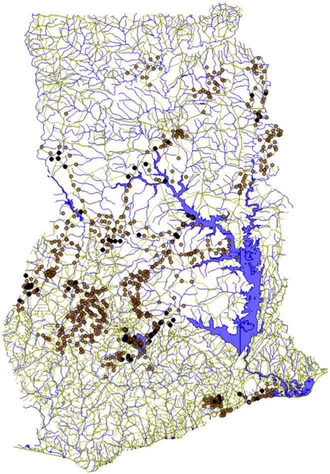
[2]White House Social and Behavioral Sciences Team

jwbowers@illinois.edu — http://jakebowers.org

[3]Statistics & Political Science @ University of Illinois

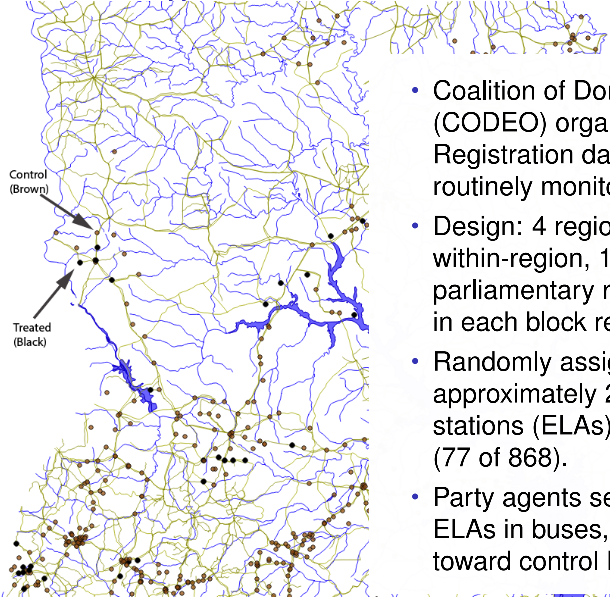
[4]Department of Political Science @ University of Michigan

Voter Registration in Ghana 2008



- Presidential and parliamentary elections in December 2008.
- 13 day voter registration exercise in August 2008.
- Estimated 800,000 people newly eligible to vote, but **2 million** new voters registered.
- Term-limited president, election expected to be very close. Decided by less than 50,000 votes out of more than 9 million votes cast.

Voter Registration in Ghana 2008



- Coalition of Domestic Election Observers (CODEO) organize registration observers. Registration day was generally **not** routinely monitored.
- Design: 4 regions (non-random); within-region, 13 blocks by 2004 parliamentary results; 1 of 3 constituencies in each block receives observers (random).
- Randomly assign observers to approximately 25% of election polling stations (ELAs) in selected constituency (77 of 868).
- Party agents seen approaching treated ELAs in buses, and then driving away toward control ELAs.

What is the true effect of the treatment assignment?

We don't know the truth...

... but I can provide a best guess of an ATE:

$ATE \equiv \bar{y}_{i1} - \bar{y}_{i0}$

$E_R(\widehat{ATE}) = ATE$

... but I can assess specific claims about the truth:

$H_0: y_{i1} = y_{i0}$

$p = ?$

What is the true effect of the treatment assignment?

We don't Know.

i	Z_i	Y_i	y_{i1}	y_{i0}
A	0	16	?	16
B	1	22	22	?
C	0	7	?	7
D	1	14	14	?
			$\bar{y}_{i1}$	$\bar{y}_{i0}$

$$\widehat{ATE} = \bar{Y}_i|Z_i=1 - \bar{Y}_i|Z_i=0$$

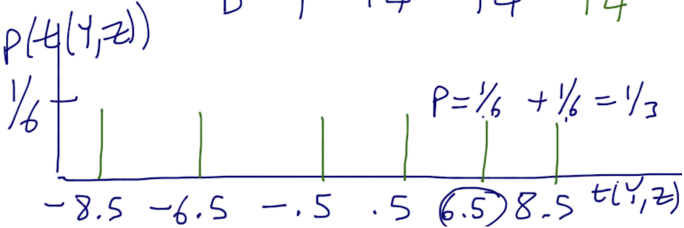
$$= \frac{22+14}{2} - \frac{16+7}{2} = 6.5$$

What is the true effect of the treatment assignment?

We don't know.

