Knowledge Generation in Business Analytics: Advertising Measurement and Optimization

Günter J. Hitsch

2019 Econometric Society Summer School in Dynamic Structural Econometrics

July 2019

This Lecture: Mostly Harmless Structural Econometrics

Two questions that motivate this lecture

Does marketing (or business analytics in general) differ from IO or applied microeconomics in general, and if so, how?

What is the role of academia in a world where companies employ economists, marketers, computer scientists, statisticians, psychologists, ... ?

Themes

1. Normative versus positive focus: How should a firm make optimal decisions?

2. Creation of generalizable knowledge that informs and aids business decisions

Illustration by example: (Traditional) Advertising

1. Jean-Pierre Dubé, Günter J. Hitsch, and Puneet Manchanda: "An Empirical Model of Advertising Dynamics" (2005), *Quantitative Marketing and Economics*
2. Bradley T. Shapiro, Günter J. Hitsch, and Anna E. Tuchman: "Generalizable and Robust TV Advertising Effects" (2019), *manuscript* (SSRN)

1. Normative Research Focus: Advertising Dynamics and Scheduling

Overview

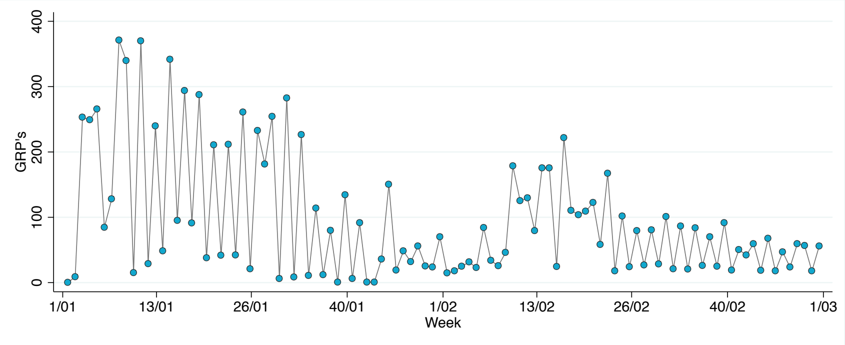
What is the optimal sequence of advertising levels?

- Particular question: How should an advertising budget be scheduled over time?

Approach this problem using a normative, “engineering” approach

- Estimate treatment effect of advertising on demand using an “empirically realistic” demand model
- Numerically solve for advertising policy
  - Require that advertising policy is optimal
  - Require robustness of predictions to strategic interactions (competition)

Example of advertising dynamics (Sprite)



GRP = gross rating point

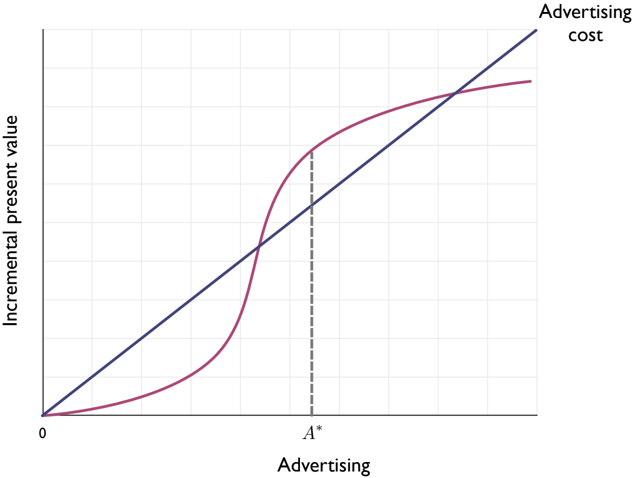
Overview

Pulsing pattern visible in many advertising series

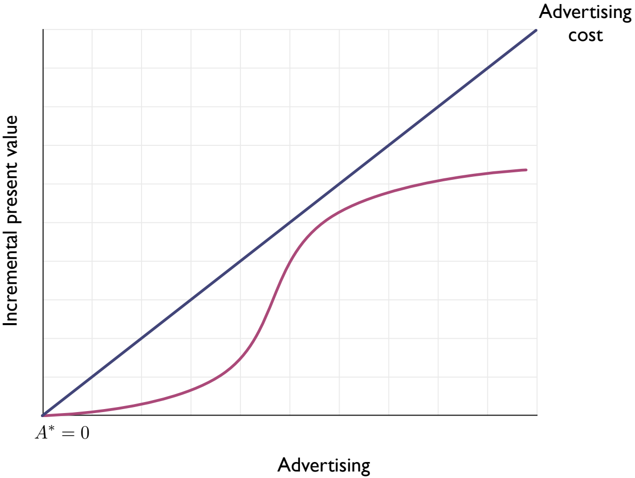
Why? — Two fundamental components of a theoretical justification:

