Online Advertising Incrementality Testing: Practical Lessons, Paid Search and Emerging Challenges

Joel Barajas[4(✉)], Narayan Bhamidipati[1], and James G. Shanahan[2,3]

[1] Yahoo Research, Sunnyvale, CA 94089, USA
narayanb@yahooinc.com
[2] Church and Duncan Group Inc, San Francisco, CA, USA
[3] UC Berkeley, Berkeley, CA, USA
[4] Amazon, Sunnyvale, CA 94086, USA
joelbz@amazon.com

Abstract. Online advertising has historically been approached as an ad-to-user matching problem within sophisticated optimization algorithms. As the research and ad tech industries have progressed, advertisers have increasingly emphasized the causal effect estimation of their ads (incrementality) using controlled experiments (A/B testing). With low lift effects and sparse conversion, the development of incrementality testing platforms at scale suggests tremendous engineering challenges in measurement precision. Similarly, the correct interpretation of results addressing a business goal requires significant data science and experimentation research expertise.

We propose a practical tutorial in the incrementality testing landscape, including:

- The business need
- Literature solutions and industry practices
- Designs in the development of testing platforms
- The testing cycle, case studies, and recommendations
- Paid search effectiveness in the marketplace
- Emerging privacy challenges for incrementality testing and research solutions

We provide first-hand lessons based on the development of such a platform in a major combined DSP and ad network, and after running several tests for up to two months each over recent years. With increasing privacy constraints, we survey literature and current practices. These practices include private set union and differential privacy for conversion modeling, and geo-testing combined with synthetic control techniques.

1 Learning Objectives and Scope

Even though there are currently solutions to evaluate the advertising effectiveness with randomized experiments, many details and recommendations rarely appear in papers. This tutorial provides a 360-degree view of the topic, from engineering designs to experiment planning and business use cases.

J. Barajas—Work done while the author was employed at Yahoo.

M. Hagen et al. (Eds.): ECIR 2022, LNCS 13186, pp. 575–581, 2022.
https://doi.org/10.1007/978-3-030-99739-7_72

The key benefits to participants include:

- Specific recommendations to the correct execution of A/B testing
- Online advertising testing engineering designs and econometric evaluation approaches
- Marketing use cases for online advertising incrementality testing
- Review of paid search effectiveness evaluation literature and challenges to operationalizing the estimations
- Review of emerging challenges fo incrementality testing within privacy constraints

Participants will:

1. Identify and formulate key approaches to measuring the effectiveness of online advertising.
2. Execute relevant statistics for hypothesis testing, power analysis in experiment planning and simulate experiment scenarios.
3. Be able to define key ingredients of an operational incrementality testing platform and their trade-offs.
4. Understand the business need for incrementality testing.
5. Identify the necessary conditions to increase the likelihood of successful test given minimum detectable lift, conversion type, test duration among others.
6. Differentiate between demand generation advertising (display, social ads) and demand capture advertising (paid search) incrementality measurement.

2 Tutorial Outline

Part 1 The basics: context and challenges [8,12,15,19]
- The problem
  - Online Advertising spend trends between performance and brand
  - Big picture problem: quarterly/yearly budget allocation
  - Budget allocation practices based on financial models
  - The need for testing combined with industry attribution practices
- How channel-level testing fits within other forms of testing
  - Real-time decision making in targeting engines
  - Tactic testing: A/B testing with last-touch attribution
  - Multi-cell testing A/B testing + Incrementality testing
  - CMO decision making at the end of the quarter/semester/year
- Business Use cases
  - Advertiser joining new partners
  - Testing to calibrate and rebase financial models
  - Media Mix Models calibration
  - Last-touch attribution multipliers
  - The Marketing component: Growth marketing vs CRM marketing

Part 2 Incrementality Testing: concepts, solutions and literature [2,3,14,22]
- Literature and Industry practices
  - Placebo based testing: practice and issues

    - Intention to treat testing
    - Ghost ads testing proposal
  - Estimation Frameworks
    - Econometric causality
    - Potential Outcomes Causal Framework
    - Pitfalls

Part 3 From concept to production: platform building, challenges, case studies [3,16]
  - Building the experiment platform journey
  - The identity graph and treatment groups
    - Cookie-based experiments
    - Device-based experiments
    - Logged-in users based experiments
    - Household-level experiments
  - User holdout design within modern Ad tech serving systems
    - The hashing functions
    - The challenge with targeting and scoring algorithms
    - How to avoid targeting bias
    - The role of look-back windows in last-touch attribution engines
  - Data Logging and Analysis

Part 4 Deployment at Scale: test cycle and case studies
  - Experiment execution cycle
    - Experiment Design and Planning
    - Intervention Execution
    - Experiment Tracking and Metrics
    - End of experiment readout
  - Case Studies
    - Insurance quotes and comparison with post-click conversions
    - Online food ordering revenue: CRM versus New audiences
    - Online acquisition signup

