Multibrand Experiments

Art B. Owen*

Stanford University

Based on joint work with:

Tristan Launay, Google

Opinions are my own, and not those of Stanford, the NSF, or Google

*This work was paid for by Google, not Stanford.

Tech sector context

Past work at Google on

- Measuring ad effectiveness
- Countering click fraud
- Internet surveying

Based on

Multibrand geographic experiments

O & Launay arXiv:1612.00503

Advertising is hard to measure

“On the near impossibility of measuring the returns to advertising”

Randall Lewis & Justin Rao (2013)

Their example

A ‘highly profitable’ campaign

that would have $R^2 = 5.4 \times 10^{-6}$

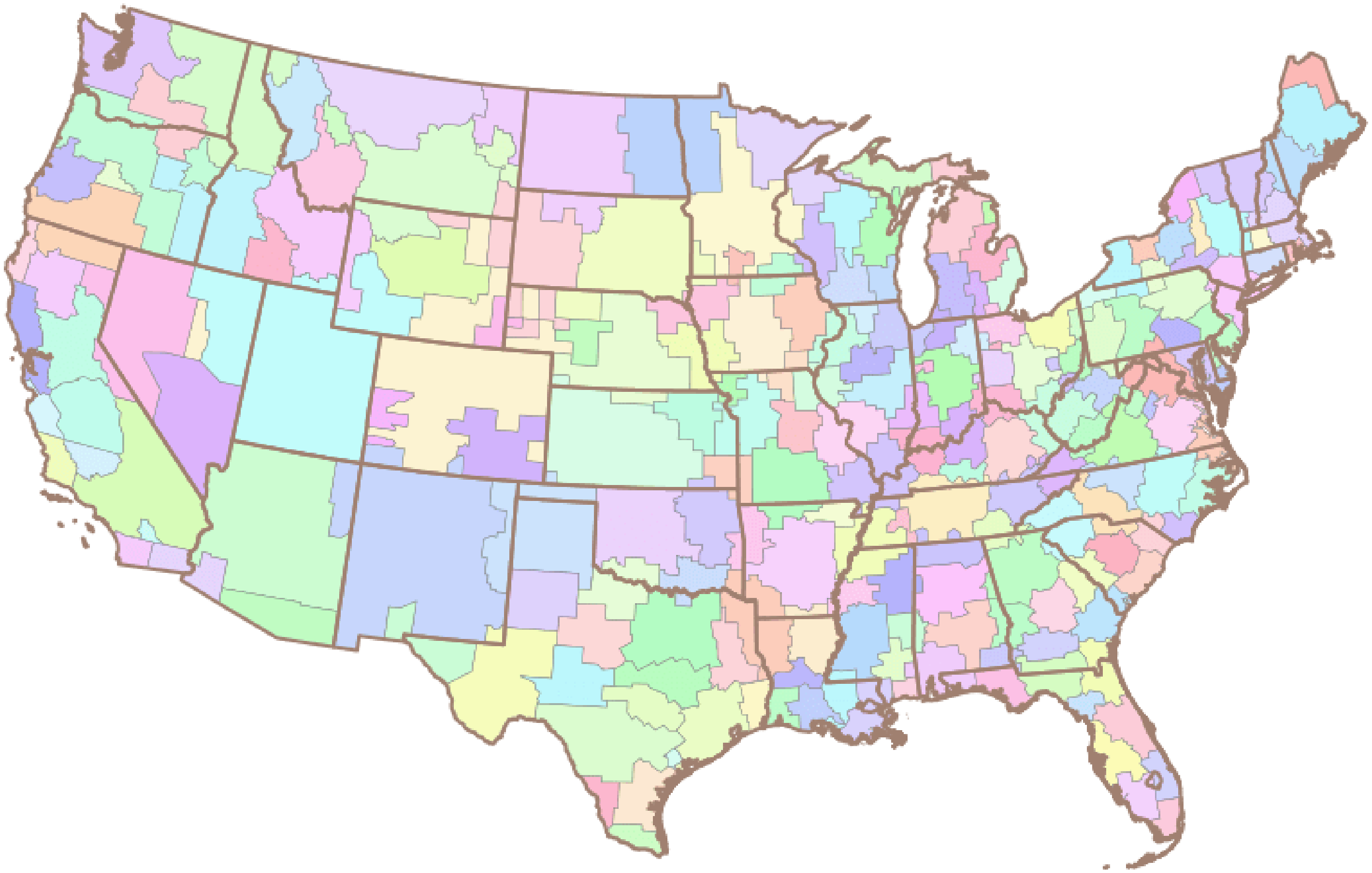
Expect low power

Other views

Maybe not always that hard. But still hard.