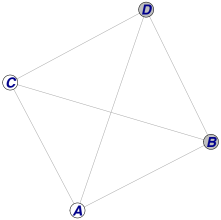
$H_0: y_{i1} = y_{i0}$

i	Z_i	Y_i	y_{i1}	y_{i0}
A	0	16	?	16
B	1	22	22	22
C	0	7	?	7
D	1	14	14	14



$p = 1/6 + 1/6 = 1/3$

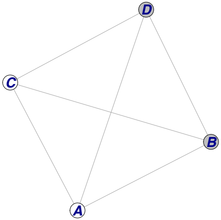
Statistical inference for counterfactual quantities with interference?



i	Z_i	Y_i	$y_{i,1100}$	$y_{i,0101}$	$y_{i,1001}$	$y_{i,0110}$	$y_{i,1010}$	$y_{i,0011}$
A	0	16	?	16	?	?	?	?
B	1	22	?	22	?	?	?	?
C	0	7	?	7	?	?	?	?
D	1	14	?	14	?	?	?	?

Lots of good work in estimation area (Sobel, Aronow, Samii, Hudgens, Ogburn, VanderWeele, Toulis, Kao, Coppock, Sicar, Raudenbush, Hong, . . .). What is the function of potential outcomes that we can estimate using observed data? The trick is to relate a given observed difference to complicated potential outcomes.

Statistical inference for counterfactual quantities with interference?



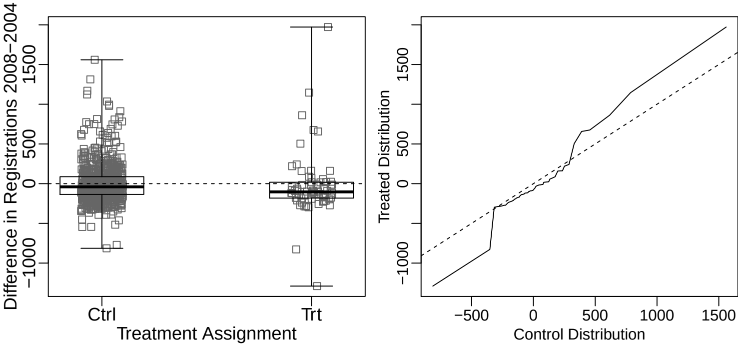
i	Z_i	Y_i	$y_{i,1100}$	$y_{i,0101}$	$y_{i,1001}$	$y_{i,0110}$	$y_{i,1010}$	$y_{i,0011}$	$y_{i,0000} \equiv y_{i,\mathbf{0}}$
A	0	16	?	16	?	?	?	?	16
B	1	22	?	22	?	?	?	?	22
C	0	7	?	7	?	?	?	?	7
D	1	14	?	14	?	?	?	?	14

The sharp null of no effects is a model of no interference:

$H_0 : y_{i,1100} = y_{i,0101} = y_{i,1001} = y_{i,0110} = y_{i,1010} = y_{i,0011} = y_{i,0000}$,
$y_{i,\mathbf{0}} = \mathcal{H}(y_{i,\mathbf{z}}, \mathbf{0}) = y_{i,\mathbf{z}}$, $p = 0.33$.

Introducing the **uniformity trial** $\equiv \mathbf{y}_{i,0000}$ (Rosenbaum, 2007).

Assessing the sharp null hypothesis of no effects.



"What is the probability of seeing as large an observed difference between the treated and control groups, if the observers had no effect at all?" $p = 0.018$

Models of Some Causal Effect

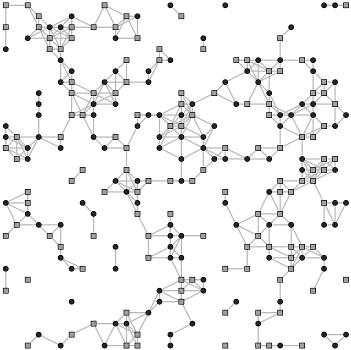


Figure: A simulated data set with 256 units and 512 connections. The $256/2 = 128$ treated units ($Z_i = 1$) are shown as filled circles and an equal number of control units ($Z_i = 0$) are shown as as gray squares.