- Advertising effect on demand is persistent
  - Over time, advertising effect depreciates
- Advertising "production function" may exhibit an S-shape
  - Effect on demand small until consumers receive a sufficiently large number of advertising messages

Pulsing: Intuition



Pulsing: Intuition



Roadmap

Let's check if this intuition is correct

- Use tools from IO, marketing, optimization, etc.
- Provide empirical application

Demand: General model structure

Demand for product j:

$$Q_{jt} = f(\boldsymbol{A}_t, \boldsymbol{a}_t, \boldsymbol{p}_t, \boldsymbol{x}_t, \boldsymbol{\epsilon}_t)$$

$\boldsymbol{A}_t = (A_{1t}, \ldots, A_{Jt})$, a vector of ad stock (goodwill) levels, and $\boldsymbol{a}_t$ is a vector of advertising flows

$\boldsymbol{p}_t$ is a corresponding vector of prices, $\boldsymbol{x}_t$ is a vector of other covariates, and $\boldsymbol{\epsilon}_t$ is a vector of demand shocks

Key idea:

- Advertising has dynamic effects
- Adstock A_{jt} is a function of current and past advertising that creates some form of intangible "brand capital"

$$A_{jt} = \phi(a_{j,t-1}, a_{j,t-2}, \ldots)$$

Concrete instance of the advertising model

In period t, adstock and the current advertising flow create theaugmented adstock:

$$A'_{jt} = A_{jt} + \psi(a_{jt})$$

$\psi(a)$ is the adstock production function

Evolution of adstock between periods:

$$A_{j,t+1} = \delta A'_{jt} + \nu_{j,t+1}$$

$0 \leq \delta < 1$ is the carryover factor, and $1 - \delta$ measures depreciation

Note:

$$A_{jt} = \sum_{k=1}^{\infty} \delta^k \psi(a_{j,t-k}) + \omega_{jt}$$

Aggregate multinomial logit demand

Augmented adstock affects utility index in a multinomial logit demand model:

$$u_{jt} = \alpha_j - \eta p_{jt} + \gamma \log(1 + A'_{jt}) + \xi_{jt}$$

Optimality and equilibrium

Firms maximize the present value of profits with respect to a sequence of prices and advertising flows

Firm j's strategy:

$$\sigma_j : \boldsymbol{A} \to (a, p)$$

$\sigma = (\sigma_1, \ldots, \sigma_J)$ is a strategy profile

Optimality requires that $\sigma_j^* = (a^*(\boldsymbol{A}_t), p^*(\boldsymbol{A}_t))$ maximizes the right-hand side of the Bellman equation

$$V_j^\sigma(\boldsymbol{A}_t) = \max_{a,p \geq 0} \left\{ \pi_j(a, p, \sigma_{-j}(\boldsymbol{A}_t), \boldsymbol{A}_t) + \beta \mathbb{E}\left[V_j^\sigma(\boldsymbol{A}_{t+1}) \right] \right\}$$

A **Markov perfect equilibrium** requires that all components of the strategy profile $\sigma^* = (\sigma_1^*, \ldots, \sigma_J^*)$ are mutually consistent

Estimation of the adstock production function

Adstock production function $\psi(a)$ is a key object in the empirical analysis

Should be able to generate pulsing without necessarily imposing pulsing as an optimal advertising strategy

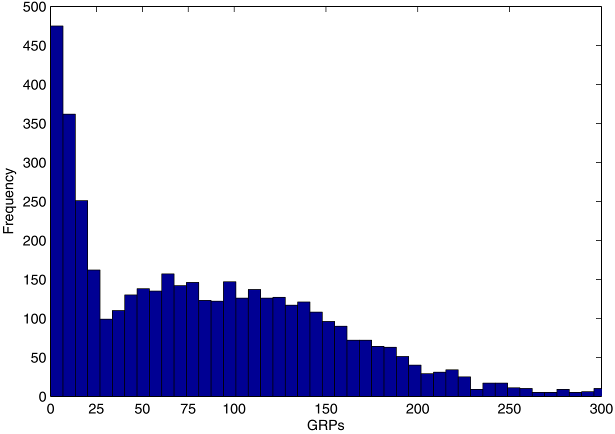
Models:

- Parametric threshold model

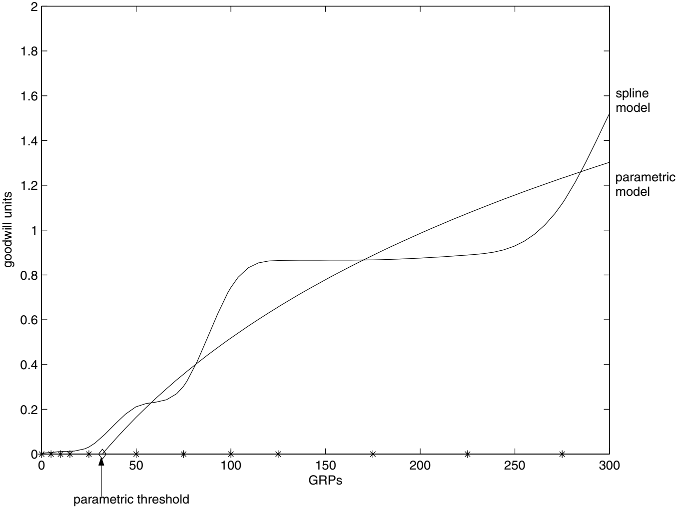
$$\psi(a) = \begin{cases} \log(1 + a - \bar{a}) & \text{if } a \geq \bar{a} \\ 0 & \text{otherwise} \end{cases}$$

- Nonparametric: $\psi(a)$ approximated using a B-spline

Is it possible to estimate the threshold, $\bar{a}$? — Make-goods



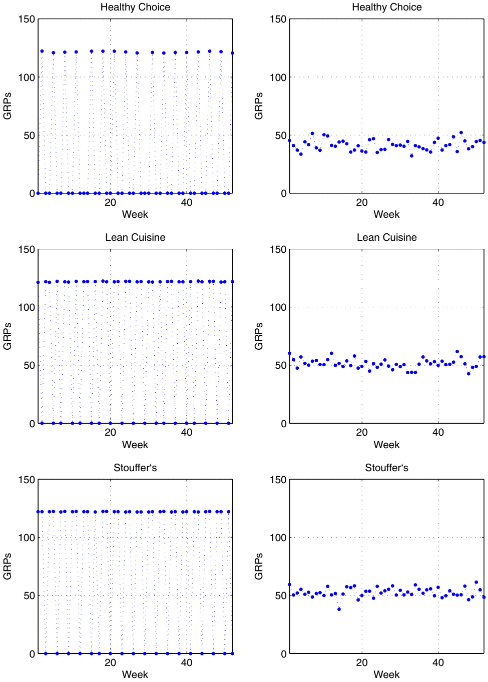
Results: Estimate of advertising production function, $\psi(a)$



Results: Predicted sequence of advertising levels

Recap of steps taken:

- Estimate empirical demand model that, as we *conjecture*, may give rise to pulsing strategies
    - Estimation makes **no assumptions on actual firm conduct**
- Conditional on the demand estimates, solve firm's dynamic advertising problem
- Make prediction robust to competitive interactions


Healthy Choice
GRPs
Week
Healthy Choice
GRPs
Week
Lean Cuisine
GRPs
Week
Lean Cuisine
GRPs
Week
Stouffer's
GRPs
Week
Stouffer's
GRPs
Week

Predicted vs. actual advertising patterns

	Data		MPE	
Brand	Average GRPs	% weeks of 0 adv.	Average GRPs	% weeks of 0 adv.
Budget Gourmet	16.5	75.9	70.8	43.3
Healthy Choice	29.2	63.5	35.7	70.0
Lean Cuisine	34.9	57.3	40.0	66.5
Stouffer's	43.3	60.5	63.9	47.0
Swanson	9.8	87.1	54.4	55.4

Note: The table displays observed advertising, and equilibrium (five firm oligopoly) advertising under the baseline demand specification.

Predicted vs. actual profits

		MPE		Zero adv.		Even adv.	
Brand	Data total profits	Total profits	% diff. wrt. actual	Total profits	% diff. wrt. actual	Total Profits	% diff. wrt. MPE
Budget Gourmet	62.47	65.12	4.25	61.39	−1.73	63.89	−1.90
Healthy Choice	41.74	42.37	1.51	41.53	−0.51	40.11	−5.33
Lean Cuisine	47.24	47.73	1.04	46.56	−1.44	44.94	−5.86
Stouffer's	66.54	67.46	1.37	63.88	−4.01	66.21	−1.85
Swanson	55.38	58.12	4.95	55.57	0.35	56.02	−3.61

Note: Total profits are reported in million dollars per year in all 18 markets. "Even" Advertising refers to a policy whereby firms reallocate the total number of predicted GRP's under MPE advertising, and schedule advertising evenly week after week.