GEO experiments

Nielsen designated marketing areas (DMAs) 2013



Source: Wikipedia, CC By. Author 7.11brown

Experiment on DMAs

Acceptably small interference

GEO experiments

Increase spend in some GEOs

Decrease in others

Observe change in revenue

Difficulty

Low power

- $\Rightarrow$ prefer large changes
- $\Rightarrow$ use up inventory
- $\Rightarrow$ auction price could move against you

Bad because

Higher cost

Concern about external validity

Solution

Experiment on B brands in G GEOs

	GEO 1	GEO 2	GEO 3	GEO 4	$\cdots$	GEO G
Brand 1	$+$	$-$	$+$	$-$	$\cdots$	$+$
Brand 2	$-$	$-$	$+$	$+$	$\cdots$	$-$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Brand B	$-$	$+$	$-$	$+$	$\cdots$	$+$

Advantages

1) Smaller expt per brand

   $\Rightarrow$ less adverse price movements

2) Pooled estimate of average effect

3) Bayes / Stein shrinkage

4) Estimate differential GEO effects

5) Easier to be cost neutral

How to sample?

Each brand should be $+$ in $G/2$ GEOs
and $-$ in $G/2$ GEOs

Conversely

Each GEO should be $+$ in $B/2$ Brands
and $-$ in $B/2$ Brands

For cost neutrality

Single brand model

One brand, GEOs g

$$Y_g^{\mathrm{post}} = \alpha_0 + \alpha_1 Y_g^{\mathrm{pre}} + \beta X_g^{\mathrm{post}} + \varepsilon_g^{\mathrm{post}}, \quad g = 1, \dots, G$$

$$\begin{aligned} Y &= \text{KPI}, \quad \text{e.g., revenue} \\ X &= \text{Treatment}, \quad \text{e.g., spending} \\ \varepsilon &= \text{Heteroscedastic error} \\ \mathrm{pre} &= \text{Background period, e.g., 8 weeks} \\ \mathrm{post} &= \text{Test period, e.g., 4 weeks} \end{aligned}$$

Notes

Used earlier by Vaver & Koehler (2011)

Fit well on some test data

minimal autocorrelations

β more interpretable than using logs

Multi-brand model

$$Y_{gb}^{\text{post}} = \alpha_{0b} + \alpha_{1b} Y_{gb}^{\text{pre}} + \beta_b X_{gb}^{\text{post}} + \varepsilon_{gb}^{\text{post}}$$

Overall return

$$\bar{\beta} = \frac{1}{B} \sum_{b=1}^{B} \beta_b$$

GEO specific

$$Y_{gb}^{\text{post}} = \alpha_{0b} + \alpha_{1b} Y_{gb}^{\text{pre}} + \left(\beta_b + \gamma_g\right) X_{gb}^{\text{post}} + \varepsilon_{gb}^{\text{post}}$$

Balanced designs

Take B and G even.

Design $Z \in \{-1, +1\}^{G \times B}$

$$
\begin{array}{c c}
 & \begin{array}{ccccc} \mathrm{GEO}\,1 & \mathrm{GEO}\,2 & \mathrm{GEO}\,3 & \cdots & \mathrm{GEO}\,G \end{array} \\
\begin{array}{c} \mathrm{Brand}\,1 \\ \mathrm{Brand}\,2 \\ \mathrm{Brand}\,3 \\ \vdots \\ \mathrm{Brand}\,B \end{array} &
\begin{bmatrix}
+ & + & - & \cdots & - \\
- & + & + & \cdots & + \\
+ & - & - & \cdots & + \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
+ & + & - & \cdots & +
\end{bmatrix}
\end{array}
$$

Marginal balance

Each row has $G/2$ $+$s

Each col has $B/2$ $+$s

Paired balance

We don't want two brands $+$ in the exact same GEOs

Nor two GEOs $+$ for the exact same brands

Ideally for: $T = +$ & $C = -$

Any two brands have

TT, TC, CT, CC

$G/4$ times each

And two GEOs have

TT, TC, CT, CC

$B/4$ times each

(Impossible)

