Simulating the Superiority of Click-Value Prediction over Conversion-Based Optimization in Digital Advertising

Authors:

Matt Butler, CEO, Bonsai.

Lydia Fern, VP Product, Bonsai.

Abstract

Digital advertising buying automation generates materially improved advertiser outcomes when trained on algorithms designed to predict the incremental advertiser value of each ad click, when compared to equivalent buying trained on historical digital ad conversion data. Predicted click-value outcomes substantially outperform current prevalent interventions on ad conversion data improvement - API-driven conversion data augmentation - with respect to improving advertiser outcomes. We prove this by simulating the theoretical limit of ad conversion data improvement, a complete retention of 'incremental' ad conversion results, and a complete elimination of ad conversion results sans incremental advertiser value. For many cases, predicted click-value accuracy need only meet accuracy standards fractionally better than random chance to ensure outperformance against even the most stringent and accurate conversion data pipelines. Finally, this method generalizes. It applies to all known advertiser situations where conversions are training data for programmatic bid management.

1. Introduction

1.1. Modern Ad Platform Bid Optimization:

Modern digital advertising platforms (Google, Meta, etc.) rely on their own proprietary bidding algorithms to automate ad buying on behalf of millions of advertisers across the globe every day. Some commercially known tools with large adoption include Google's *Smart Bidding* product and Facebook's *Bid Strategies* products. These products execute fundamental bidding theory, often called "value-based bidding", with speed and scale across billions of unique ad interactions available for advertiser auction. The speed and scale of ad platform operated bid algorithms have captured the value arbitrage opportunities available due to speed and scale. While value-based bidding isn't a particularly complicated algorithm with a meaningful moat, there's significant variance across the levers in which value-based bidding decisions can be applied. The impact mirrors those seen in financial markets for firms with access to hardware and software systems capable of executing high-frequency trading versus their analog counterparts.

Though industry insiders popularize the notion that ad platform bid automation has completely captured any remaining value margin on behalf of advertisers, this ignores one significant limitation to these proprietary bid management algorithms. All of them require a feed of training data. This feed of integer or floating-point data is commonly called "conversion data". It's collected via tracking pixel, or sent from advertiser systems to ad platforms via API connection. Customer actions that trigger a 'conversion' vary by industry, but the most common include actual online or offline transactions at a point-of-sale, either via eCommerce website or through a data connection to physical points-of-sale, subsequent to an ad interaction, most often an 'ad click'.This record of customer purchases become 'conversions', and the transaction values translate into 'conversion values', and these data are sent back to ad platforms along with their matched ad click identifiers[1] in regular batch or streaming data pipelines. From that point forward, they function as training datasets for proprietary ad platform bidding algorithms.

Sending 'conversion data' to ad platforms for bid automation purposes has become the industry standard, and is defined by platforms themselves as the best-in-class approach to training bid algorithms to buy ads that will generate the most future advertiser value. With respect to sophistication, ad platforms regularly communicate and champion a basic principle: the more conversion data you can collect and send, the better.

1.2 The Problem with Conversion-Based Optimization

Three critical issues pervade the conversion-based training dataset model for modern bid optimization: One, consumer privacy concerns, are frequently discussed, yet rarely addressed in the United States, because economic incentives of a relatively small list of large companies outweigh negative externalities disbursed amongst millions of consumers. The other two: ***conversion data sparseness*** and directional ***predictive inaccuracy***, are rarely discussed, and not yet widely understood.

First, conversion data is inherently **sparse**. Because consumer purchase decisions are complicated and require many independent datapoints, clicks outnumber conversions by a wide margin in almost every vertical market. 'Conversion rates' are marketing industry jargon for describing this relationship, and simply refer to the quantity of purchases per period versus the requisite number of advertising clicks that occur prior to those purchases. 'Conversion rates' vary by industry, but are typically lower than 5%, and can be as small as 0.1% in some high-ticket, high-consideration industries like housing, automotive, travel, and others.

This isn't necessarily good or bad independent of bid management, but it's a particular issue for bidding algorithms that require training data (conversion data) to optimize outcomes for advertisers. Clicks that don't 'convert' don't return any information at all in the training data. This means that for an advertiser in an industry with a standard conversion rate of 1%, 99 out of 100 clicks provide no training data at all to the bid algorithm.

Bid algorithms, therefore, are largely forced to treat that absence of data as a signal that those clicks are value-less. It cannot reach that conclusion too soon, as it would immediately circumvent the algorithm's ability to reach customers who are indeed in a successful purchase journey prior to completion, but it cannot wait too long, as it would mean that the algorithm's wasting ad impressions on consumers who will never successfully convert.

