CausalImpact: Measuring the impact of market interventions

Kay H. Brodersen, Fabian Gallusser, Nicolas Remy, Jim Koehler, and Steven L. Scott

Google

May 20, 2015

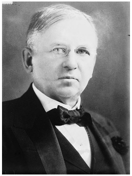
Overview

Summary: We use Bayesian time series regression to model the counterfactual when measuring the impact of market interventions.

Measuring advertising effectiveness is a tricky business

I know that half my advertising dollars are wasted.
I just don't know which half.

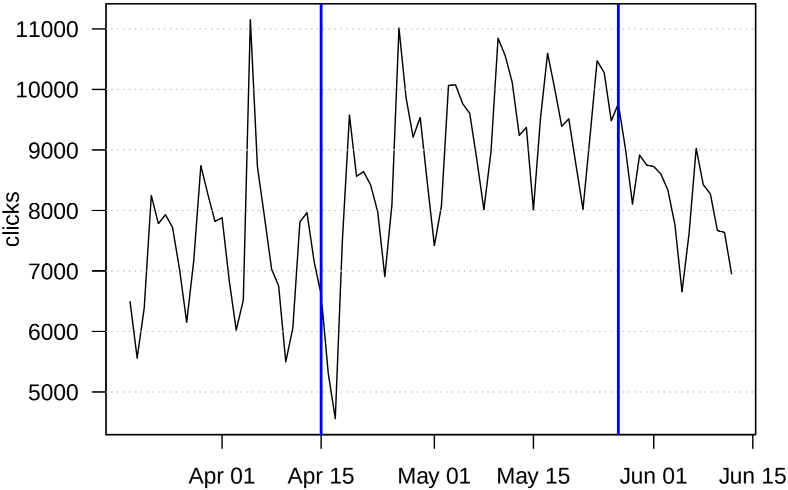
John Wanamaker



- One of the basic promises of online advertising is measurement.
- It is supposed to be easy.
  - Change something (e.g. increase bid on Google).
  - Look to see how many incremental ad clicks you get.
- You'd like to know what would have happened if you hadn't advertised.
  (Lots of potential confounders)

Example

Real Google advertiser. 6-week ad campaign. Random shift added to both axes.



Problem statement

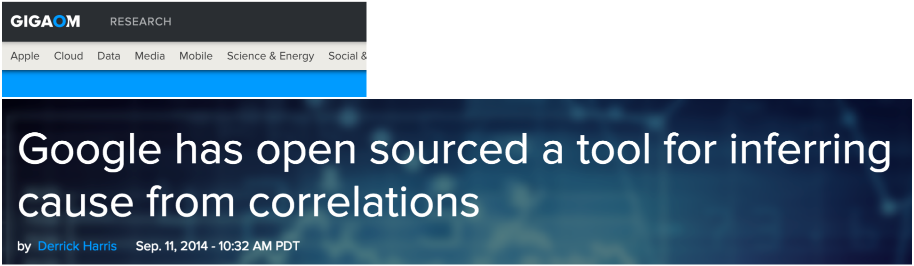
- An actor engages in a market intervention.
  - Has a sale.
  - Begins (or modifies) an advertising campaign.
  - Introduces (or adopts) a new product.
- Other similar actors don't engage in the intervention.
  - This is an important limitation!
  - Can't use this technique to gague the effect of Christmas sales.
- We have data on both the actor and the similar actors prior to the intervention.

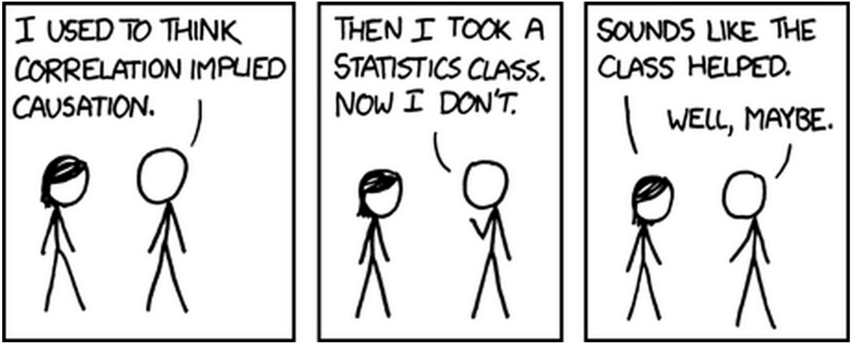
- Question: What was the effect of the intervention?
  - Total change to the bottom line.
  - How quickly did changes begin to occur?
  - How quickly did the effect begin to die out?

