Measuring online ad effectiveness

Kay H. Brodersen
Hal R. Varian
Google,Inc.

May 31, 2017

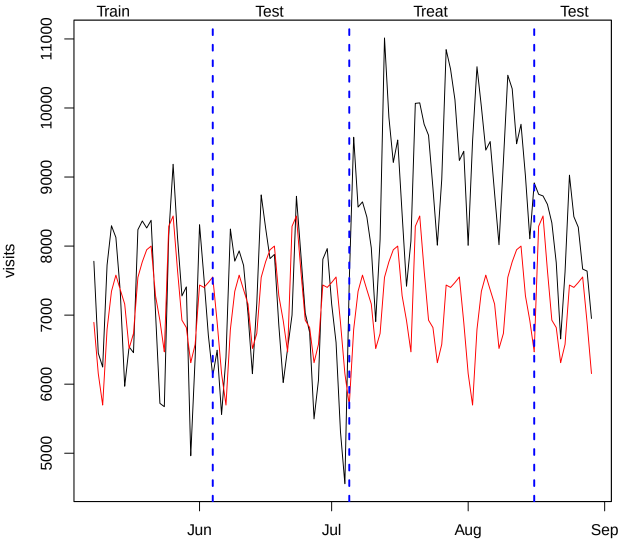
Estimating online ad effectiveness

1. Apply treatment: change ad spend, bid, budget, creative, etc.
2. Compare to counterfactual: what would have happened without experiment?
3. Counterfactual cannot be observed, so it must be estimated
4. One way to estimate the counterfactual is to use a control group randomly chosen from the population
    4.1 Treat some subjects, not others
    4.2 Treat in some geos, not others
    4.3 Treat in some times, not others
5. But this is costly since we *think* the treatment is good, but the control group is not treated

How to reduce cost of experiment

- Use multi-armed bandit (or other sequential testing). See Steven L. Scott, "A modern Bayesian look at the multi-armed bandit," Google Research.
- Use a "synthetic control." [Abadie et al, 2010], "Synthetic Control Methods for Comparative Case Studies," *JASA*.
- In our context, a synthetic control is just a predictive model for the counterfactual
- Another motivation: may be interested in impact of experiment on a *single advertiser* as subject
- In such cases it is natural to use time-based experiment

Hypothetical example of train-test-treat-compare



Bayesian Structural Time Series

We will do this in a time series context using BSTS (available from CRAN.) BSTS combines:

- **Kalman filter.** Accounts for seasonality and trend
- **Spike-and-slab regression.** Automated selection of predictors
- **Bayesian model averaging.** Avoids overfitting, accounts for model uncertainty.

Described in Scott-Varian [2013,2014], Brodersen et. al. [2013]. Related to "interrupted regression", "synthetic controls".

Kalman filter

- Two important time series models
    - Random walk: $y_t = y_{t-1} + e_t$, best prediction is y_{t-1}
    - Constant mean: $y_t = \mu_0 + e_t$, best prediction is $\bar{y}$.
- State space model nests these two models:

$$
\begin{aligned}
y_t &= \mu_t + v_t \\
\mu_t &= \mu_{t-1} + w_t
\end{aligned}
$$

-
    - If $\text{var}(w) = 0$, this is random walk
    - If $\text{var}(v) = 0$, this is constant mean
    - Best prediction for $E\mu_t = m_{t-1} + k_t(y_t - m_{t-1})$
    - ...where k_t depends on $\text{var}(w)$ and $\text{var}(v)$
- Basic Structural Model includes level, local trend, seasonal and regression components