Pair balance

Design is $Z \in \{-1, +1\}^{B \times G}$

Single balance

B balanced rows

$$\implies Z\mathbf{1}_G = \mathbf{0}_B$$

G balanced cols

$$\implies Z^{\mathsf{T}}\mathbf{1}_B = \mathbf{0}_G$$

Balance pairs of brands

$$\frac{1}{B}Z^{\mathsf{T}}Z = I_G \qquad \text{orthogonal}$$

$$\implies B \leqslant G - 1$$

Cannot also have $G \leqslant B - 1$

We must relax about balance

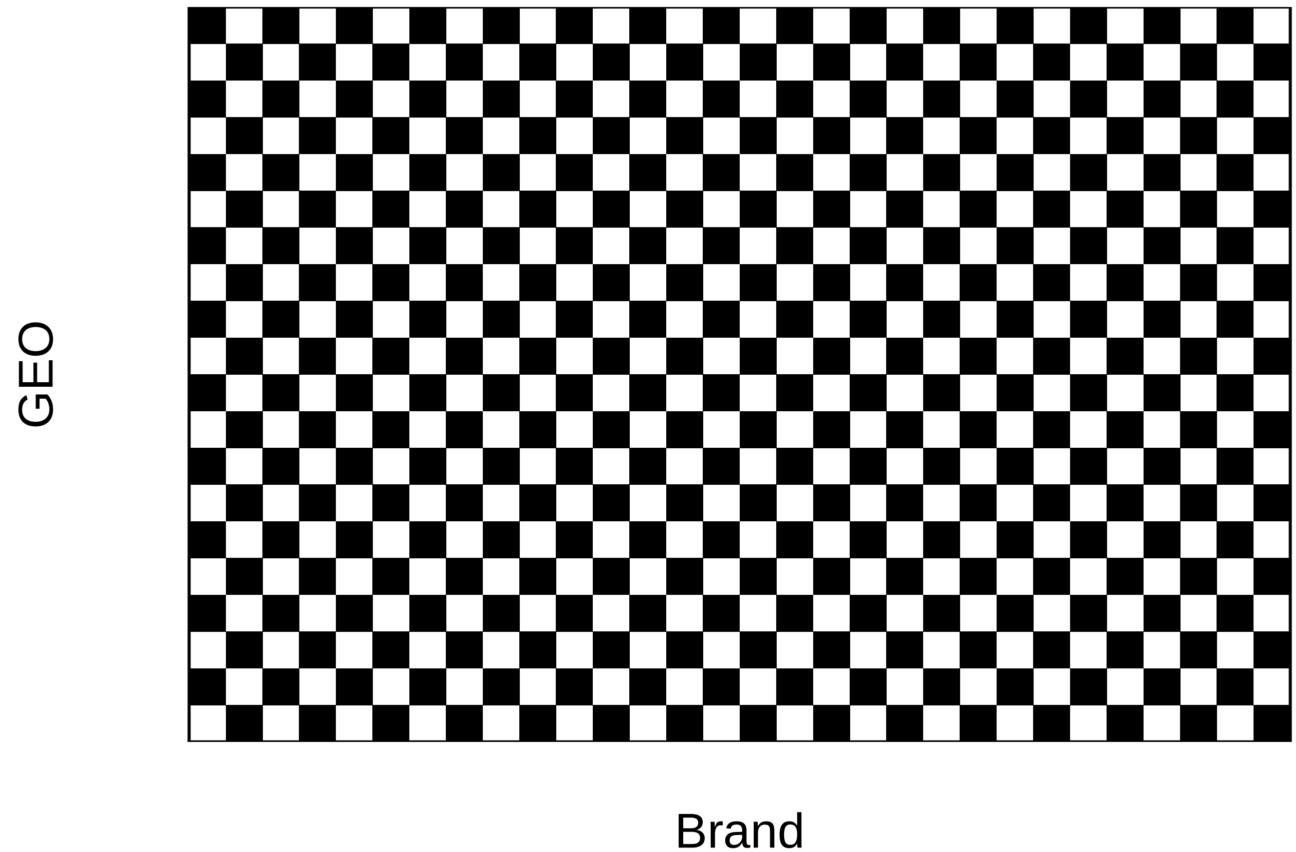
Markov design walk

Based on Diaconis & Gangolli (1995)

