Online Advertising Incrementality Testing

Industry Practical Lessons And Emerging Challenges

Joel Barajas, Narayan Bhamidipati, James G. Shanahan

November, 2021

yahoo!

CIKM 2021 1-5 NOVEMBER

Who we are ...

Joel Barajas
Sr Research Scientist
Yahoo! Research,
Sunnyvale, CA, USA

Linkedin

Narayan Bhamidipati
Sr. Director, Research
Yahoo! Research,
Sunnyvale, CA, USA

Linkedin

James G. Shanahan
Church and Duncan
Group Inc
UC Berkeley, CA, USA

Linkedin

Tutorial Parts

1. The basics: context and challenges
2. Incrementality Testing: concepts, solutions and literature
3. From concept to production: platform building, challenges, case studies
4. Deployment at Scale: test cycle and case studies
5. Emerging trends: identity challenges, industry trends and solutions

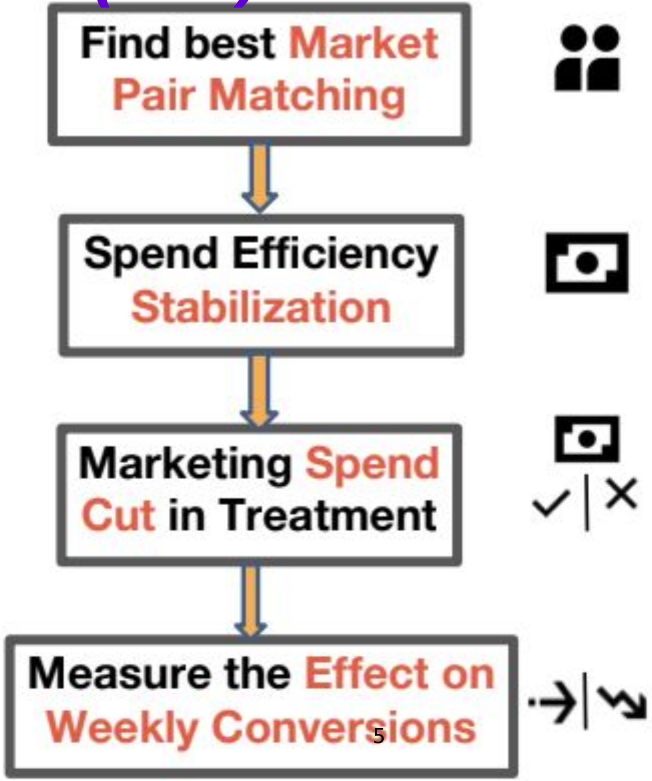
yahoo!

Geo-Testing

Testing with aggregate time series and geo testing units

Controlled Geo-Experiment + Synthetic Control

Barajas et al. (2020), Blake et al. (2015), Abadie et al. (2010)



Without user level holdout, market pair testing is a viable solution

- Typical incrementality testing for advertisers when the ad network **does not support user-level holdouts**

yahoo!

Causal Estimation: Synthetic Control

Barajas et al. (2020)

Bayesian Structural framework with **time series and a regression component** from the control market conversions to predict the treatment conversions (**synthetic control**).

Structural Equation

$$
\begin{aligned}
y_t^{(treat)} &= \overbrace{F_t \theta_t}^{\text{Time Series trend}} + \overbrace{x_t^{(control)T} \beta}^{\text{Control Market Predictor}} + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma^2), \\
\theta_t &= G_t \theta_{t-1} + \omega_t, \qquad \omega_t \sim N(0, W),
\end{aligned}
$$

Market Best Match: A/A tests

Barajas et al. (2020)

Best Pair selection given the conversion and estimation method

Algorithm 1 Control/Treatment Market Pair Selection

Ω: Set of Markets to consider
Φ: Set of placebo intervention times
Δ_t: Time length of historical data
Δ_{t_i}: Time after placebo intervention
for all treatment market: $m \in \Omega$ **do**
  for all control market: $n \in \{\Omega - m\}$ **do**
    for all intervention time: $d \in \Phi$ **do**
      Fit the synthetic control model of Eq 1:
      Find Θ^s, $s = 1, \ldots, N_s$, given $\{y^{(m)}_{d-\Delta_t:d-1}, x^{(n)}_{d-\Delta_t:d-1}\}$
      Predict $\hat{y^s}^{(m)}_{t_i}$, $\forall s \in \{s = 1, \ldots, N_s\}$ after intervention, $\forall t_i \in \{d, \ldots, d + \Delta_{t_i}\}$
      Estimate Credible Intervals (CI) $lift_{cum(d+\Delta_{t_i})}$, Eq 3
    end for
  end for
  $n^*, d^* \leftarrow$ tightest CI that include $lift_{cum(d+\Delta_{t_i})} = 0$
  Append best control/treatment/time $V = \{V, (m, n^*, d^*)\}$
end for
return V