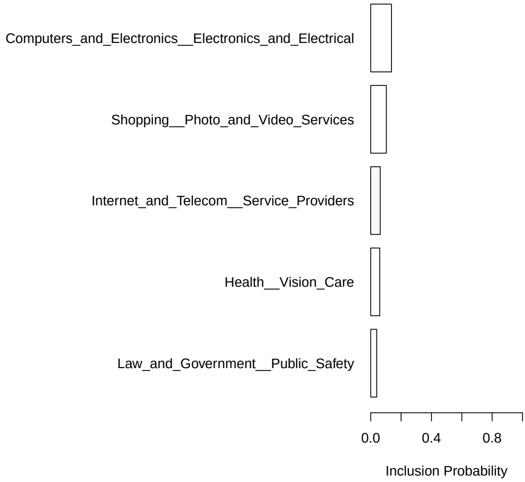
Spike and slab regression

- Let $\gamma = (\gamma_1, \ldots, \gamma_n)$ indicate probability of inclusion
- Conditional on inclusion have prior on coefficient β_i
- Multiply prior times likelihood function and sample from posterior using MCMC
- Draw Kalman parameters, probability of inclusion, coefficient values, predicted value of y_t
- Repeat 5000 times
- Result: Posterior probability of inclusion, distribution of coefficients, posterior distribution of forecast
- Includes model uncertainty, which is necessary due to large number of possible models

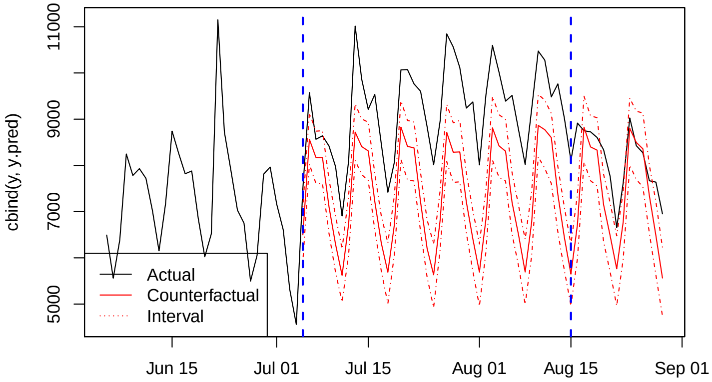
Code from BSTS

```
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
```