Models of Some Causal Effect

$$\mathcal{H}(\mathbf{y_0}, \mathbf{z}, \beta, \tau) = \left[\beta + (1 - z_i)(1 - \beta)\exp\left(-\tau^2 \mathbf{z}^T \mathbf{S}\right)\right] \mathbf{y_0} \qquad (1)$$

$$\mathcal{H}(\mathbf{y_z}, \mathbf{0}, \beta, \tau) = \left[\beta + (1 - z_i)(1 - \beta)\exp\left(-\tau^2 \mathbf{z}^T \mathbf{S}\right)\right]^{-1} \mathbf{y_z} \equiv \mathbf{y_0} \qquad (2)$$

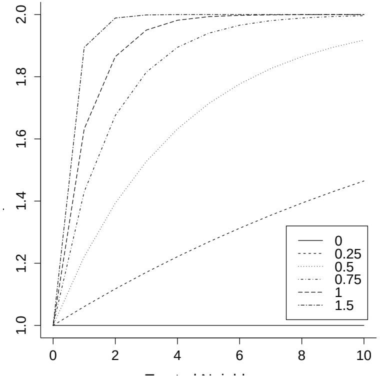


Figure: Growth curve of spillover effects for the expression $\beta + (1 - \beta)\exp\left(-\tau^2 \mathbf{z}^T \mathbf{S}\right)$ as the number of treated neighbors, $\mathbf{z}^T \mathbf{S}$, increases for $\beta = 2$ and a selection of τ values.

Different Test Statistic Proposals

"Select a test statistic [$\mathcal{T}$] that will be small when the treated and control distributions in the adjusted data ... are similar, and large when the distributions diverge." (Rosenbaum 2002, §2.4.4)

So far:

KS If the empirical cumulative distribution function (ECDF) of the treated units is F_1 and the ECDF of the control units is F_0 then the KS test statistic is $\mathcal{T}(\mathbf{y_0}, \mathbf{z})_{\text{KS}} = \max_{i=1,\ldots,n} [F_1(y_{i,\mathbf{0}}) - F_0(y_{i,\mathbf{0}})]$, where $F(x) = (1/n) \sum_{i=1}^{n} I(x_i \leq x)$ records the proportion of the distribution of x at or below x_i.

KS+Net Collect Z_i,$\mathbf{z}^T\mathbf{S}$ (number of treated neighbors), and $\mathbf{1}^T\mathbf{S}$ (number of neighbors) into a matrix $\mathbf{X}$, and fit the $y_{i,\mathbf{0}}$ as a linear function of $\mathbf{X}$ with coefficients β. Replaces $y_{i,\mathbf{0}}$ with $e_{i,\mathbf{0}} = y_{i,\mathbf{0}} - (\mathbf{1}^T\mathbf{S})\hat{\beta})$ in the expression above.

SSR+Net Define the SSR test statistic as:

$$\mathcal{T}(\mathbf{y_0}, \mathbf{z})_{\text{SSR}} \equiv \sum_i (y_{i,\mathbf{0}} - \mathbf{X}\hat{\beta})^2 \tag{3}$$

Power differences among test statistics

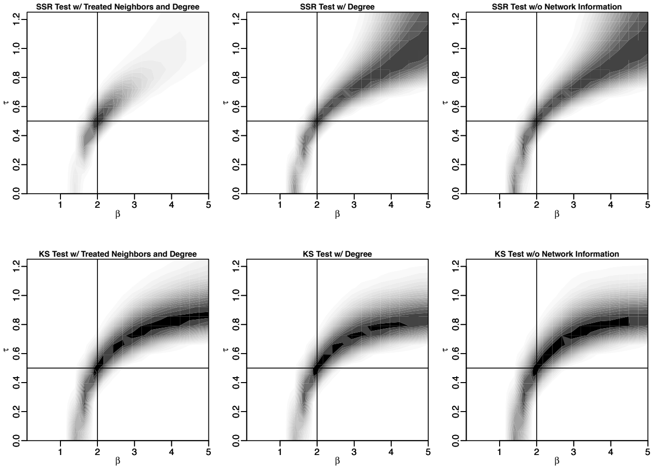


Figure: Proportion of *p*-values less than .05 for randomization tests of joint hypotheses about τ and β. Darker values mean less rejection. Truth is at $\tau = .5, \beta = 2$. All tests reject the truth no more than 5% of the time at $\alpha = .05$. All simulations, not Normal approximations.

Sharp No-Effects Data, Spillover Model

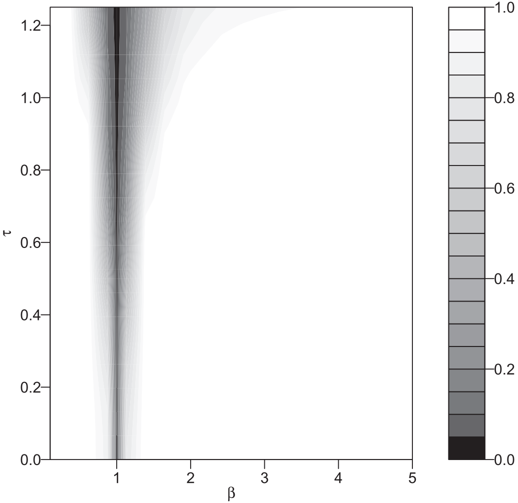


Fig. 15 Proportion of $p < .05$ for joint hypotheses about β and τ as defined in equation (6) when the true model is the sharp null model (setting $\beta = 1$ in the simulation engine as described in Section 5.1.6).