Reducing training data sparseness is a long-standing technical issue that parties have been attempting to solve for decades. 'Attribution models' are methods by which sparse conversion data are made slightly-less sparse, by taking conversions and separating value into fractional segments across multiple ad interactions, when ad interactions can be joined together through consumer identifiers[2]. Common

---

[1] Matchable ad identifiers include: cookie, click, or personal customer identifiers, such as email or phone number

[2] Common, legal examples include cookie,web browser, and device identifiers.

examples are rules-based attribution models, such as linear attribution, and probabilistic models, such as Google's Data-Driven Attribution model. These techniques reduce data sparseness, but do so only marginally. It's common to see these techniques increase the measurable training dataset (conversion data with values > 0) by three to five times. This would still leave the unscored sample of clicks in the prior example at 95%[3].

Only recently discussed in the industry zeitgeist is the second issue: ***predictive inaccuracy.*** The trouble with simply sending a log of purchases as a measure of advertiser value is that a purchase needs to be 'incremental' to the advertiser's existing customer outcomes today. Said differently, if ads are shown to people who are already in the process of buying, it generates an ad cost, but doesn't increment the advertiser's outcomes. Ads that aren't driving incremental outcomes create a net-loss of advertiser value.

Conversion datasets cannot differentiate between purchases that are incremental, and therefore useful to bid algorithms, and conversions that aren't incremental, which train bid algorithms to waste advertiser funds.

Should the underlying conversion data be more non-incremental than incremental, ad platform bid algorithms could conceivably make advertiser outcomes worse than having no bid management at all.

1.3. Faulty Fixes: Enhancing Conversion Data

Ad platforms champion a remedy to the issues outlined in section 1.2: sending additional customer outcome data to ad platforms via API connection.

This mainly addresses the sparseness issue with conversion data, but can also be seen as a method to address the predictive accuracy of said data. An example of how this intervention addresses sparseness is the inclusion of physical store sales data in training datasets, by sending this data via API. Many businesses experience more transactions in physical stores than they do on eCommerce websites. An example of how this intervention addresses predictive accuracy is the ability to send special financial information to ad platforms for each customer outcome via API, say the 'contribution margin' associated with a transaction, often not available in real-time via ad pixel data collection, but reconcilable and sharable after a purchase is evaluated against a business' pricing, cost-of-goods sold, and other operational expense information.

Third parties offer a lesser-used, but additional method to address predictive accuracy. "Incrementality attribution" remains a specious practice with limited evidence of efficacy, yet theoretically offers advertisers a method by which each individual conversion value is 'adjusted' according to the relative likelihood (or measured reality) that each individual conversion was an 'incremental' business outcome.

In practice, while this will increase the signal *quality* substantially, this would allow only for *training signal volume reduction* to bid algorithms, because the data being eliminated would be conversions measured as non-incremental, or values adjusted downward to the incremental fraction of each conversion value total. While this improves data *purity*, it doesn't solve the problem of data *sparsity*.

1.4. A New Paradigm: Predicted Click-Value (pCV)

A superior approach is to change the training data target entirely. Instead of training on a *few, "perfect" conversion signals*, platforms should train on *many, "good enough" click-value signals*. This method both

[3] 1% signal strength becomes a 5% signal strength for the purposes of bid algorithm training

capitalizes on the power of predictive analytics and advanced machine learning models, and represents the logical extension of the existing intent of multi-touch attribution methods.

Fractional attribution attempts to spread conversion values across multiple ad interactions, as both intuition and research conclude that consumer conversion journeys are multi-interaction. These ‘advanced’ attribution methods aren’t limited by theoretical backing: industry practitioners have advocated for measuring value of ad interactions prior to the ‘last-interaction’ for quite some time, and practitioners would agree that a perfect attribution model would reflect the relative incremental value of each and every ad interaction towards future customer outcomes.

However, technical limitations doom this approach: cookie deprecation, device proliferation, data storage and compute requirements are among the myriad of reasons the notion of capturing and keeping a perfectly complete record of every single ad interaction’s position in every single customer outcome hopelessly ineffective.

Our key insight relies on reminding the purpose of conversion data in the first place: its core function is as a proxy dataset, representing advertiser value, training automated bidding systems to learn which next ad interactions are most worth purchasing. Opening the aperture wider, it turns out that predictive models, not perfect observational records, hold the most promising approach to optimizing the business impact of modern ad platform bid algorithms.

We call this method ‘predicted click value’, or pCV Our definition of pCV is as follows:

A model that predicts the incremental business value of every single ad click, not just clicks in observable, attributable conversion paths.