The "CausalImpact" model for counterfactual imputation

- Use data in the pre-treatement period to build a flexible time series model for the series of interest.
- Forecast the time series over the intervention period given data from the pre-treatment period.
    - Can use contemporaneous regressors in the forecast.
    - Model fit is based on pre-treatment data.
    - Deviations from the forecast are the "treatment effect."
- Generalizes "difference in differences" and "synthetic controls."

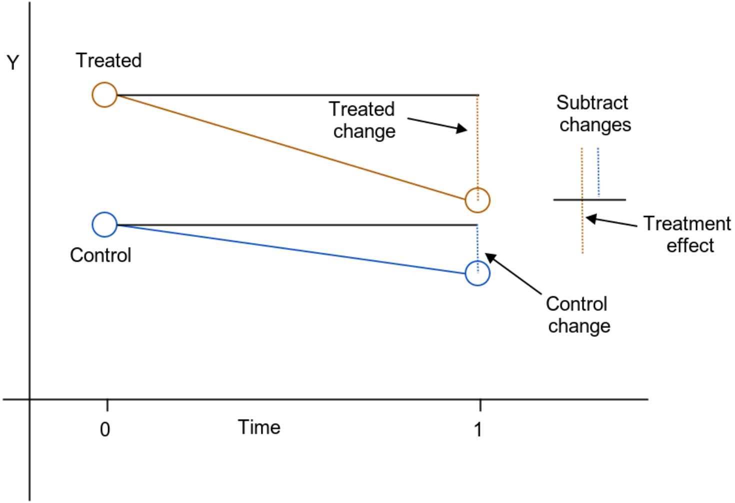
Problem solved!





Difference in differences

An old trick from econometrics. Only measures at two points.

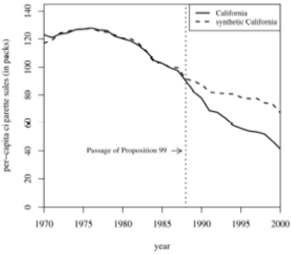


Synthetic controls

A more sophisticated counterfactual model than DnD

Abadie *et al.* (2003, 2010) suggested *synthetic controls* as counterfactuals.

- Weighted averages of untreated actors used to forecast actor of interest.
- Weights ($0 \leq w_i \leq 1$) estimated so that "synthetic control" series matches actor's series in pre-treatment period.
- Difference from forecast is estimated treatment effect.



Good Allows multiple controls, captures temporal effects.

Bad Scaling issues (California vs. Rhode Island), sign constraints (negative correlations?), other time series?
Time series signals ignored (left as "unexplained variance").

Outline

Introduction

Structural time series

Case study

Structural time series models

Can combine time series behavior with contemporaneous predictors

Observation equation

$$y_t = Z_t^T \alpha_t + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, H_t)$$

- y_t is the observed data at time t.
- Z_t and H_t are structural parameters (partly known).
- α_t is a vector of latent variables called the "state".