Constant-Effects Data, Spillover Model

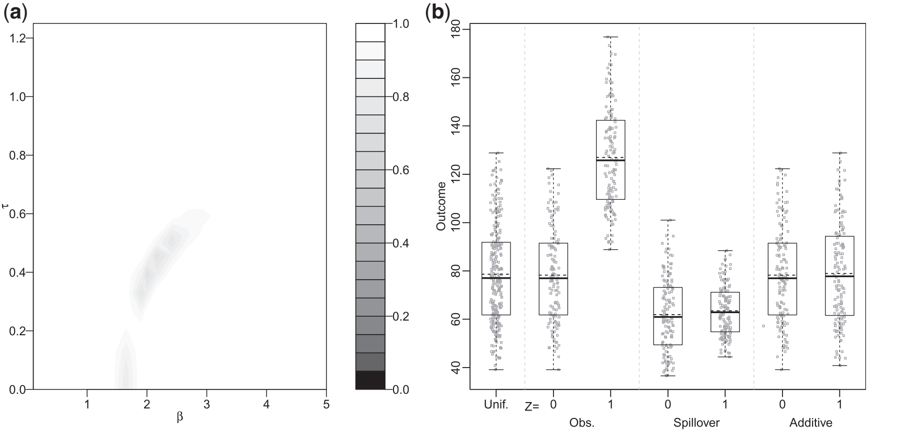


Fig. 17 Simulation results for data generated by the true additive model (equation (8), $\alpha = 48$) and tested using the spillover model (equation (6)). (a) Proportion of hypotheses rejected for hypotheses generated by the spillover model applied to the data generated by the additive model. Minimum achieved at $\beta = 2$ and $\tau = 0.395$. (b) Example observed data (from one of the 1000 trials) and adjustments implied by the true additive model and the spillover model at $\beta = 2$ and $\tau = 0.395$.

What can we learn from this experiment?

In general: Why do so many new democracies seem so unstable? Why so many coups and protests and civil wars?

One specific idea: Ethnic parties slow consolidation of democratic norms.

How might this dynamic play out in Ghana 2008?

- Party agents are in charge of registering voters (honestly and dishonestly). They mobilize potential voters (for example, in buses). They get paid for fraud (in part).
- Party agents want to register as many people using as few resources as possible (and with as little risk as possible). They know that many voters in Ghana (where the political parties are strongly associated with particular ethnicities):
  - Prefer to have a co-ethnic in office who is more likely to favor them than a non-co-ethnic politician
  - Believe that co-ethnic leaders matters for local public goods
  - Anticipate a close election, citizens may not report registration fraud
- So, agents may target ethnically-homogeneous areas where it's less likely they'll be reported
- *Alternatively:* potential reporting by ordinary citizens may not be a concern, and distances/resources may be a more important factor.

What can we learn from this experiment?

In general: Why do so many new democracies seem so unstable? Why so many coups and protests and civil wars?

One specific idea: Ethnic parties slow consolidation of democratic norms.

How might this dynamic play out in Ghana 2008?

- Party agents are in charge of registering voters (honestly and dishonestly). They mobilize potential voters (for example, in buses). They get paid for fraud (in part).
- Party agents want to register as many people using as few resources as possible (and with as little risk as possible). They know that many voters in Ghana (where the political parties are strongly associated with particular ethnicities):
  - Prefer to have a co-ethnic in office who is more likely to favor them than a non-co-ethnic politician
  - Believe that co-ethnic leaders matters for local public goods
  - Anticipate a close election, citizens may not report registration fraud
- So, agents may target ethnically-homogeneous areas where it's less likely they'll be reported
- *Alternatively:* potential reporting by ordinary citizens may not be a concern, and distances/resources may be a more important factor.

What can we learn from this experiment?

In general: Why do so many new democracies seem so unstable? Why so many coups and protests and civil wars?

One specific idea: Ethnic parties slow consolidation of democratic norms.

How might this dynamic play out in Ghana 2008?