1.5. Hypothesis & Study Objective

- **Hypothesis 1:** Bidding algorithms trained on pCV data generate materially better incremental advertiser outcomes than algorithms trained on standard conversion data.
- **Hypothesis 2:** pCV will *also* outperform algorithms trained on "theoretically perfect" incremental conversion data (i.e., the best-case scenario for the “fix” from **1.3**).
- **Objective:** To test these hypotheses using a simulation that models the optimization outcomes of these three different bidding inputs over time.

2. Methodology (Simulation Design)

2.1. Model Overview

The objective of this simulation is to compare the cumulative incremental business outcomes generated by three distinct bidding optimization training datasets over a sequence of discrete time periods $t \in \{1, 2, ..., T\}$. The distinct bid optimization training datasets are as follows:

- pCV (Predicted Click Value)
- $1P$ (Enhanced Conversions)
- $3P$ (Traditional Conversions)

The model iteratively applies a per-period efficiency gain of the bidding algorithm based on the quality of the training signal, Q_t, available to each strategy, as generated by each respective training dataset.

2.2. Variable Definitions and Formalization

To ensure mathematical rigor, the simulation variables are formalized as follows:

- N_t **(Ad Click Volume):** The total number of ad interactions purchased in period t.
- C_t **(Measured Conversions):** The volume of conversion events reported via standard tracking mechanisms (e.g., pixels or API).
- I_t **(Incremental Outcomes):** The subset of conversions that represent true incremental business value.
- γ **(Signal Precision):** The predictive accuracy of the model in identifying a click with incremental advertiser value.
- Q_t **(Average Click Quality):** The normalized value signal derived from the previous period's training data.

2.3. Theoretical "Winning" Conditions

The core mathematical evaluation compares the threshold at which pCV outperforms conversion-based alternatives.

- pCV vs. $1P$: The pCV model is considered superior when:
  $$\gamma > \frac{I_t}{N_t} + 0.5$$
  This inequality posits that pCV wins when its predictive accuracy exceeds random chance by a margin greater than the base incrementality rate.
- pCV vs. $3P$: The pCV model is considered superior when:
  $$\gamma > \frac{I_t}{N_t} + 0.5 - \left(\frac{C_t - I_t}{N_t}\right)$$
  This accounts for the penalty introduced by non-incremental conversions present in standard ad conversion datasets.

2.4. Scope and Model Constraints

While this simulation demonstrates the directional superiority of high-density training signals, it operates under specific constraints to isolate the impact of training data on bidding logic. We are not intending to measure the optimal marginal benefit, nor calculate the specific economic impact of one training dataset versus another. Because of this, the following dynamics are not captured in our simulation:

- **Absence of Auction Dynamics:** The simulation does not model changes in Cost-Per-Click (CPC) or competitive bidding response. In a live environment, increased Q_t typically improves ad relevance, potentially lowering CPCs and further magnifying the lift in I_t. This omission ensures the model estimates remain conservative.
- **Asymptotic Limits of Click Quality:** The model assumes Q_t can improve linearly over the simulation window. To avoid unrealistic extrapolation of click quality, the simulation is limited to a discrete period of $T = 27$ iterations.
- **Channel Independence:** The simulation isolates single-channel impact and does not account for

cross-channel attribution effects or multi-channel marketing dynamics.

- **Neutrality of Unscored Clicks:** A significant assumption is that "unscored" clicks (where no signal is present) have a zero-sum impact on the algorithm (0). In practice, a prolonged absence of signal in value-based bidding systems is often interpreted as a negative signal, which would theoretically disadvantage the sparse $1P$ and $3P$ models further than shown here.

---

3. Simulation

3.1. Parameter Initialization

To initialize the simulation at $t = 1$, we define the baseline environment using the following fixed parameters:

- $N_1 = 10,000$
- $C_1 = 100$
- $Q_1 = 1.0$
- $I_1 = 51$
- $\gamma = .51$

The initial period establishes a baseline relationship between ad clicks, resulting ad conversions, and incremental business conversions contributed. It's assumed for the purposes of simulation that the specific amount of incremental business conversions contributed is known[4].

The simulation establishes that valuable ad clicks driving real-world outcomes are inherently valuable, independent of whether a tracking signal captures them. Conversely, bidding algorithms optimize solely toward the provided signal, regardless of its validity.