Predictors selected by BSTS



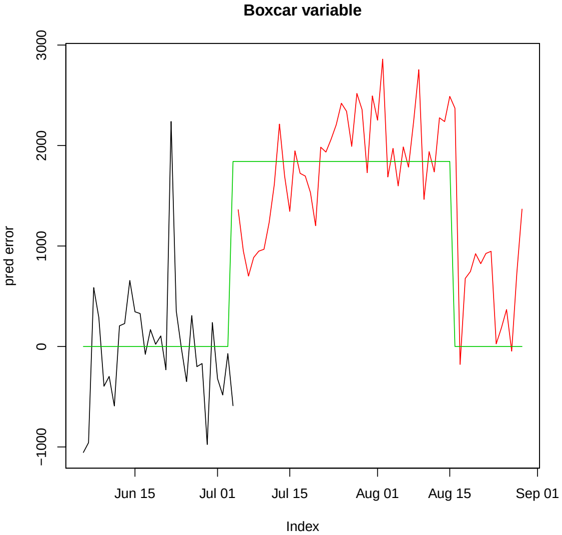
Estimate of treatment effect



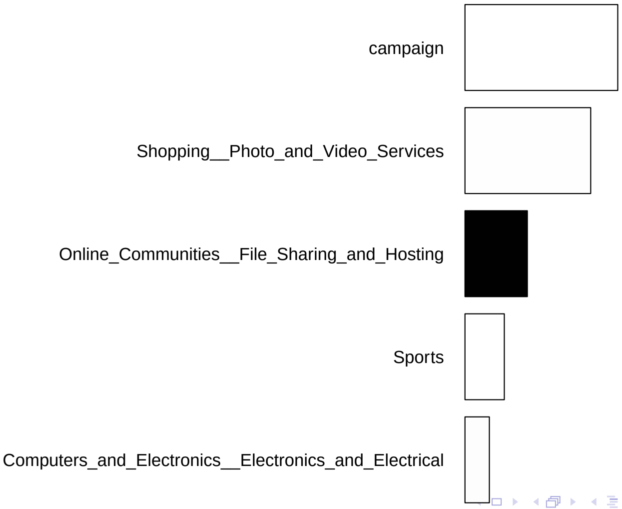
Alternative approaches

1. Use alternative model for impact of ad campaign such as parallel shift?
   - Benefit: Can use all the data to estimate
   - Cost: Restrictive functional form; may miss ramp-up or hysteresis
2. Use alternative estimation technique?
3. Use alternative models for seasonality and trend?

1. Use parallel shift for ad impact



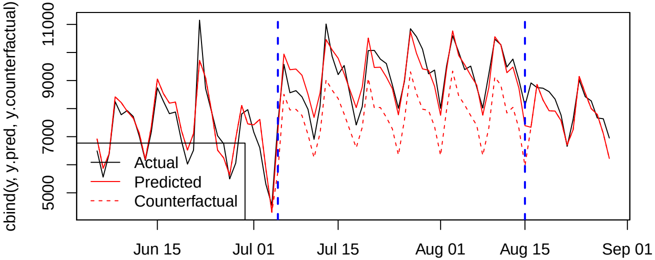
Boxcar indicator variable for campaign



2. Alternative estimation: linear model

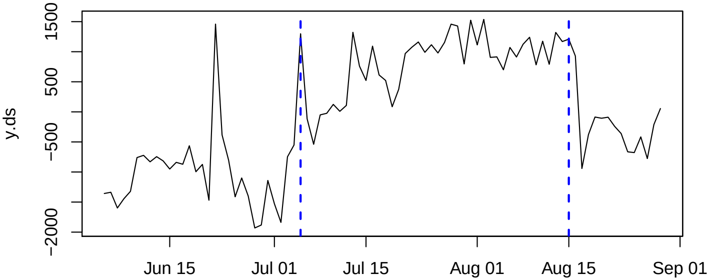
Drop Kalman filter, just use simple linear model

- First spike was a news story about CEO
- July 4 holiday dummy
- Top two categories from Google Trends as regressors



Deseasonalized the data first

Deseasonalize by fitting model with holiday regressor + day-of-week dummies.

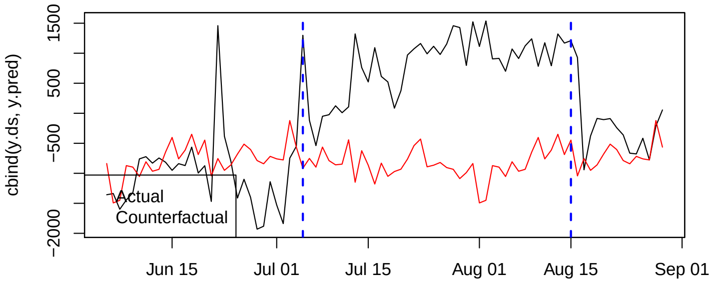


Alternative approaches to seasonality

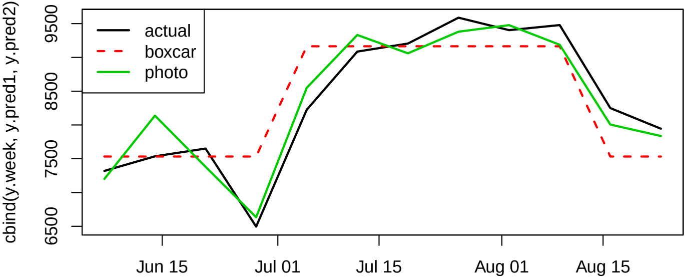
1. Make no adjustment for seasonality (since predictor already has appropriate seasonality)
2. Deseasonalize both predictor and outcome
    - Use boxcar regressor
    - Use extrapolation

3. Alternative seasonality: detrend first

Deseasonalize by fitting model with holiday regressor + day of week dummies.