Start with a checkerboard and perturb it

Each row/col is half $+1$'s

Multibrand design after 0 steps



Perturbations

Pick two rows and two columns

If you get

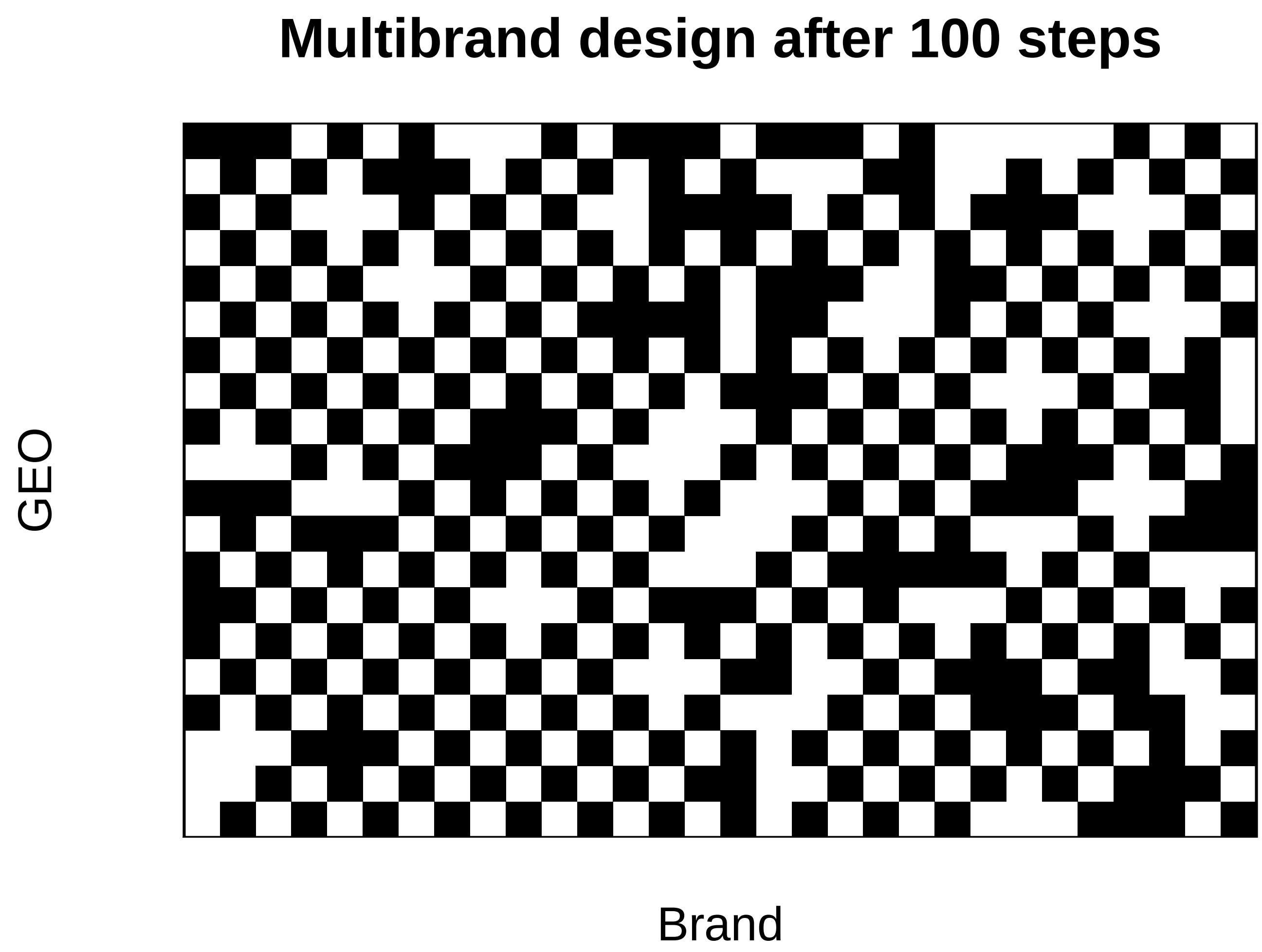
$$\begin{bmatrix} + & - \\ - & + \end{bmatrix} \xleftrightarrow{\text{or}} \begin{bmatrix} - & + \\ + & - \end{bmatrix}$$

Then swap.