3.2. Simulation Procedure

For each subsequent period $t + 1$, the algorithm applies the optimization rate α derived from the bidding value signals[5] S_t accrued in the previous period for each of the three distinct training dataset groups: pCV, $1P$ and $3P$. The cumulative signal S_t is derived from three distinct click classifications:

- **Correctly Scored (ps):** Weight $+1$
- **Incorrectly Scored (ns):** Weight -1
- **Unscored/Neutral (zs):** Weight 0

Clicks are 'bought' for each respective group, and the resulting 'click quality' derives from the bidding algorithm's optimization towards the cumulative value signal accrued during t. Cumulative value signal comprises three segments:

1. Clicks contributing to real incremental business outcomes where value signal is positive are "correctly scored" clicks, ps, carrying a positive score 1.

---

[4] In practice, establishing this measurement requires non-trivial inferential analysis.

[5] Pixel or CAPI collected Ad Conversions, MTA-derived 1P Incremental Conversions, or Predicted click-value (pCV)

2. Clicks not contributing to real incremental business outcomes where value signal is positive are "incorrectly scored" clicks, ns, carrying a negative score -1.

3. Clicks without a value signal have a neutral score zs, 0, regardless of real business outcome.

The cumulative signal equation is defined as:

$$S_t = (ps \cdot 1) + (ns \cdot -1) + (zs \cdot 0)$$

The resulting average click quality for each respective group is calculated as $Q_{t+1} = \frac{S_t}{N_t}. + Q_t$.

This quality metric is then applied to the baseline incremental outcomes to determine realized business value:

$$I_{t+1} = I_1 \cdot Q_{t+1}$$

For each of the three distinct training dataset groups: pCV, $1P$ and $3P$. The procedure iterates for 27 periods ($T = 27$).

The simulation establishes the premise that valuable ad clicks driving real-world business outcomes are inherently valuable on their own, owing none of their value to whether or not a value signal successfully measures them. Conversely, value signals[6] are inherently arbitrary, representing true incremental value only when efforts to test their efficacy prove successful inferentially, using statistical techniques, or observationally, using advanced data collection and validation techniques[7]. This means that bid algorithms optimize to the value signal, regardless of validity, and the subsequent success or failure of the bidding algorithm to improve real world business outcomes relies solely on the relationship strength and size between value signal and realized business value. Such relationships can be positive, neutral, or negative. Therefore, bid algorithms improve performance, have no effect, or degrade performance.

3.3. Key Findings

- **Finding 1 (pCV vs. $1P$):** The pCV-trained model generated a **17.8% lift** in incremental business outcomes (I_t) over the Enhanced Conversions $1P$ model after 27 periods. As illustrated in the results, the trajectory of Q_t for pCV increases at a significantly steeper slope than both conversion-based groups.
- **Finding 2 ($1P$ vs. $3P$):** The Enhanced Conversions model $1P$ produced only a **6.6% lift** over the baseline Traditional Conversions model $3P$. This marginal gain suggests that industry efforts to refine sparse conversion data yield diminishing returns.
- **Finding 3 (Compounding Density):** Despite pCV generating more false positives (ns) due to lower precision, its average click quality (Q_t) increased by **52%** over the simulation window (See *Figure 1)*. This is attributed to the compounding effect of learning from $N_t = 10,000$ data points per period, whereas the sparse $1P$ accuracy increased by only 14% due to signal scarcity. The signal volume gap is visualized in *Figure 2*.

---

[6] Again, Pixel or CAPI collected Ad Conversions, MTA-derived 1P Incremental Conversions, or Predicted Click Value (pCV)

[7] In practice, both methods are often required.