- Party agents are in charge of registering voters (honestly and dishonestly). They mobilize potential voters (for example, in buses). They get paid for fraud (in part).
- Party agents want to register as many people using as few resources as possible (and with as little risk as possible). They know that many voters in Ghana (where the political parties are strongly associated with particular ethnicities):
  - Prefer to have a co-ethnic in office who is more likely to favor them than a non-co-ethnic politician
  - Believe that co-ethnic leaders matters for local public goods
  - Anticipate a close election, citizens may not report registration fraud
- So, agents may target ethnically-homogeneous areas where it's less likely they'll be reported
- *Alternatively:* potential reporting by ordinary citizens may not be a concern, and distances/resources may be a more important factor.

Formal decision-theoretic models

k The total number of agents.

t The total number of "ticks" or time periods, in which agents can visit ELAs.

τ The number of false registrants an agent can add to an unobserved ELA.

Distance-minimizing model In each "tick," agents go to the nearest ELA by road distance (start at ELA nearest the most others); if they encounter an observer, immediately move to the nearest ELA from there. Cannot revisit ELAs. Implies starting at ELA that are close to others.

Ethnic homogeneity-seeking model In each "tick," agents only consider moving to an ELA with $F \leq \alpha$ where F is ethnic fractionalization and α is a percentile of F within a constituency. Among these, move to the closest ELA by road distance, and move again if encounter observer. Implies starting at ELA with lowest F.

Uniformity: model without observers

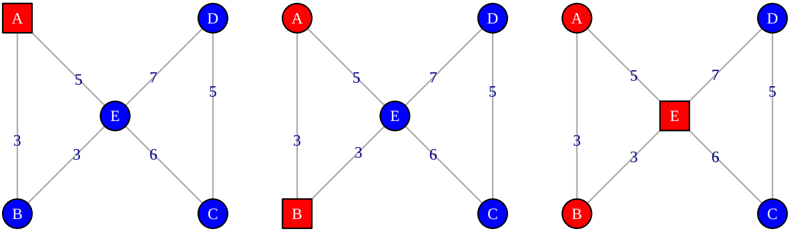


Figure: Agent movement rules when no observers are encountered. Squares indicate the agent's current location. Red ELAs are visited. Blue ELAs are not yet visited. From left to right: 1) $t = 0$, the agent starts at A, 2) agent selects B as closest ELA, 3) Agent moves to E in final period.

Experiment: model with observers

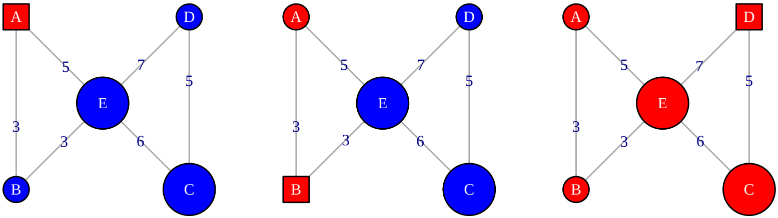


Figure: Agent movement rules when observers are present. Squares indicate the agent's current location. Red ELAs are visited. Blue ELAs are not yet visited. The large circles indicate observer ELAs. From left to right: 1) $t = 0$, the agent starts at A, 2) agent selects B as closest ELA, 3) Agent moves to E, but as an observer is present, immediately moves to C, again encounters an observer, and finally stops at D.

Assessing these models of party agents

"Which combinations of parameters would have been surprising given the observed data?"

- Pick a set of values for the parameters k, τ, t, and α to determine a path for our agents through the road network. Each set of parameters generates one sharp hypothesis as an output of the agent-based model.
- A *p*-value records the information our data and design provide against these hypotheses given the test statistic (here the Kolmogorov-Smirnov test statistic).
- *Q:* How to interpret a 4d confidence region?
- *A:* Our approach: Focus on a composite quantity, $T = \sum_{i=1}^{n}(Y_i - y_{i0})$, where y_{i0} is the number of registrations implied by the model under the uniformity trial.(inspired by Rosenbaum's attributable effects, $A = \sum_i Z_i \tau_i$) If the *minimum* *p*-value for all hypotheses that make up a given T is greater than .05, then this T is in the confidence set.