Conclusions: Normative prediction of advertising dynamics and scheduling

Empirical framework shows that advertising pulsing strategies can be optimal

Framework is practically applicable

- Show a firm how to optimally schedule advertising

Comparison to positive work in economics (IO)

Economics (IO) approach

- Make assumptions on firm conduct (optimality/equilibrium) when estimating structural parameters
- Counterfactual: Change a parameter/determinant of market structure and resolve for equilibrium

Normative marketing approach

- No assumptions on firm conduct made when estimating parameters (demand system)
- Counterfactual: Optimal/equilibrium outcome versus observed firm conduct

Think of a more general model of firm conduct, where learning how to make optimal decisions is costly

2. Creation of Generalizable Knowledge: Generalizable and Robust TV Advertising Effects

Overview

Estimate TV advertising effects and predict ROI's for sample of 288 brands

- Sample accounts for vast majority of TV advertising among products sold in supermarkets

Assess robustness to alternative estimation/identification strategies

Attempt to obtain generalizable results

Background

What value can academia (marketing analytics and business economics) provide to make business decisions?

One key task:

- Provide generalizable results
- $\Rightarrow$ prior distribution that is useful for decision-making

Generalizable results as a prior

In this example: Provide generalizable and robust TV advertising effect estimates that serve as a **prior distribution on advertising effectiveness**

Use this prior as an input in decision-making

- What ad effects can be expected in general?
- Obtain more precise product-specific effects to increase the value from an ad campaign
- Anti-trust — is advertising pro or anti-competitive?

Vice versa, decision-making with a flawed prior can have detrimental effects

Generalizability and the research process in marketing and IO (economics)

Case study approach

Generalization based on meta-analyses (mostly in marketing)

Publication bias undermines generalizability

Large estimates, estimates that differ statistically from zero, and "surprising" results are systematically more likely to be selected into publication

- Null effects not interesting
- Reviewer 2: "We already know that"
- Reviewer 3: "Why are these results surprising?"

Prospect of rejection affects research process:

- File-drawer problem
- p-hacking

Consequence: Published results may not be generalizable

- A meta-analysis does not solve the problem if based on results that through selection (effect size, statistical significance) survived the publication process

Publication bias begets publication bias!

This research

Designed to avoid publication bias

Clear and verifiable research protocol to choose population of brands

- All estimates are presented irrespective of results

What does the literature tell us about TV ad elasticities?

Some randomly chosen case-studies of specific categories or brand

- Antidepressants = 0.045 (Shapiro 2018)
- E-cigarettes = 0.08 (Tuchman 2018)
- Frozen entrees = 0.11 (Dubé, Hitsch, and Manchanda 2005)

Meta-analyses of published estimates

- Assmus et al. (1984): Mean = 0.22
- Sethuraman et al. (2011): Mean = 0.12

Multi-product studies with randomized advertising treatment

- Lodish et al. (1995) split cable experiments: Mean = 0.13 (0.05 for established products)

Research design

Basic model structure

Linear model, estimated separately for each brand:

$$\log(Q_{st}) = \boldsymbol{\beta}^T \log(1 + \boldsymbol{A}_{d(s)t}) + \boldsymbol{\alpha}^T \log(\boldsymbol{p}_{st}) + \boldsymbol{\eta}^T \boldsymbol{x}_{st} + \epsilon_{st}$$

$$\boldsymbol{A}_{d(s)t} = \sum_{\tau=t-L}^{t} \delta^{t-\tau} \boldsymbol{a}_{d(s)\tau}$$

Observations at store-week level, s, t

- $d(s)$ ... DMA where store s is located
- Q ... total volume sold (in equivalent units)
- $\boldsymbol{a}$ ... advertising (GRP) flows
- $\boldsymbol{A}$ ... ad stock
- $\boldsymbol{p}$ ... prices
- $\boldsymbol{x}$ ... includes feature and display advertising (in some of the specifications)

Identification of causal advertising effects

Concern: Advertising and unmeasured confounders (demand components) are statistically dependent

- Presumption that advertisers can predict unmeasured demand components and schedule (target) advertising accordingly

Main strategies to argue that statistical association is causal

1. Baseline specification: Rich set of store (market), season, and time fixed effects
2. Shapiro (2018) border strategy

Baseline specification

$$\log(Q_{st}) = \boldsymbol{\beta}^T \log(1 + \boldsymbol{A}_{d(s)t}) + \cdots + \gamma_s + \gamma_{\mathcal{S}(t)} + \gamma_{\mathcal{T}(t)} + \epsilon_{st}$$

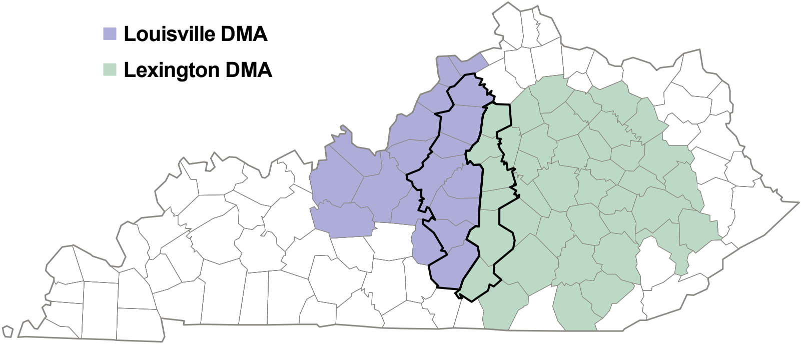
- Includes store, season, and time fixed effects