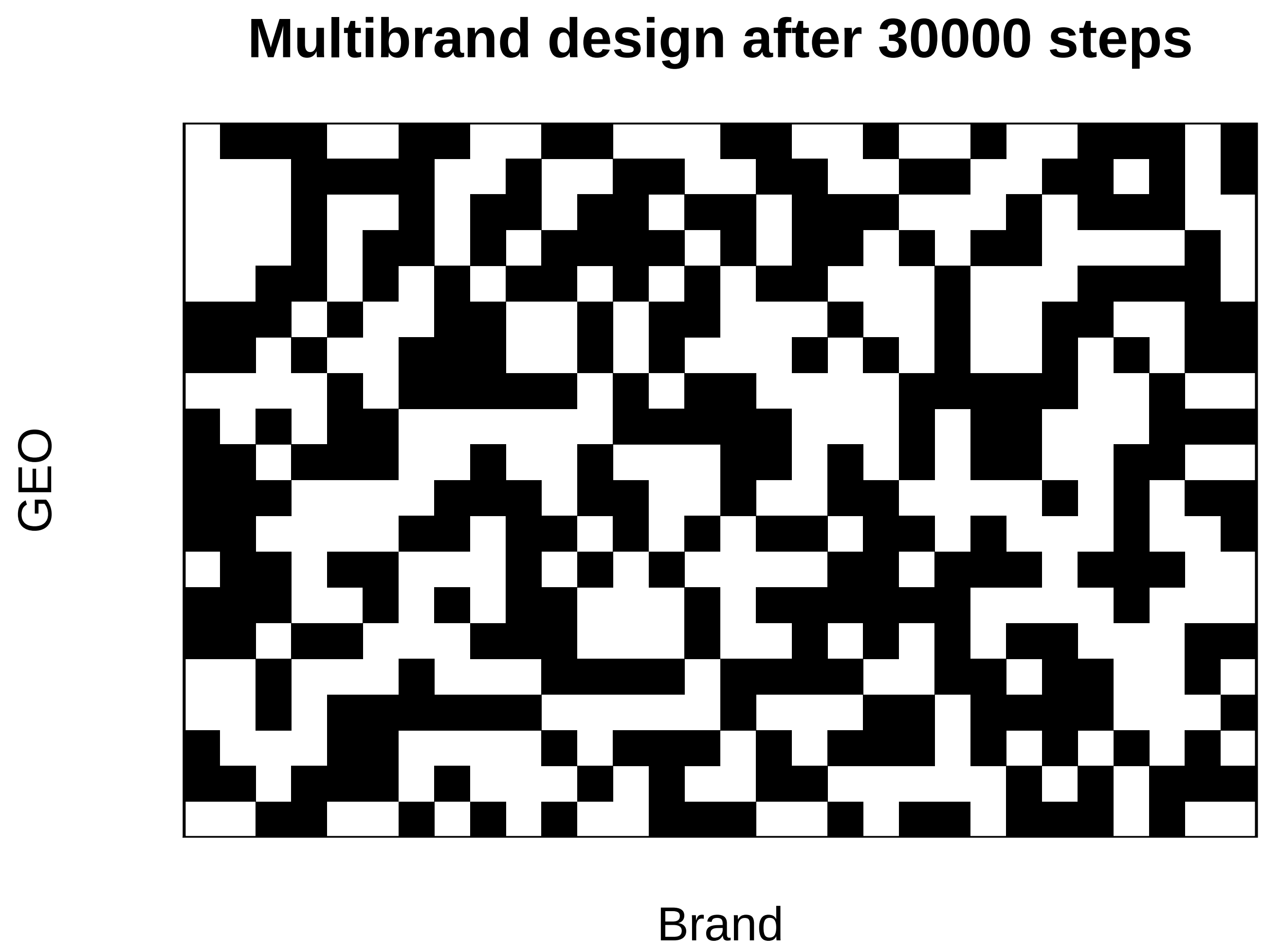
Rows and columns still balanced.

Stationary distribution uniform on matrices with row and column sums of zero

After 100 steps



After 30,000 steps



Convergence

There are 16 possible 2×2 binary matrices

2 are flippable

So $\mathbb{P}(\text{any flip}) \sim 1/8$ (higher at first!)