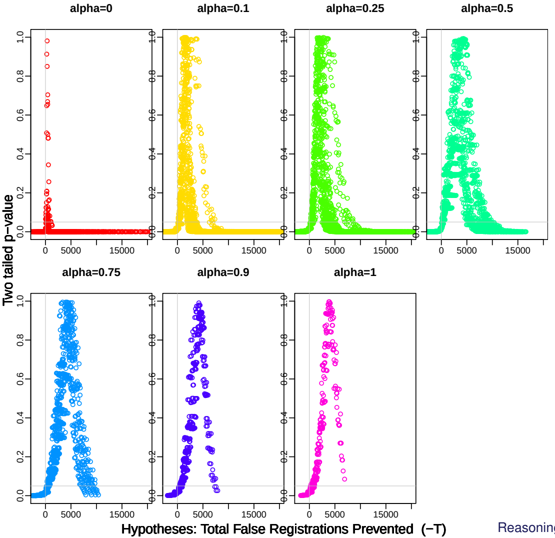
Assessing these models of party agents

"Which combinations of parameters would have been surprising given the observed data?"

- Pick a set of values for the parameters k, τ, t, and α to determine a path for our agents through the road network. Each set of parameters generates one sharp hypothesis as an output of the agent-based model.
- A *p*-value records the information our data and design provide against these hypotheses given the test statistic (here the Kolmogorov-Smirnov test statistic).
- *Q:* How to interpret a 4d confidence region?
- *A:* Our approach: Focus on a composite quantity, $T = \sum_{i=1}^{n}(Y_i - y_{i0})$, where y_{i0} is the number of registrations implied by the model under the uniformity trial.(inspired by Rosenbaum's attributable effects, $A = \sum_i Z_i \tau_i$) If the *minimum* *p*-value for all hypotheses that make up a given T is greater than .05, then this T is in the confidence set.

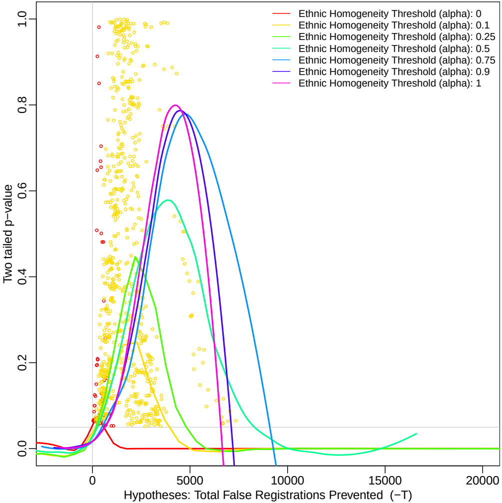
Testing the models

Notes: $\alpha = 0$ (only the most homogeneous ELAs available), $\alpha = .1$ (only 10% of ELAs available),. . ., $\alpha = .9$ (nearly all ELAs available).



Testing the models

Notes: $\alpha = 0$ (only the most homogeneous ELAs available), $\alpha = .1$ (only 10% of ELAs available),. . ., $\alpha = .9$ (nearly all ELAs available).



Investigating Model Implications

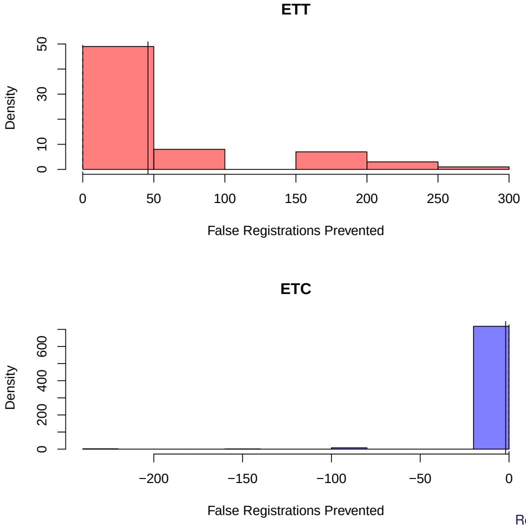
- Consider a model that maximizes the *p*-value:
  - $(\alpha = 0.25, k = 137, t = 9, \tau = 48)$
  - 137 agents only visit ELAs with less than 25th percentile in homogeneity.
  - Each agent visits 9 ELAs.
  - Each agent inflicts 48 false registrations per ELA visited.
- Use the model and the parameters to recover the *uniformity trial*:

$$\widetilde{\mathbf{y_0}} = \mathcal{M}(\mathbf{y_z}, \mathbf{z}, \alpha = 0.25, k = 137, t = 9, \tau = 48)$$

- Compare the uniformity trial to the observed outcome to establish treatment effects:

$$\widetilde{\mathbf{ETT}} = \mathbf{y_z} - \widetilde{\mathbf{y_0}} \mid z_i = 1$$
$$\widetilde{\mathbf{ETC}} = \mathbf{y_z} - \widetilde{\mathbf{y_0}} \mid z_i = 0$$

What Unit Level Treatment Effects Are Implied by Our Best Model?