Identifying assumptions

- Advertising can be easily adapted to markets and modified for seasons
- Advertising bought in advance (several months in the case of network advertising) $\Rightarrow$ makes targeting less predictable, time-varying demand components difficult or impossible
- In particular, advertising scheduling is "coarse" and does not adapt to local, time-varying demand components

If true, residual variation in advertising, conditional on fixed effects, is as good as random

Border strategy



$$\log(Q_{st}) = \boldsymbol{\beta}^T \log(1 + \boldsymbol{A}_{d(s)t}) + \cdots + \gamma_s + \gamma_{\mathcal{S}(t)} + \gamma_{\mathcal{B}(s,t)} + \epsilon_{st}$$

- Includes store, season, and border-specific time fixed effects

Identifying assumptions

- Advertising can possibly target local, high-frequency demand components
- Demand shocks across the border are equal, but advertising is different
    - Costs, advertising slot availability, ...

If true, residual variation in advertising, conditional on fixed effects, is as good as random

Potential issues with border strategy

- Are the border effects representative of the effect in the whole DMA?
- Statistical power (theoretically ambiguous)

Data

Data sources

Demand models estimated at brand level

Nielsen RMS scanner data

- Store/week level
- 2010 - 2014
- 12,671 stores in final sample
- Quantities, prices
- Feature and display advertising in subset of stores (17 percent)

Nielsen Ad-Intel data

- DMA-level advertising occurrences and viewership data (program ratings)
- Set top box in top 25 DMA's (LPM — local people meter markets)
- Diary entries in bottom 185 DMA's
- Media types
  - Cable TV
  - Network TV
  - Syndicated TV
  - Spot TV
- Use occurrences and viewership data to construct GRP's at DMA-week level

Nielsen Homescan household panel data

- More than 60,000 panelists across the U.S.
- Records of products bought on shopping trips
- Prices, quantities
- Projection factors make sample of households nationally representative

Homescan data used to predict national sales volume for ROI calculations

Advertising measure: Gross rating points (GRP's)

1. Occurrence and impression data, market universe estimates $\Rightarrow$ ratings $=$ percentage of TV viewing households in DMA who are exposed to an ad
2. Sum ratings across all brand ad occurrences within a market-week $\Rightarrow$ GRP

Choice of brands in analysis sample

Generalizability requires a clear protocol to choose the brands included in the analysis

- Choose top 500 RMS brands based on revenue
- Drop brand if merge with Ad-Intel data fails
    - There is little or no advertising for such brands

Final sample contains 288 brands

- Largely the universe of brands sold in supermarkets that are major advertisers

Results

Interpretation of advertising effect β

Simplified demand model:

$$\log(Q_t) = \beta \cdot \log(1 + A_t)$$

$$A_t = \sum_{\tau=t-L}^{t} \delta^{t-\tau} a_\tau$$

β approximates the ad stock elasticity if A_t is large

$$\frac{\partial Q_t}{\partial A_t} \frac{A_t}{Q_t} = \beta \frac{A_t}{1 + A_t} \approx \beta$$

To simplify the exposition assume that advertising is constant at $a_t = a$ in all periods t

Long-run effect of advertising

1. Elasticity with respect to a permanent increase in advertising level a

$$\frac{dQ_t}{da}\frac{a}{Q_t} \approx \beta$$

2. Effect on total, long-run demand from increase in advertising in period t

$$\left(\frac{\partial}{\partial a_t}\sum_{\tau=t}^{t+L} Q_\tau\right)\frac{a_t}{Q_t} \approx \beta$$

Short-run effect of advertising

$$\frac{\partial Q_t}{\partial a_t}\frac{a_t}{Q_t} \approx (1-\delta)\cdot\beta$$