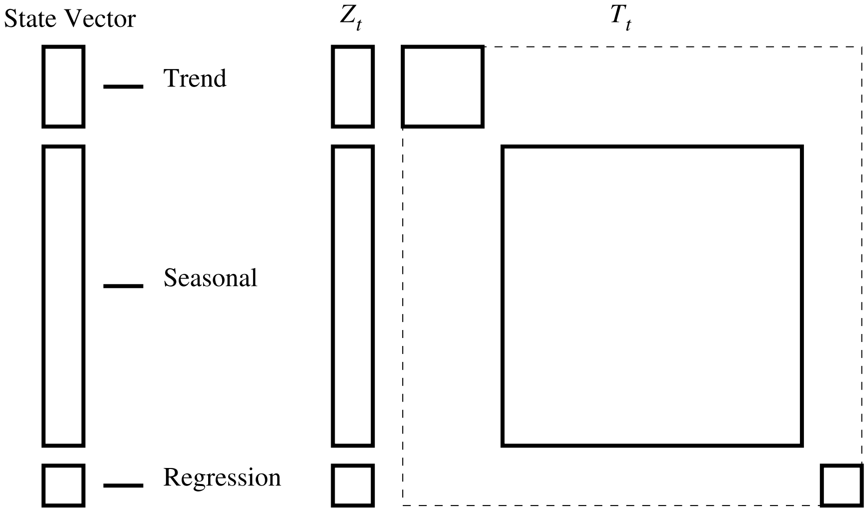
Transition equation

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \qquad \eta_t \sim \mathcal{N}(0, Q_t)$$

- T_t, R_t, and Q_t are structural parameters (partly known).
- η_t may be of lower dimension that α_t.

Structural time series models are modular

Add your favorite trend, seasonal, regression, holiday, etc. models to the mix



A good default model

The model with S seasons can be written

$$
\begin{aligned}
y_t &= \underbrace{\mu_t}_{\text{trend}} + \underbrace{\gamma_t}_{\text{seasonal}} + \underbrace{\beta^T \mathbf{x}_t}_{\text{regression}} + \epsilon_t \\
\mu_t &= \mu_{t-1} + \delta_{t-1} + u_t \\
\delta_t &= \delta_{t-1} + v_t \\
\gamma_t &= -\sum_{s=1}^{S-1} \gamma_{t-s} + w_t
\end{aligned}
$$

This is the "basic structural model" with an added regression effect.

- Trend: "level" μ_t + "slope" δ_t.
- Seasonal: $S - 1$ dummy variables with time varying coefficients. Sums to zero in expectation.
- Regression: Spike and slab prior to handle sparsity.

The default model written in bsts code

```
y <- my.data$ResponseVariable

ss <- AddLocalLinearTrend(
    list(),    ## No previous state specification.
    y)         ## Peek at the data for scaling.

ss <- AddSeasonal(
    ss,            ## Adding state to ss.
    y,             ## Peek at the data for scaling.
    nseasons = 7)  ## 7 "seasons" for day of week effect

model <- bsts(y ~ .,           ## regression formula like 'lm'
              state.specification = ss,  ## time series spec
              niter = 1000,              ## MCMC iterations
              data = my.data,
              expected.model.size = 1)   ## spike-slab
```

MCMC

- The model parameters are $\theta = \{\sigma_\epsilon, \sigma_u, \sigma_v, \sigma_w, \beta\}$.
- The state is $\alpha = \{\alpha_1, \ldots, \alpha_n\}$.

- MCMC algorithm:
  - Draw α given $\mathbf{y}$, θ
    - Kalman filter "forward filter - backward sampler" [Carter and Kohn(1994)], [Frühwirth-Schnatter(1995)], [de Jong and Shepard(1995)], [Durbin and Koopman(2002)].
    - Draws α directly
  - Draw θ given α.
    - Given α, then $[\sigma_u], [\sigma_v], [\sigma_w], [\beta, \sigma_\epsilon]$ are conditionally independent.
    - Independent priors on the time series σ's. Boring.
    - "Spike and slab" prior on β handles sparsity when there are many potential controls.

Other potential models

There is a lot of modeler's choice at play here.

- The "default model" is robust, fast, scalable, and (nearly) automatic.
  - Local level vs local linear trend?
  - Seasonality?
- There are many other approaches we could have taken instead.
  - Time varying regression coefficients.
  - "Intervention analysis" (dummy variable for intervention period).
  - Dynamic factor models.
  - Other "sparse" priors.
  - Spike and slab on state components.
- All of these would have been reasonable too.