This flips 4 pixels

Flip about $4 \times 1/8 = 1/2$ pixel per step on average

$30{,}000$ steps $\sim 15{,}000$ pixel flips

$15{,}000/(20 \times 30) = 25$ flips for each pixel on average

Random balance

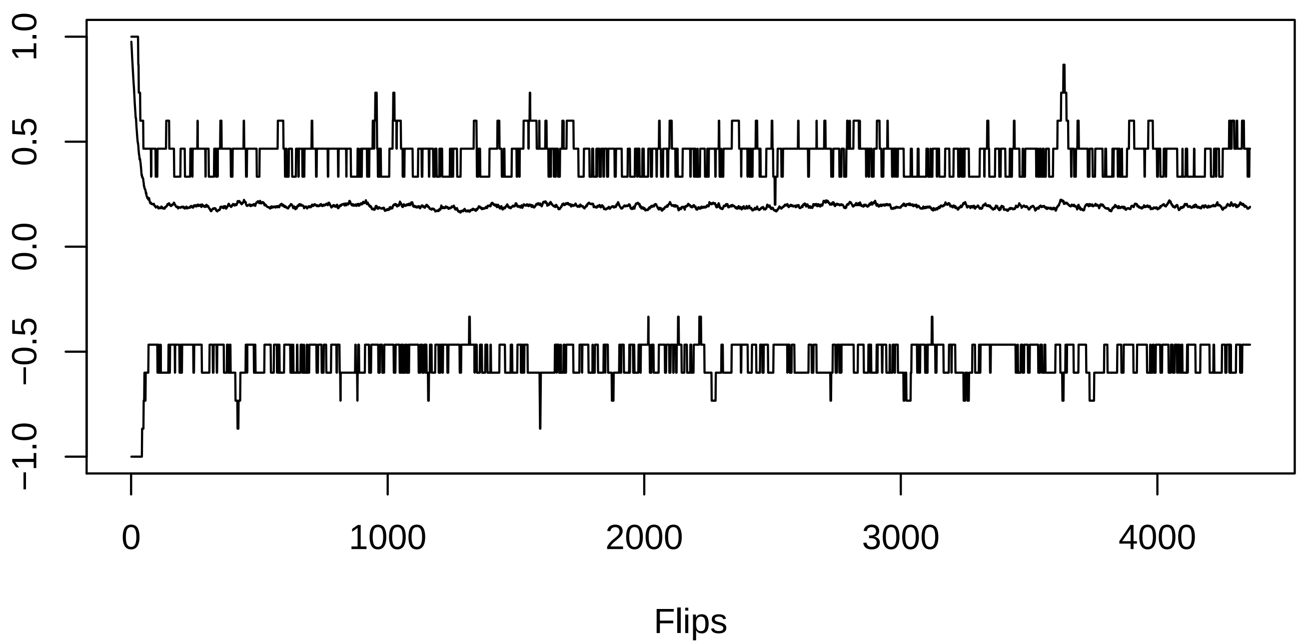
Pairwise correlations:

brands b, b' $\rho_{bb'}$

GEOs g, g' $\rho_{gg'}$

$$\rho_{bb'} = \frac{1}{G}\sum_{g=1}^{G} Z_{bg}Z_{bg'} \qquad \rho_{gg'} = \frac{1}{B}\sum_{b=1}^{B} Z_{bg}Z_{b'g}$$

Min, max and rms GEO correlations



Classic designs

Not so effective

Balanced incomplete block

$$\begin{array}{c} \\ B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \\ B_6 \end{array} \begin{array}{c} \begin{array}{cccc} G_1 & G_2 & G_3 & G_4 \end{array} \\ \begin{bmatrix} + & + & \cdot & \cdot \\ + & \cdot & + & \cdot \\ + & \cdot & \cdot & + \\ \cdot & + & + & \cdot \\ \cdot & + & \cdot & + \\ \cdot & \cdot & + & + \end{bmatrix} \end{array}$$

Exact opposites:

$$B_1 = -B_6 \quad B_2 = -B_5 \quad B_3 = -B_4$$

Submatrices of Hadamard matrices

Also awkward. See O & Launay

Design fun

Can we make $Z \in \{-1,+1\}^{B\times G}$ with

Rows	Cols
half $\pm$	half $\pm$
no two identical	no two identical
no two opposite	no two opposite

Constraints

B, G even

$B \leqslant \binom{G}{G/2}/2$

$G \leqslant \binom{B}{B/2}/2$

Impossible for $\min(B,G) \leqslant 4$ Theorem 2 O & Launay

I.e., $\min(B,G) \geqslant 6$

$$B = G = 6$$

$$\begin{pmatrix} + & + & + & \cdot & \cdot & \cdot \\ + & + & \cdot & + & \cdot & \cdot \\ + & \cdot & \cdot & \cdot & + & + \\ \cdot & + & \cdot & \cdot & + & + \\ \cdot & \cdot & + & + & + & \cdot \\ \cdot & \cdot & + & + & \cdot & + \end{pmatrix}$$

$$B = G = 8$$

$$\begin{pmatrix} + & + & + & + & \cdot & \cdot & \cdot & \cdot \\ + & + & \cdot & \cdot & \cdot & \cdot & + & + \\ + & \cdot & + & \cdot & + & + & \cdot & \cdot \\ + & \cdot & \cdot & + & \cdot & + & + & \cdot \\ \cdot & + & + & + & + & \cdot & \cdot & \cdot \\ \cdot & + & \cdot & \cdot & + & + & \cdot & + \\ \cdot & \cdot & + & \cdot & + & \cdot & + & + \\ \cdot & \cdot & \cdot & + & \cdot & + & + & + \end{pmatrix}$$

New designs

Can also do

$$(B, G) \to (B + 4, G + 4)$$
$$(B, G) \to (B + 8, G + 4)$$

Get some large correlations this way

So we used Markov chain

Does multibrand work?