Choice of carry-over factor δ

Calibrated at $\delta = 0.9$

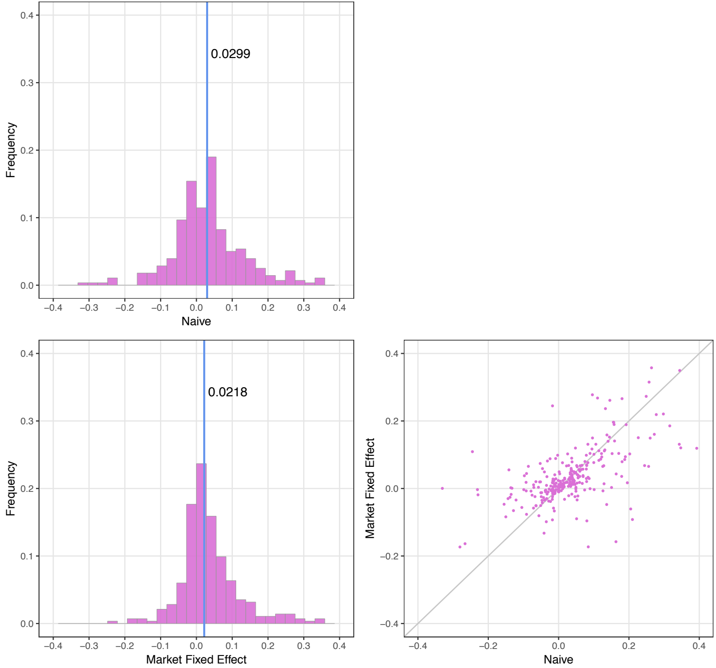
- E.g. Dubé, Hitsch, and Manchanda (2005)

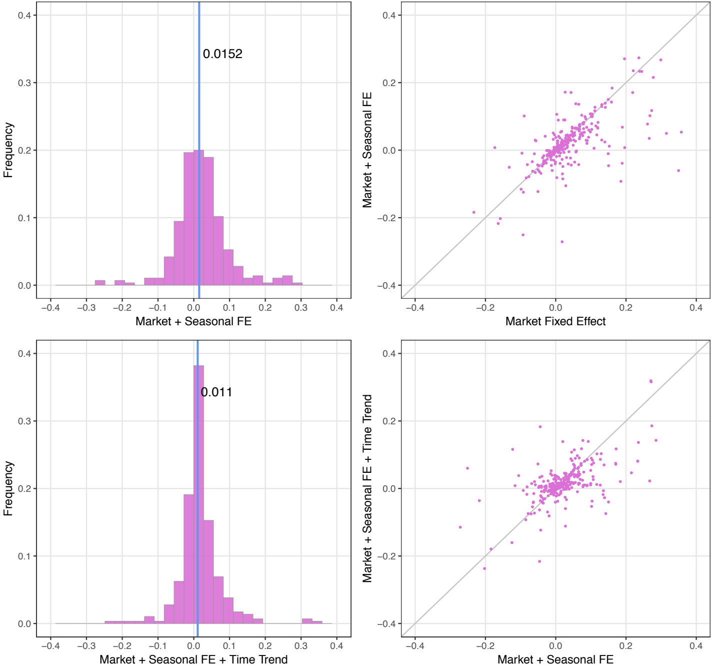
Robustness to this assumption and estimation of δ more generally discussed in the paper

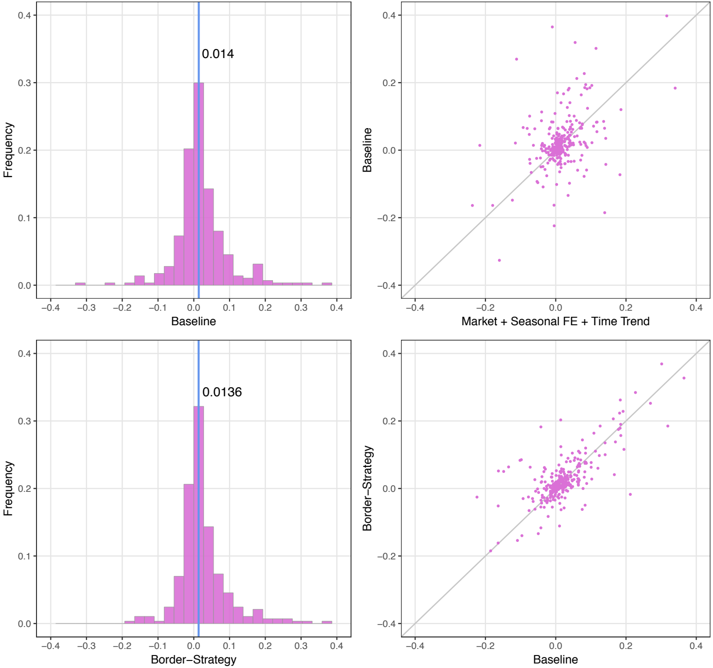
Interactive results

Publicly available

https://advertising-effects.chicagobooth.edu


0.0299
Frequency
Naive
0.0218
Frequency
Market Fixed Effect
Market Fixed Effect
Naive


0.0152
Frequency
Market + Seasonal FE
Market + Seasonal FE
Market Fixed Effect
0.011
Frequency
Market + Seasonal FE + Time Trend
Market + Seasonal FE + Time Trend
Market + Seasonal FE

