International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Empirical Analyses of the Bias in User Based Randomized Control Trials Created Due to Difference in Unit of Randomization and Unit of Measurement in Digital Ad Delivery Platforms

Varun Chivukula

varunvenkatesh88@berkeley.edu

Abstract

Incrementality measurement in advertising is considered as the gold standard for determining the causal impact of ads. However, misalignment between the units of randomization and measurement introduces biases that significantly distort lift outcomes. This paper examines such challenges when the unit of randomization is an ad platform user, but the unit of conversion measurement aligns with advertiser accounts. Using a probabilistic model, we quantify the understatement in measured lift caused by this misclassification, identify drivers, and propose potential solutions.

Keywords: Causal inference, Digital Ad platforms, Randomized control trials, Misclassification Bias, Halo effect

Introduction

Digital ad platforms rely on randomized control trials (RCTs) to estimate the causal effect of advertisements. Ideally, the units of randomization and measurement should align to prevent contamination. However, in scenarios where these units differ, such as when platform-level randomization intersects with account-based conversion measurement, significant bias can arise. This discrepancy manifests in misclassification, where conversions are inaccurately attributed to control or test groups.

For instance, consider an ad platform conducting a randomized test where users are the randomized units, while an advertiser measures conversions at the household level. This scenario often leads to cases where a household is split between test and control groups due to multiple linked users. This paper models and quantifies the understatement in lift measurement resulting from such misclassification, providing a structured analysis.

The Problem: Misclassification Bias

Scenario

When ad platforms randomize users while advertisers measure conversions at an account level (e.g., household accounts), discrepancies arise. A household may have one user in the test group exposed to ads and another in the control group, resulting in misaligned conversion attribution. Consequently, conversions driven by test-group exposure may erroneously appear in the control group's metrics.

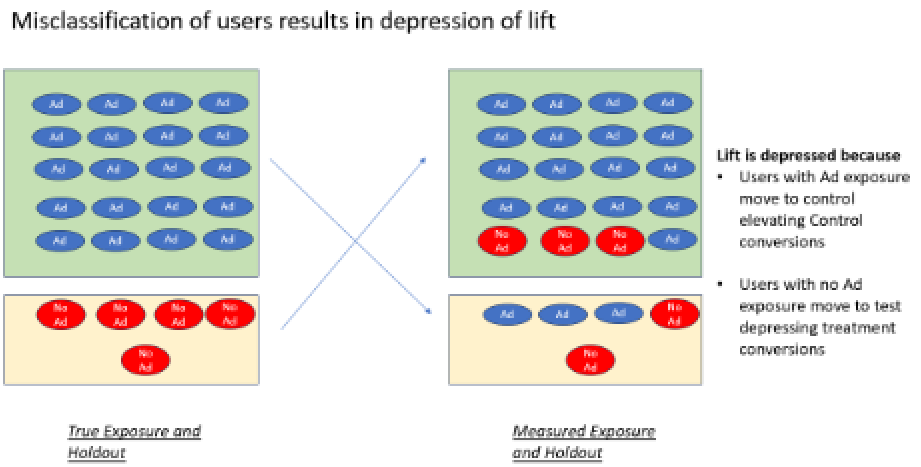
Example

Consider in table 1 two advertiser accounts (A1, A2), each linked to two platform users (M1–M4) under a 50:50 randomized control setup. Account holders may be assigned to different groups, but conversions are aggregated at the account level, leading to potential contamination.

Advertiser Household	Ad Platform User	Assignment (Ad Platform)	Where Advertiser conversion is counted in RCT	Key on Ad Platform Side	Key on Advertiser Side
A1	M1	Test	Test	Ph1	Ph1
A1	M2	Control	Test	Ph2	Ph1
A2	M3	Test	Control	Ph3	Ph4
A2	M4	Control	Control	Ph4	Ph4

Table 1.

This effect will depress lift as illustrated below as seen in Fig 1.



Methodology: Modeling Misclassification Bias

Scenario Overview

Consider an experimental setup where:

- The ad platform randomizes at the individual user level.
- The advertiser aggregates conversions at the household (or account) level.

Each account typically contains multiple users. Misclassification arises when users within the same account are allocated to different experimental groups (test/control) but their conversions are aggregated under a single account, leading to improper attribution. 2.

- N: Total number of advertiser accounts.
- U: Average number of users per account.
- P_{test}: Proportion of users in the test group.
- $P_{control}$: Proportion of users in the control group $(1 - P_{test})$.
- C_{base}: Baseline conversion rate (control group).
- L_{true}: True lift induced by ad exposure $(L_{true} > 0)$.

For simplicity, assume that $P_{test} = P_{control} = 0.5$ in a balanced experimental design. Let M denote the probability of misclassification within an account, driven by the distribution of users between groups.

Impact of Misclassification

The measured lift ($L_{measured}$) is derived from observed test and control conversions:

$$L_{measured} = \frac{C_{test} - C_{control}}{C_{control}},$$

where:

- C_{test}: Conversions observed in the test group.
- $C_{control}$: Conversions observed in the control group.

Under misclassification, the true test-group conversions ($C_{test,true}$) are attenuated by contributions from control-assigned users within the same account, and vice versa:

$$C_{test} = C_{test,true} \times (1 - M) + C_{control,true} \times M.$$

$$C_{control} = C_{control,true} \times (1 - M) + C_{test,true} \times M.$$

This contamination reduces the numerator in $L_{measured}$, systematically understating lift.