Of course: Bayes, Stein

How well?

Simulation

Advantages of simulation

- finite sample
- explainable to users
- customizable to details
- can avoid some assumptions

Disadvantage

case by case; less generality

Single brand

Effectiveness $\beta = 5.0$

$G = 20$ GEOs

Heavy tailed GEO sizes

Gamma distributed Y_g^{pre}

Treatment $X = 1\%$ of Y_g^{pre} [repeated with 0.5%]

Gamma distributed noise

Weighted least squares

Results

1) Low power

   30% at $X = 0.5$% 80% at $X = 1$%

2) Wide confidence intervals

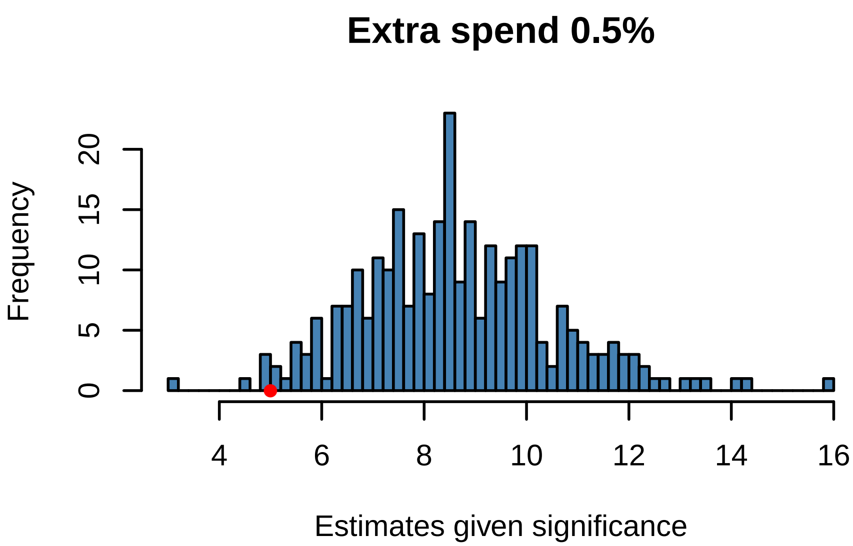
3) Significant winners' curse

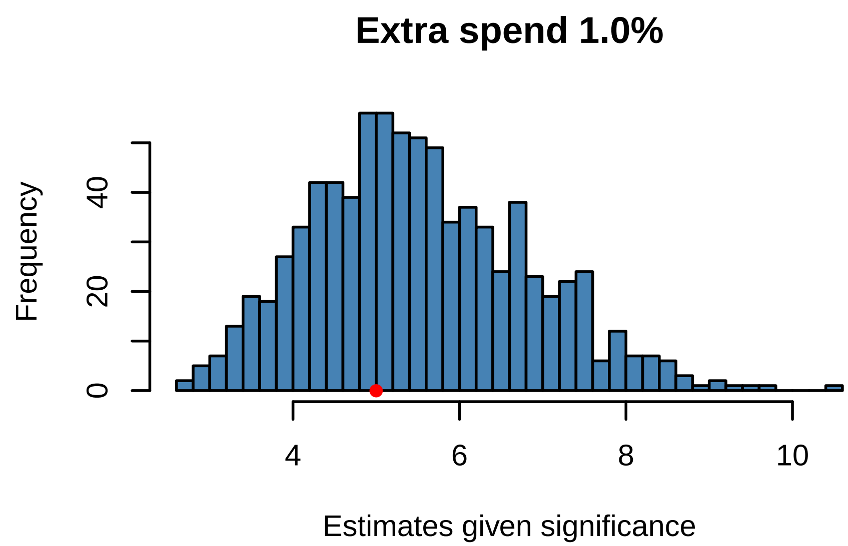
   $\hat{\beta}/\beta$ large given H_0 rejected

Winners' curse

True $\beta = 5$

Statistically significant $\hat{\beta}$ biased





Multibrand

$B = 30$ brands for $G = 20$ GEOs

$\beta_b \sim \mathcal{N}(5, 1)$

Shrinkage

"SURE" of Xie, Kuo, Brown (2012)

$$\tilde{\beta}_b = \frac{\lambda}{\text{Var}(\hat{\beta}_b) + \lambda}\hat{\beta}_b + \frac{\text{Var}(\hat{\beta}_b)}{\text{Var}(\hat{\beta}_b) + \lambda}\hat{\bar{\beta}}$$