Part 5 Paid Search Incrementality Testing [7,9,21]
  - Evaluating Demand Capture Channels
    - The challenge with demand capture ads in paid search
    - Organic versus paid search results
    - The effects on the search marketplace
  - Tehniques with Aggregate Data
    - Differences-in-Differences
    - Synthetic Control

Part 6 Emerging trends: identity challenges, industry trends and solutions [1,6,7]
  - Advertisers Testing without Ad Network holdouts
    - Spend as experiment intervention
    - Methodologies: Time series based testing
  - Geo-testing
    - Geo units specification
    - Geo unit treatment assignment

    - The power of A/A tests in the experiment design
  - Emerging challenges with user ids
    - Private set Intersection
    - Differential Privacy
    - Identity fragmentation challenges

3 Authors Biography

Joel Barajas, Sr Research Scientist, has over 11 years of experience in the online advertising industry with research contributions at the intersection of Ad tech, Marketing Science, and Experimentation. He has experience with Ad load personalization and experimentation in a publisher marketplace. Within Marketing Data Science, he has supported regular budget allocation and Media Mix Models in multi-channel advertising. With a PhD dissertation focussed on ad incrementality testing, his published work has appeared in top outlets including INFORMS Marketing Science Journal, ACM CIKM, ACM WWW, SIAM SDM. He led the science development and marketing analytics of the incrementality testing platform in a multidisciplinary team. He currently oversees most incrementality tests in Verizon Media ad network (previously yahoo!) and DSP (previously AOL advertising.com). Joel also leads the science development in CTV and linear TV measurement modeling. He holds a B.S. (with honors) in Electrical and Electronics Engineering from the Tecnológico de Monterrey, and a PhD in Electrical Engineering (with emphasis on statistics) from UC Santa Cruz.

Narayan Bhamidipati, Sr Director of Research, has over 14 years of experience in Computational Advertising and Machine Learning. He currently leads a team of researchers focused on providing state-of-the-art ad targeting solutions to help ads be more effective and relevant. This includes creating various contextual targeting products to reduce the company's reliance on user profiles and help improve monetization in a more privacy aware world. Alongside that, Narayan ensures that the user profile based ad targeting products continue to improve despite the decline of tracking data. In addition, Narayan is keen on developing the most accurate ad effectiveness measurement platform which would help the company attract more revenue by proving the true value of the ad spend on our platforms. He holds B.Stat(Hons), M.Stat and PhD(CS) degrees, all from the Indian Statistical Institute, Kolkata.

Dr. James G. Shanahan has spent the past 30 years developing and researching cutting-edge artificial intelligence systems, splitting his time between industry and academia. For the academic year 2019–2020, Jimi held the position of Rowe Professor of Data Science at Bryant University, Rhode Island. He has (co) founded several companies that leverage AI/machine learning/deep learning/computer vision in verticals such as digital advertising, web search, local search, and smart cameras. Previously he has held appointments at AT&T (Executive Director of Research), NativeX (SVP of data science), Xerox Research (staff research scientist), and Mitsubishi. He is on the board of Anvia, and he also advises several high-tech startups including Aylien, ChartBoost, DigitalBank, LucidWorks, and others. Dr. Shanahan received his PhD in engineering mathematics and computer vision from the University of Bristol, U. K. Jimi has been involved with KDD since 2004 as an author, as a tutorial presenter, and as a workshop co-chair; he has actively been involved as a PC/SPC member over the years also.

4 List of References by Topic

4.1 The need for Incrementality Testing Solutions

- *A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook* by Gordon *et al.* (2019) [12]
- *Do display ads influence search? Attribution and dynamics in online advertising* by Kireyev *et al.* (2016) [15].
- *Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment* by Li and Kannan (2014) [19].
- *Evaluating online ad campaigns in a pipeline: causal models at scale* by Chan *et al.* (2010) [8].

4.2 Incrementality Testing Solutions

- *Incrementality Testing in Programmatic Advertising: Enhanced Precision with Double-Blind Designs* by Barajas and Bhamidipati (2021) [3]
- *Ghost ads: Improving the economics of measuring online ad effectiveness* by Johnson *et al.* (2017) [14].
- *Experimental designs and estimation for online display advertising attribution in marketplaces* by Barajas *et al.* (2016) [2].
- *Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising* by Lewis *et al.* (2011) [18].

4.3 Causal Inference

- *Causal inference using potential outcomes: Design, modeling, decisions* by Rubin (2005) [22]
- *Principal stratification in causal inference* by Frangakis and Rubin (2002) [11].
- *Bayesian inference for causal effects in randomized experiments with noncompliance* by Imbens and Rubin (1997) [13].

4.4 Operationalization and Practical Recommendations

- *Incrementality Testing in Programmatic Advertising: Enhanced Precision with Double-Blind Designs* by Barajas and Bhamidipati (2021) [3]
- *Trustworthy online controlled experiments: A practical guide to a/b testing* by Kohavi *et al.* (2020) [16].
- *The unfavorable economics of measuring the returns to advertising* by Lewis *et al.* (2015) [17].