Summary and Next Steps

- Specific counterfactual models and randomized experiments provide a side-step around some of the difficulties of causal inference on networks.
- New questions raised:
  - How to choose test statistics for multidimensional sharp-hypothesis testing? Are there multi-dimensional “effect increasing” characteristics that we can assess for a given model?
  - Are there general classes of scientific/counterfactual models?
  - How should we interpret and display results?

This presentation contains work from:
Bowers, J.; Fredrickson, M.; Aronow, P. “Research Note: A more powerful test statistic for reasoning about interference between units.” (R & R *Political Analysis*)

Ichino, N.; Bowers, J.; Fredrickson, M.; Grady, C. “Ethnicity and Electoral Fraud in New Democracies: Modelling Political Party Agents in Ghana.” (Working paper).

Bowers, J.; Fredrickson, M.; and Panagopoulos, C. 2013. “Reasoning about Interference Between Units: A General Framework.” *Political Analysis*

Fisherian Inference Algorithm

1. Write a model ($\mathcal{H}(\mathbf{y_0}, \mathbf{z}, \theta)$) converting uniformity trial into observed data (i.e. a causal model).
2. Solve for $\mathbf{y_0}$: $\mathcal{H}(\mathbf{y_z}, \mathbf{0}, \theta_0) = \mathbf{y_0}$
3. Select $\mathcal{T}$ (KS, Anderson-Darling, Network-weighted SSR, etc. . .).
4. Compute *p*-values for substantively meaningful range of θ.

Highlights of the Method

The sharp null allows models of interference Interference need not be an assumption.

Testing procedure not estimation procedure We are not estimating causal effects but testing hypotheses about causal effects.

Background stuff No probability models of outcomes; large-samples not required although Normal approximations are available and useful for some $\mathcal{T}$ and $\mathcal{H}$ (Hansen and Bowers 2008, 2009); and some analytic results for some $\mathcal{T}$ and $\mathcal{H}$ enable exact intervals in large-samples too (Aronow and Samii 2012).

On models

- Models are mathematical functions, multiple functions can have similar adjustments to the data.
- Assessing more than one model may enhance insight (Rosenbaum, 2010).
- Our method can help eliminate the implausible, not accept the plausible.

Conclusions/Summary

New Questions

- Where do models of counterfactuals come from? Do we have advice about going from words to math?
- Math has its own logic. Some expressions for models may not be sensitive to changes in parameters. How can we assess what a given model is telling? How can we go from math to words before testing hypotheses?
- The KS-statistic is low powered for tail-differences. Recall that we are testing $t(\mathcal{H}(), \mathbf{z})$ not just $\mathcal{H}()$. Some results might tell us that our test is low powered against certain alternatives more than that we have identified a region of plausibility. How to find a better test statistic?
- How can this work learn from other modes of statistical inference and other representations of causal inference? What are the connections to ATE and other estimation frameworks (Spatial Econometrics, Network Analysis (ERGMs), etc. . .)?

New Questions

- Where do models of counterfactuals come from? Do we have advice about going from words to math?
- Math has its own logic. Some expressions for models may not be sensitive to changes in parameters. How can we assess what a given model is telling? How can we go from math to words before testing hypotheses?
- The KS-statistic is low powered for tail-differences. Recall that we are testing $t(\mathcal{H}(), \mathbf{z})$ not just $\mathcal{H}()$. Some results might tell us that our test is low powered against certain alternatives more than that we have identified a region of plausibility. How to find a better test statistic?
- How can this work learn from other modes of statistical inference and other representations of causal inference? What are the connections to ATE and other estimation frameworks (Spatial Econometrics, Network Analysis (ERGMs), etc. . .)?

New Questions

- Where do models of counterfactuals come from? Do we have advice about going from words to math?
- Math has its own logic. Some expressions for models may not be sensitive to changes in parameters. How can we assess what a given model is telling? How can we go from math to words before testing hypotheses?
- The KS-statistic is low powered for tail-differences. Recall that we are testing $t(\mathcal{H}(), \mathbf{z})$ not just $\mathcal{H}()$. Some results might tell us that our test is low powered against certain alternatives more than that we have identified a region of plausibility. How to find a better test statistic?
- How can this work learn from other modes of statistical inference and other representations of causal inference? What are the connections to ATE and other estimation frameworks (Spatial Econometrics, Network Analysis (ERGMs), etc. . .)?