Use weekly data



Summary

	method	estimate
1	bsts-extrap	1830.43
2	bsts-boxcar	1362.88
3	bsts-boxcar-all-predictors	1279.05
4	bsts-boxcar-top-predictors	1327.06
5	lm-boxcar	1434.57
6	lm-extrap	1289.19
7	not deseasonalized	1393.41
8	deseasonalized-boxcar	1300.67
9	deseasonalized-extrap	1298.37
10	week-boxcar	1248.61

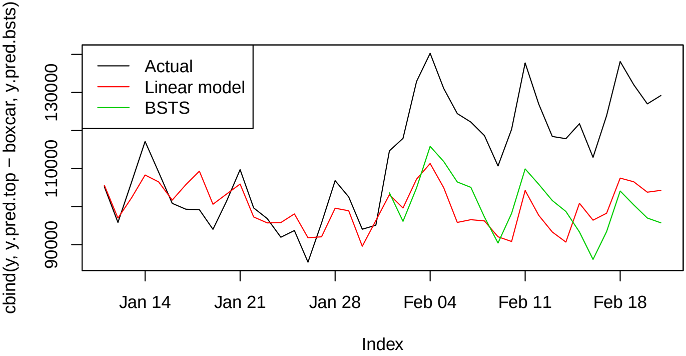
What about revenue?

- Ad clicks may cannibalize search clicks
- May want to look at total number of clicks (i.e., visitors)
- But ad clicks may be worth more or less than search clicks, so really want revenue (or profit)
- Can model ad revenue, search revenue separately or together

Examine a different advertiser . . .

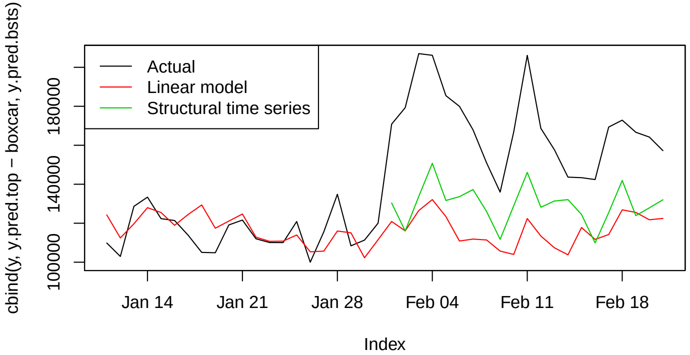
BSTS: Visits actual and counterfactual

Uses the BSTS extrapolation model and a linear regression



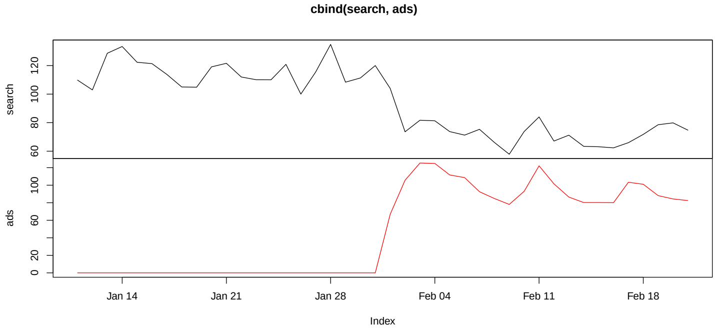
BSTS: Revenue actual and counterfactual

Uses the BSTS extrapolation model and a linear regression



Revenue cannibalization

When campaign begins organic visitors fall but overall ad revenue increases.



Kids you *can* do this at home

Kay Brodersen, Fabien Gallusser, Jim Koehler, Nicholas Remy, Steven Scott, "Inferring causal impact using Bayesian structural time series," *Annals of Applied Statistics*, vol 9, 2015, 247–274.

CRAN: `CausalImpact`