Estimation Parameters: Markets, Intervention times, Train/follow-up Length

A/A test estimation: Given Causal Estimation Framework

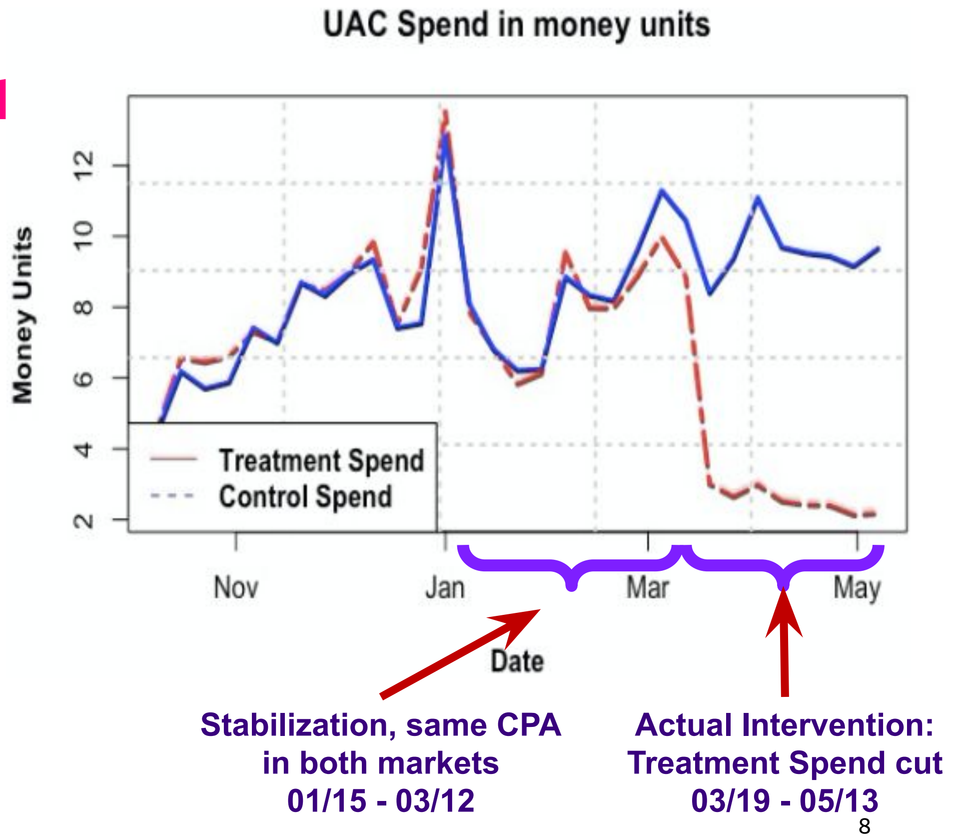
Filter Best Pairs and Parameters: tightest and interval with zero effect

UAC Incrementality: Intervention
Barajas et al. (2020)

Given **treatment/control pairs and the estimation method**, we execute the experiment

1. Cost-per-attributed-signup (CPA) **stabilization** both groups

2. **Suspend** spend for treatment market

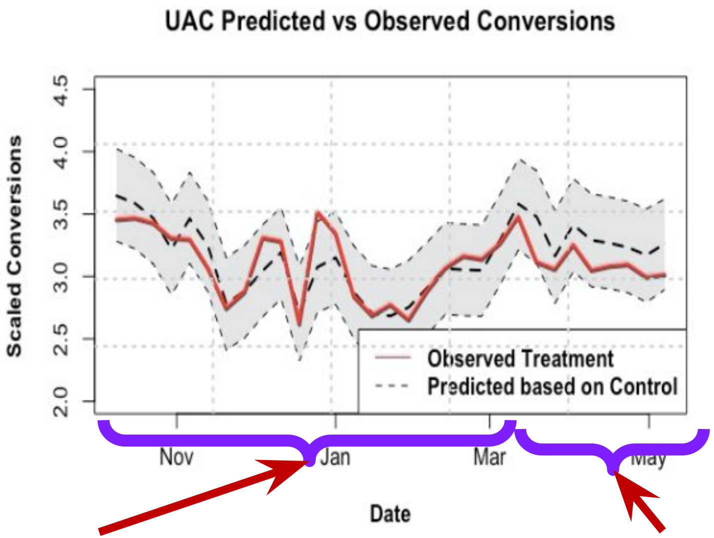


yahoo!

UAC Incrementality: Effect on Weekly Conversions

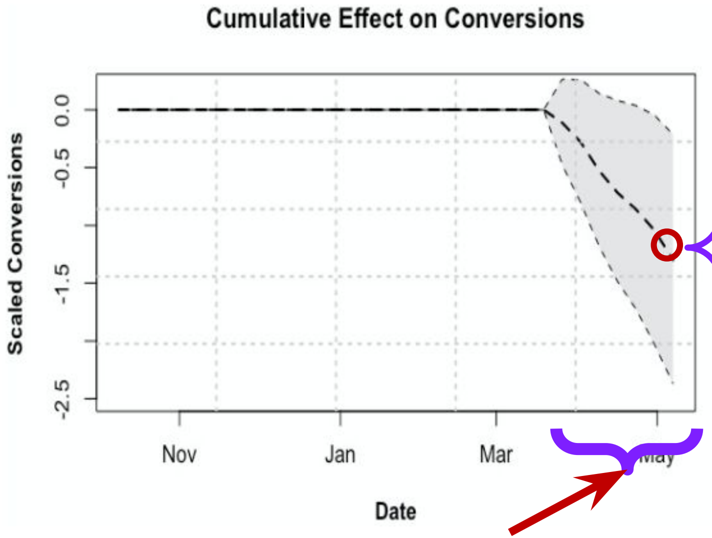
Barajas et al. (2020)

Consistently lower predictive (synthetic control) treatment conversions than observed



Training period:
09/09 - 03/12

Intervention period.
Predictive vs observed
03/12 - 03/15



Cumulative effect and 95% credible intervals

Conversion Lift:
-6.57%

Spend Lift:
-72.33%

Cost per Incremental Scaled conversion:
39.30 money units

yahoo!

Limitations and Caveats

Limitations and Caveats

1. Comparisons between aggregate market conversions **require large intervention effects (spend)** since we are **unable to identify users not exposed to the ads** leading to less precision.

2. Rigorously designed experiments provide **valuable data to build channel cost curves of incremental conversions** and to calibrate Media Mix Models for optimal spend allocation

3. **Testing during holidays is noisy and problematic**, which is a big limitation compared to user holdout testing

Thank you!!!

Feedback welcome.

joel.barajas__AT__yahooinc.com

yahoo!