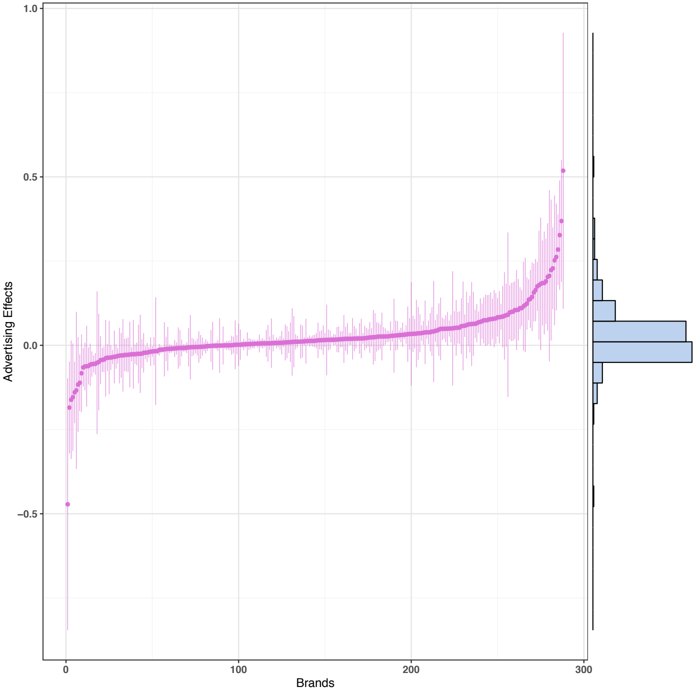

0.014
Frequency
Baseline
Baseline
Market + Seasonal FE + Time Trend
0.0136
Frequency
Border−Strategy
Border−Strategy
Baseline

Comparison of estimates across specifications

Estimates of advertising effect β

	Median	Mean
Naive	0.0299	0.0415
+ store fixed effects	0.0218	0.0467
+ season fixed effects	0.0152	0.0251
+ time trend	0.0110	0.0171
+ time fixed effects	0.0140	0.0233
Border strategy	0.0136	0.0258

Results robust to specifications once we adjust for store (market) and season


1.0
0.5
0.0
−0.5
Advertising Effects
0
100
200
300
Brands

Magnitude of estimates

For reference: Sethuraman et al. (2011) meta analysis shows mean effect of 0.12

We document substantially smaller effects

Furthermore:

- 26.4/24.3 percent of estimates (baseline/border strategy) are positive and statistically different from zero at a 5 percent level
- 66.3/68.4 percent of estimates not statistically different from zero

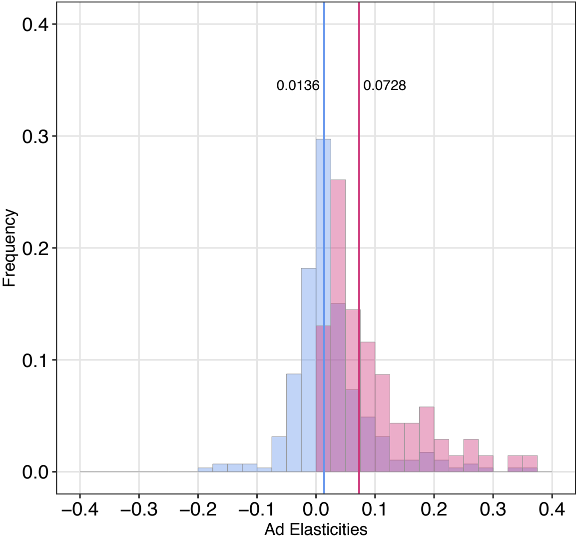
Subpopulation of brands: Power to detect ad effect

Estimates with 50 percent power to detect an effect of $\beta = 0.05$

- Median/mean of 0.0073 and 0.0083
- Similar percentages
  - Positive and statistically significant
  - Not statistically different from zero

Small estimates and incidence of null-effects not just driven by lack of statistical power

Subpopulation of brands: Positive and significant ad effect



Estimates that are positive and "significant," i.e. $p < 0.05$

- 15 percent of all brands
- Median/mean of 0.0728 and 0.1015

Publication bias?