4.5 Paid Search Incrementality Testing

– *Consumer heterogeneity and paid search effectiveness: A large-scale field experiment* by Blake *et al.* (2015) [7].
– *Sponsored Search in Equilibrium: Evidence from Two Experiments* by Moshary (2021) [21]
– *Effectiveness of Paid Search Advertising: Experimental Evidence* by Dai and Luca (2016) [9].

4.6 Geo-testing and Synthetic Control and Identity Challenges

– *Advertising Incrementality Measurement using Controlled Geo-Experiments: The Universal App Campaign Case Study* by Barajas *et al.* (2020) [6]
– *Consumer heterogeneity and paid search effectiveness: A large-scale field experiment* by Blake *et al.* (2015) [7].
– *Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program* by Abadie *et al.* (2010) [1].
– *The identity fragmentation bias* by Lin and Misra (2020) [20].

References

1. Abadie, A., Diamond, A., Hainmueller, J.: Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. J. Am. Stat. Assoc. **105**(490), 493–505 (2010)
2. Barajas, J., Akella, R., Holtan, M., Flores, A.: Experimental designs and estimation for online display advertising attribution in marketplaces. Mark. Sci. **35**(3), 465–483 (2016)
3. Barajas, J., Bhamidipati, N.: Incrementality testing in programmatic advertising: enhanced precision with double-blind designs. In: Proceedings of the Web Conference 2021, pp. 2818–2827. WWW 2021, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3442381.3450106
4. Barajas, J., Bhamidipati, N., Shanahan, J.G.: Online advertising incrementality testing and experimentation: industry practical lessons. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 4027–4028. KDD 2021, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3447548.3470819
5. Barajas, J., Bhamidipati, N., Shanahan, J.G.: Online advertising incrementality testing: practical lessons and emerging challenges. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 4838–4841. CIKM 2021, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3459637.3482031
6. Barajas, J., Zidar, T., Bay, M.: Advertising incrementality measurement using controlled geo-experiments: the universal app campaign case study (2020)
7. Blake, T., Nosko, C., Tadelis, S.: Consumer heterogeneity and paid search effectiveness: a large-scale field experiment. Econometrica **83**(1), 155–174 (2015)
8. Chan, D., Ge, R., Gershony, O., Hesterberg, T., Lambert, D.: Evaluating online ad campaigns in a pipeline: causal models at scale. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 7–16. KDD 2010, ACM, New York, NY, USA (2010). https://doi.org/10.1145/1835804.1835809,http://doi.acm.org/10.1145/1835804.1835809

9. Dai, D., Luca, M.: Effectiveness of paid search advertising: Experimental evidence. Technical report, Harvard Business School (October 2016). workin Paper No. 17–025
10. Farahat, A., Shanahan, J.: Econometric analysis and digital marketing: how to measure the effectiveness of an ad. In: Proceedings of the sixth ACM International Conference on Web Search and Data Mining, pp. 785–785 (2013)
11. Frangakis, C., Rubin, D.: Principal stratification in causal inference. Biometrics **58**(1), 21–29 (2002). https://doi.org/10.1111/j.0006-341X.2002.00021.x
12. Gordon, B.R., Zettelmeyer, F., Bhargava, N., Chapsky, D.: A comparison of approaches to advertising measurement: evidence from big field experiments at Facebook. Mark. Sci. **38**(2), 193–225 (2019)
13. Imbens, G.W., Rubin, D.B.: Bayesian inference for causal effects in randomized experiments with noncompliance. Ann. Stat. **25**(1), 305–327 (1997). http://www.jstor.org/stable/2242722
14. Johnson, G.A., Lewis, R.A., Nubbemeyer, E.I.: Ghost ads: improving the economics of measuring online ad effectiveness. J. Mark. Res. **54**(6), 867–884 (2017). https://doi.org/10.1509/jmr.15.0297
15. Kireyev, P., Pauwels, K., Gupta, S.: Do display ads influence search? Attribution and dynamics in online advertising. Int. J. Res. Mark. **33**(3), 475–490 (2016)
16. Kohavi, R., Tang, D., Xu, Y.: Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing. Cambridge University Press, Cambridge (2020)
17. Lewis, R.A., Rao, J.M.: The unfavorable economics of measuring the returns to advertising *. Q. J. Econ. **130**(4), 1941–1973 (2015)
18. Lewis, R.A., Rao, J.M., Reiley, D.H.: Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising. In: Proceedings of the 20th International Conference on World Wide Web, pp. 157–166. WWW 2011, ACM, New York, NY, USA (2011). https://doi.org/10.1145/1963405.1963431,http://doi.acm.org/10.1145/1963405.1963431
19. Li, H.A., Kannan, P.: Attributing conversions in a multichannel online marketing environment: an empirical model and a field experiment. J. Mark. Res. **51**(1), 40–56 (2014)
20. Lin, T., Misra, S.: The identity fragmentation bias (2020)
21. Moshary, S.: Sponsored search in equilibrium: evidence from two experiments. Available at SSRN 3903602 (2021)
22. Rubin, D.B.: Causal inference using potential outcomes. J. Am. Stat. Assoc. **100**(469), 322–331 (2005). https://doi.org/10.1198/016214504000001880