For individual estimates $\hat{\beta}_b$

Pick λ

"SURE" criterion

minimize an unbiased estimate of $\sum_{b=1}^{B}(\tilde{\beta}_b - \beta_b)^2$

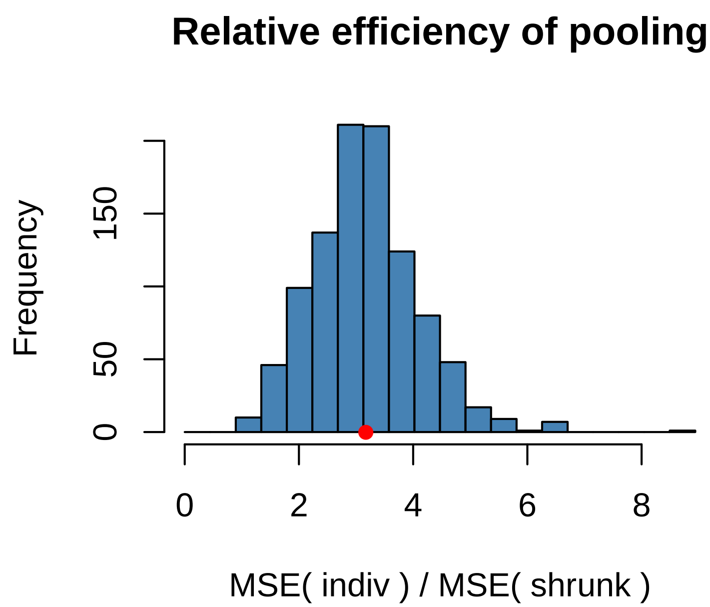
Plug in

Use $\widehat{\text{Var}}(\hat{\beta}_b)$

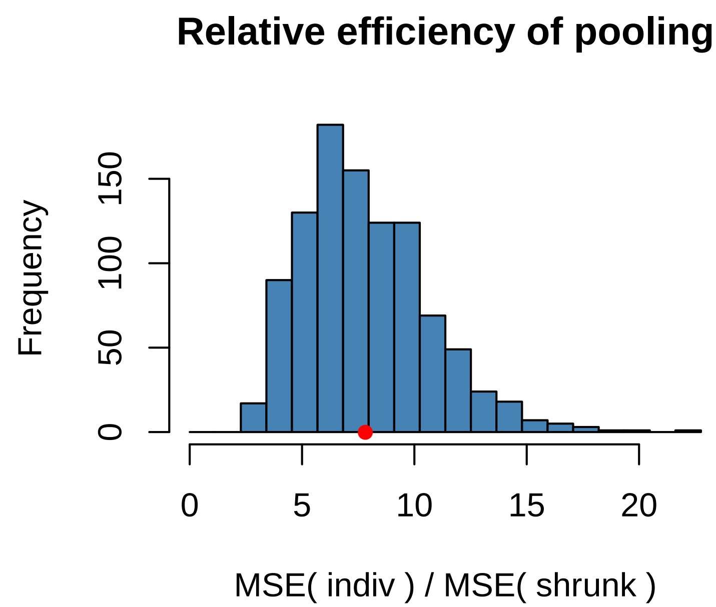
Efficiency

$$\frac{\text{Single MSE}}{\text{Pooled MSE}} = \frac{\frac{1}{B}\sum_{b=1}^{B}(\hat{\beta}_b - \beta_b)^2}{\frac{1}{B}\sum_{b=1}^{B}(\tilde{\beta}_b - \beta_b)^2}$$

1000 simulations



(a) Experimental spend 1%.



(b) Experimental spend 0.5%.

Fully Bayesian

Used STAN

Similar outcomes

Other tech sector things

Tie-breakers

Reward all top customers

no bottom customers

randomize in between

Better accuracy vs regression discontinuity

Joint with Hal Varian Google

Dan Kluger, Harrison Li, Tim Morrison, Michael Baiocchi, Minh Nguyen Stanford

Pooled data

Small excellent sample $\longrightarrow \hat{\beta}_S$

Enormous other sample $\longrightarrow \hat{\beta}_L$

Shrink $\hat{\beta}_S$ towards $\hat{\beta}_L$

Joint with Aiyou Chen, Min Liu Google

Thanks

- Tristan Launay, co-author
- Jon Vaver, Jim Koehler, David Chan, Qingyuan Zhao, comments
- Google for allowing this to be shared
- Ramesh Johari, Stefan Wager, Guido Imbens
- Elizabeth Wilsey
- Stanford Causal Science Center
- Stanford Data Science