New Questions

- Where do models of counterfactuals come from? Do we have advice about going from words to math?
- Math has its own logic. Some expressions for models may not be sensitive to changes in parameters. How can we assess what a given model is telling? How can we go from math to words before testing hypotheses?
- The KS-statistic is low powered for tail-differences. Recall that we are testing $t(\mathcal{H}(), \mathbf{z})$ not just $\mathcal{H}()$. Some results might tell us that our test is low powered against certain alternatives more than that we have identified a region of plausibility. How to find a better test statistic?
- How can this work learn from other modes of statistical inference and other representations of causal inference? What are the connections to ATE and other estimation frameworks (Spatial Econometrics, Network Analysis (ERGMs), etc...)?

What can we learn from this experiment?

In general: Why do so many new democracies seem so unstable? Why so many coups and protests and civil wars?

One specific idea: Ethnic parties slow consolidation of democratic norms.

How might this dynamic play out in Ghana 2008?

- Party agents are in charge of registering voters (honestly and dishonestly). They mobilize potential voters (for example, in buses). They get paid for fraud (in part).
- Party agents want to register as many people using as few resources as possible (and with as little risk as possible). They know that many voters in Ghana (where the political parties are strongly associated with particular ethnicities):
  - Prefer to have a co-ethnic in office who is more likely to favor them than a non-co-ethnic politician
  - Believe that co-ethnic leaders matters for local public goods
  - Anticipate a close election, citizens may not report registration fraud
- So, agents may target ethnically-homogeneous areas where it's less likely they'll be reported
- *Alternatively:* potential reporting by ordinary citizens may not be a concern, and distances/resources may be a more important factor.

What can we learn from this experiment?

In general: Why do so many new democracies seem so unstable? Why so many coups and protests and civil wars?

One specific idea: Ethnic parties slow consolidation of democratic norms.

How might this dynamic play out in Ghana 2008?

- Party agents are in charge of registering voters (honestly and dishonestly). They mobilize potential voters (for example, in buses). They get paid for fraud (in part).
- Party agents want to register as many people using as few resources as possible (and with as little risk as possible). They know that many voters in Ghana (where the political parties are strongly associated with particular ethnicities):
  - Prefer to have a co-ethnic in office who is more likely to favor them than a non-co-ethnic politician
  - Believe that co-ethnic leaders matters for local public goods
  - Anticipate a close election, citizens may not report registration fraud
- So, agents may target ethnically-homogeneous areas where it's less likely they'll be reported
- *Alternatively:* potential reporting by ordinary citizens may not be a concern, and distances/resources may be a more important factor.

What can we learn from this experiment?

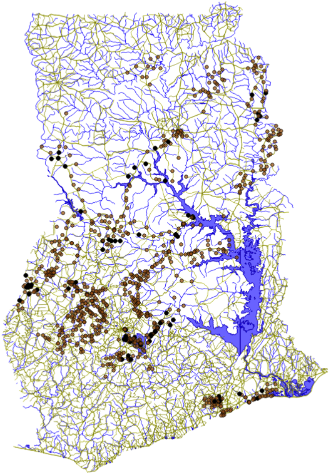
In general: Why do so many new democracies seem so unstable? Why so many coups and protests and civil wars?

One specific idea: Ethnic parties slow consolidation of democratic norms.

How might this dynamic play out in Ghana 2008?

- Party agents are in charge of registering voters (honestly and dishonestly). They mobilize potential voters (for example, in buses). They get paid for fraud (in part).
- Party agents want to register as many people using as few resources as possible (and with as little risk as possible). They know that many voters in Ghana (where the political parties are strongly associated with particular ethnicities):
  - Prefer to have a co-ethnic in office who is more likely to favor them than a non-co-ethnic politician
  - Believe that co-ethnic leaders matters for local public goods
  - Anticipate a close election, citizens may not report registration fraud
- So, agents may target ethnically-homogeneous areas where it's less likely they'll be reported
- *Alternatively:* potential reporting by ordinary citizens may not be a concern, and distances/resources may be a more important factor.

Voter Registration in Ghana 2008



- Registration data from the Electoral Commission of Ghana
- Ethnicity data from the 2000 Population and Housing Census, Ghana Statistical Service.
- Digitize map of feeder roads from the Ministry of Roads and Highways (no Accra in this paper so far)