Simulation Results

A simulation with the following parameters illustrates the impact:

- $N = 10,000$ accounts.
- $U = 2$ users/account.
- $C_{base} = 1\%$, (L_{true} = 3.6%.

Key findings:

1. **True Lift (L_{true})**: 3.6%.
2. **Observed Lift ($L_{measured}$)**: 1.78%.
3. **Understatement**: 49%.

As U increases (e.g., larger household sizes), the probability of misclassification rises, further suppressing $L_{measured}$.

Probabilistic Framework

Further when we simulate the understatement by varying the ratio of ad platform users per advertiser household we see that the understatement increases with as the ratio increases as seen in Fig 2.

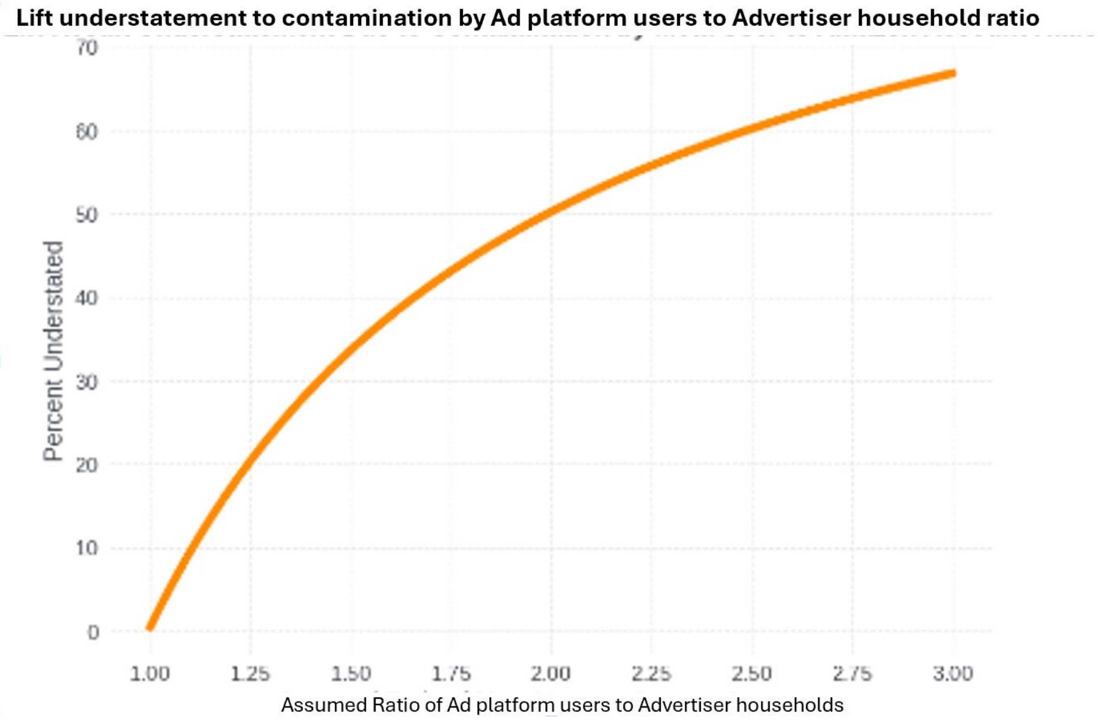


Fig 2.

Findings and Implications

1. **Impact of Household Size:** Larger households increase the probability of misclassification, exacerbating understatement.
2. **Halo Effects:** Spillover effects within households (e.g., one user influencing another) further distort causal attribution.

Methodological Adjustments

1. **Randomization Alignment:** Randomizing at the account level eliminates intra-account misclassification. While this approach simplifies measurement, it requires platforms to adopt advertiser-specific data structures.
2. **Cohort Analysis:** Stratify results based on account size to estimate and adjust for bias. For example, single-user accounts serve as a baseline for uncontaminated lift.
3. **Enhanced Measurement Systems:** Platforms should facilitate finer-grained tracking of user-account mappings during experiments to more accurately attribute conversions.

Conclusion

This study underscores the critical need to align randomization and measurement units in incrementality testing. Misclassification, driven by household structures and platform-advertiser discrepancies, significantly understates the true lift of advertising campaigns. By implementing short- and long-term strategies, both platforms and advertisers can improve measurement accuracy and better capture the causal effects of their campaigns.

References

1. Bakshy, E., Eckles, D., & Bernstein, M. (2014). *Design and Analysis of Ad Experiments on Digital Platforms*. Proceedings of the ACM.
2. Johnson, G. A., & Lewis, R. A. (2020). *The Economics of Advertising Experiments*. Journal of Marketing Research.
3. Athey, S., & Imbens, G. (2017). *The State of Applied Econometrics: Causality and Policy Evaluation*. Journal of Economic Perspectives.
4. Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press.
5. Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.
6. Thomke, S. H. (2020). *Experimentation Works: The Surprising Power of Business Experiments*. Harvard Business Review Press.
7. Lewis, R. A., & Rao, J. M. (2015). *On the Near Impossibility of Measuring the Returns to Advertising*. Quarterly Journal of Economics.
8. Xu, Y., Chen, N., & Thorpe, M. (2020). *Challenges in Causal Inference for Online Advertising: Theoretical and Practical Issues*. Journal of Applied Econometrics.
9. Varian, H. R. (2014). *Big Data: New Tricks for Econometrics*. Journal of Economic Perspectives.
10. Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.