- Suppose a *hypothetical* publication process selected brands (case studies) from our population based on positive and highly significant estimates
- Resulting advertising effects will then be comparable to estimates in extant literature

Economic significance — Return on investment (ROI)

Predicting ROI's

1. Choose a base week t
2. Change advertising by Δa_t in period t only
3. Calculate effect on profits in period $t \leq \tau \leq t + L$

$$\Delta \pi_{d\tau} = m \cdot p_{d\tau} \cdot \Delta Q^H_{d\tau}$$

   - Q^H_{dt} ... volume in market d calculated using Homescan data
   - m ... manufacturer margin as a percentage of final retail price

4. Sum over all markets to obtain effect on national profits, $\Delta \pi_\tau$

ROI:

$$\mathrm{ROI}_t = \frac{\sum_{\tau=t}^{t+L} \Delta \pi_\tau - c\Delta a_t}{c\Delta a_t}$$

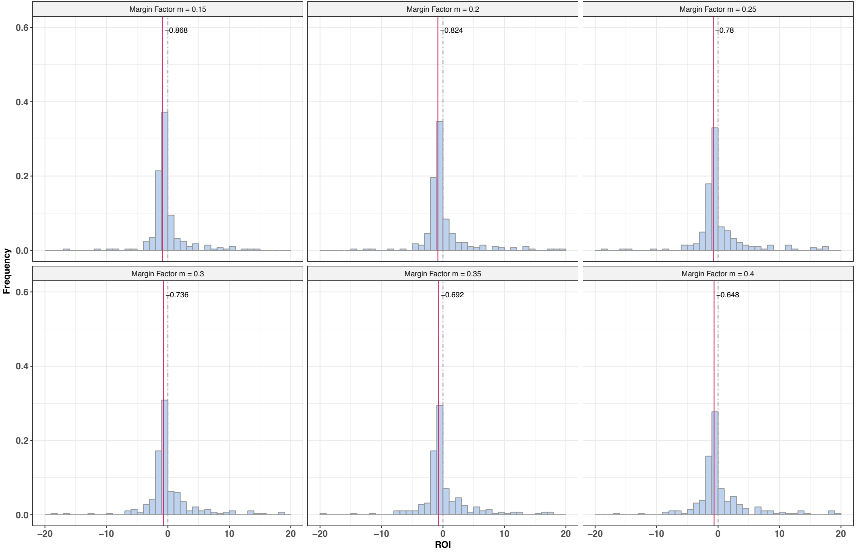
- c ... cost per GRP

Manufacturer margin m

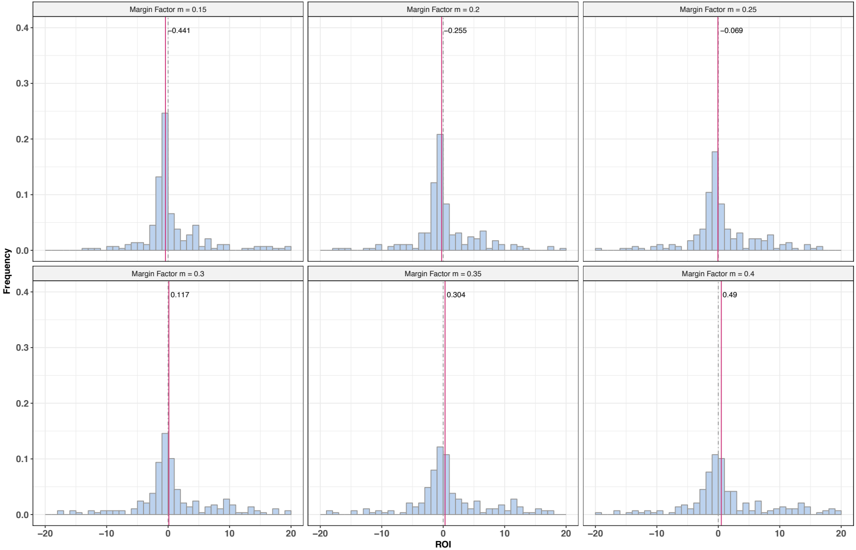
Unknown

We provide ROI predictions for a range of plausible values of m

Average ROI of advertising in weeks when $a_{jt} > 0$



Average ROI of all observed advertising



Conclusions

Mean elasticities among 288 brands between 0.0233 and 0.0258

- Much smaller than prior benchmarks, such as 0.12 in Sethuraman et al. (2011) meta-analysis

About two thirds of estimates not statistically different from zero at 5 percent level, about one quarter positive and "significant"

- Null-effects not an artifact due to lack of statistical power

Mean elasticity is 0.10 if we condition only on positive, "significant" ($p < 0.05$) estimates

- Consistent with publication bias

Preliminary results on ROI's indicate over-spending on TV advertising

- Need large, unmeasured long-run advertising effect to justify observed ad spending

Unfinished work (separate paper) documents challenges in pinning down advertising carry-over effects

Generalizable results as a prior

It appears that the publication process has failed to provide a prior on advertising effects that is useful for decision-making

This research

- Tries to correct the record on advertising

Summary of this lecture

Two questions that motivated this lecture

Does marketing (or business analytics in general) differ from IO or applied microeconomics in general, and if so, how?

- Normative versus positive research approach

What is the role of academia in a world where companies employ economists, marketers, computer scientists, statisticians, psychologists, ... ?

- Create methods
- Create generalizable knowledge