Why we settled on this approach

- Simple to understand, implement, and automate.
- Works with limited data.
- "Pure" in terms of potential outcomes
  - Y_{t0}: outcome at time t under control.
  - Y_{t1}: outcome at time t under treatment.
  - Model is based on Y_{t0}'s, and not "polluted" with Y_{t1}'s.

Outline

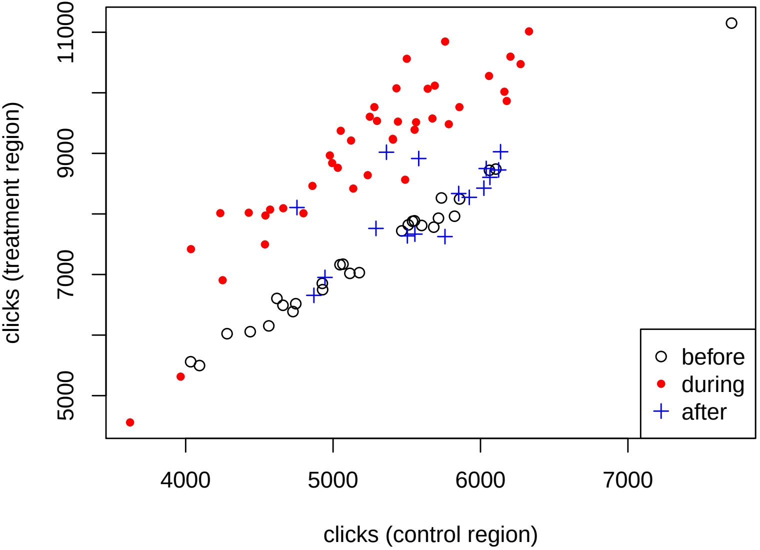
Case study

A Google advertiser ran a marketing experiment.

- Google search ads ran 6 weeks.
- Response is total search related visits to the site.
  - Native search clicks.
  - Ad clicks.
- 95 of 190 "designated marketing areas" received the ads. (DMA's are areas that can receive distinct TV ads).

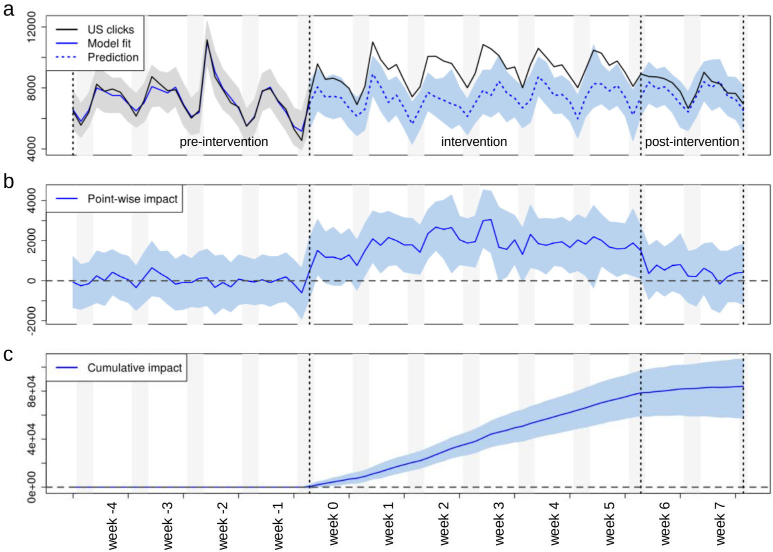
This particular advertiser ran an experiment

Plot shows clicks from treated vs untreated geos. Each dot is a time point.



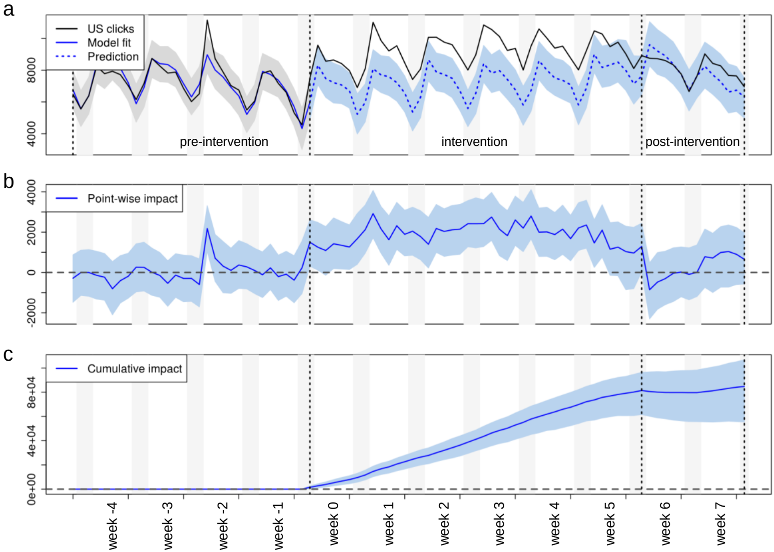
Case study

Google advertiser. Treated vs. Untreated regions



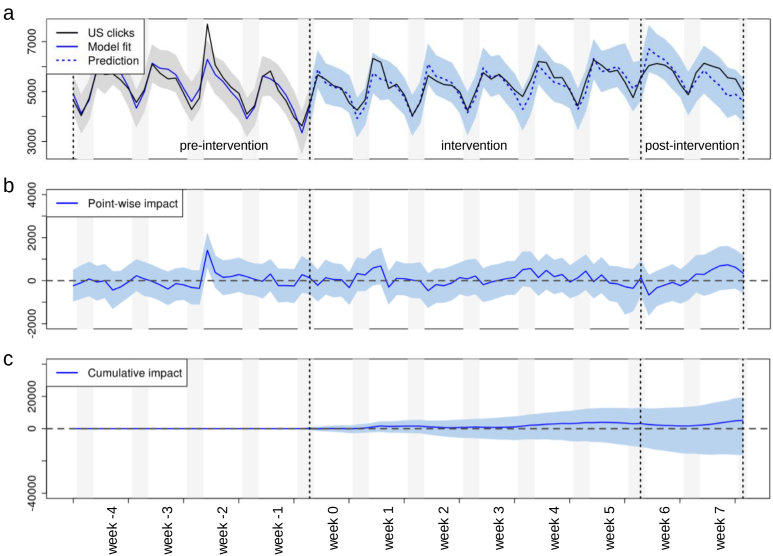
Case study

Google advertiser. Competitor's clicks as predictors



Case study

Google advertiser. Untreated regions. Competitor's sales as predictors



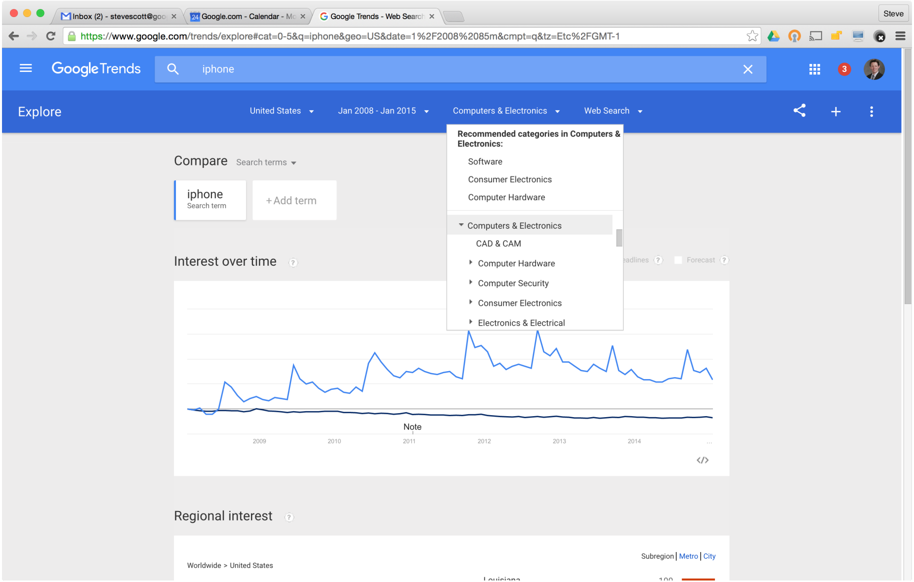
Case study

Summary

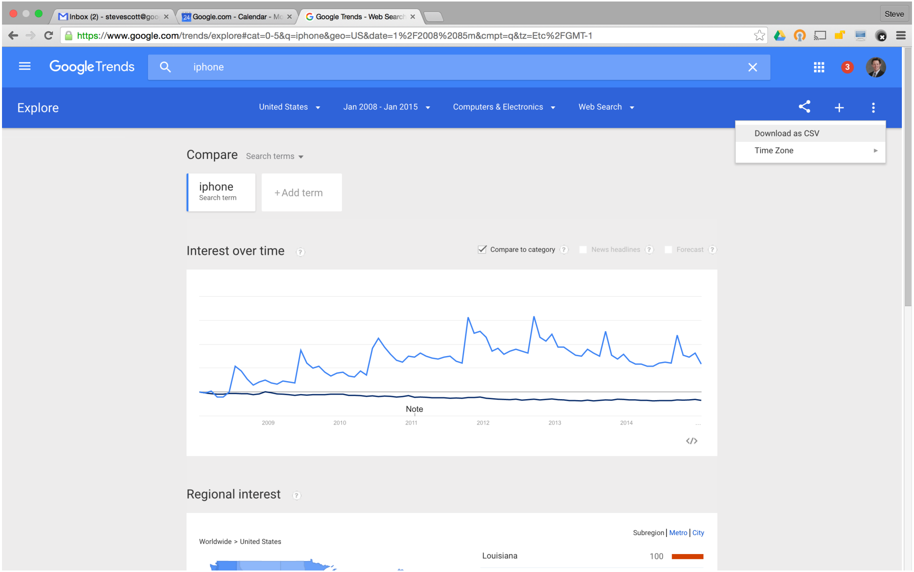
	Clicks	%	95% Interval
vs. Untreated (1)	84,100	20	(15, 26)%
vs. Competitors (2)	84,800	21	(13, 26)%
A-A (placebo) test	8,000	2	(-5, 6)%

- Need experimental data to do analysis 1.
- Analysis 2 is observational, but replicates the experimenatal results.
- Using Google trends (instead of competitor information) gets about the same results.
  - Google trends are publicly available, while competitor clicks are not.
  - Many more potential controls for Google trends. Spike and slab variable selection / model averaging is useful for selecting appropriate control groups.

What if you don't have competitor information?



Google trends “categories” are good industry proxies.



Conclusion

Nice features of CausalImpact:

- Handy way of measuring the impact of market interventions.
- Gives "shape" as well as magnitude (bought with wider SE's).
- Works with arbitrary predictor time series (Google trends!)

Limitations:

- It would be nice to have a diagnostic of when it doesn't work. (I.e. regime change in the X's).
- Like any causal model, you still need exogenous variation to measure causal effects.

R packages:

- CausalImpact
- bsts

References

Abadie, A and Gardeazabal, J. (2010)
The economic costs of conflict: A case study of the Basque Country.
American Economic Review. **93**, 113–132.

Abadie, A. Diamond, A. and Hainmueller, J. (2010).
Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program.
Journal of the American Statistical Association **105**, 493–505.

Carter, C. K. and Kohn, R. (1994).
On Gibbs sampling for state space models.
Biometrika **81**, 541–553.

de Jong, P. and Shepard, N. (1995).
The simulation smoother for time series models.
Biometrika **82**, 339–350.

Durbin, J. and Koopman, S. J. (2002).
A simple and efficient simulation smoother for state space time series analysis.
Biometrika **89**, 603–616.

Frühwirth-Schnatter, S. (1995).
Bayesian model discrimination and Bayes factors for linear Gaussian state space models.
Journal of the Royal Statistical Society, Series B: Methodological **57**, 237–246.

George, E. and McCulloch R. (1997).
Approaches for Bayesian Variable Selection.
Statistica Sinica **7**, 339–374.