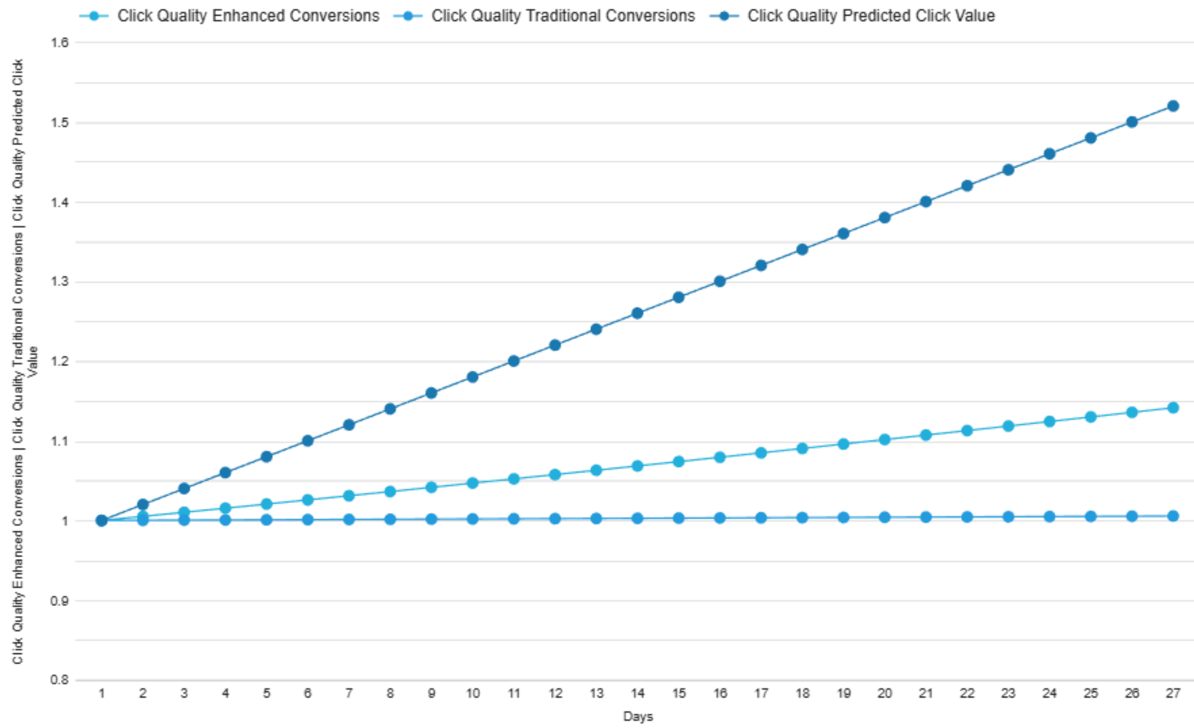

Figure 1: Q_t *(Click Quality) over Time:* pCV *versus Enhanced Conversions* $1P$ *(via API) and Traditional Conversions* $3P$ *(pixel or API)*

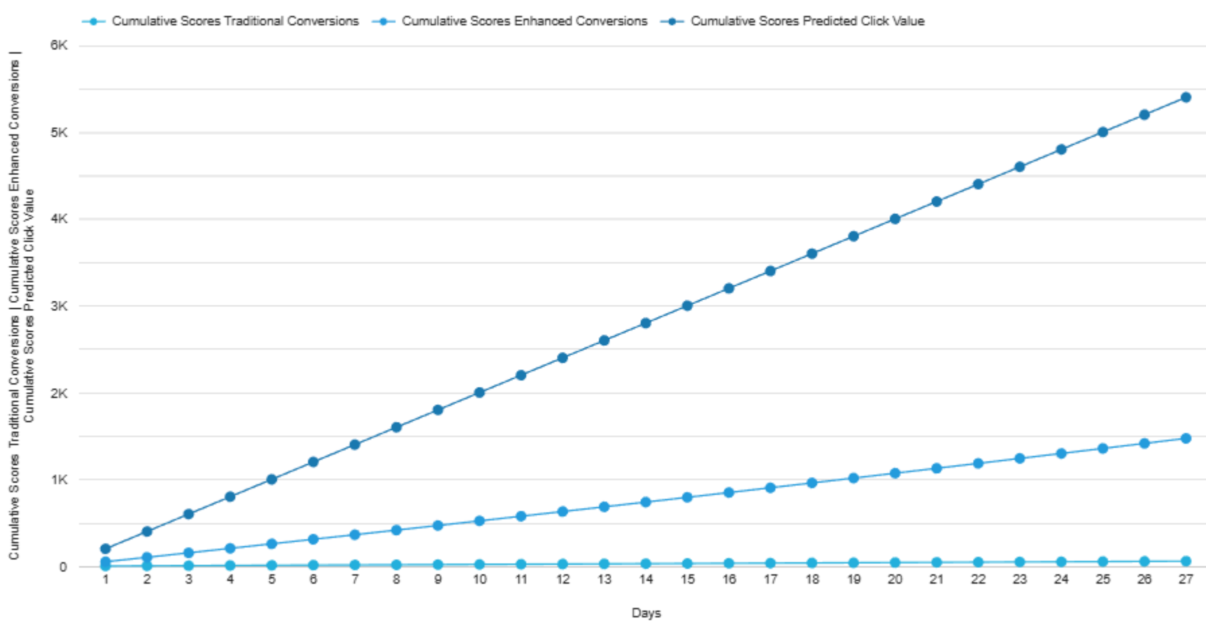

Figure 2: S_t (*Cumulative value signal quantity) over time:* pCV *versus* $1P$ *(via API) and Traditional Conversions data* $3P$ *(pixel or API). Positive numbers → real value improvement*

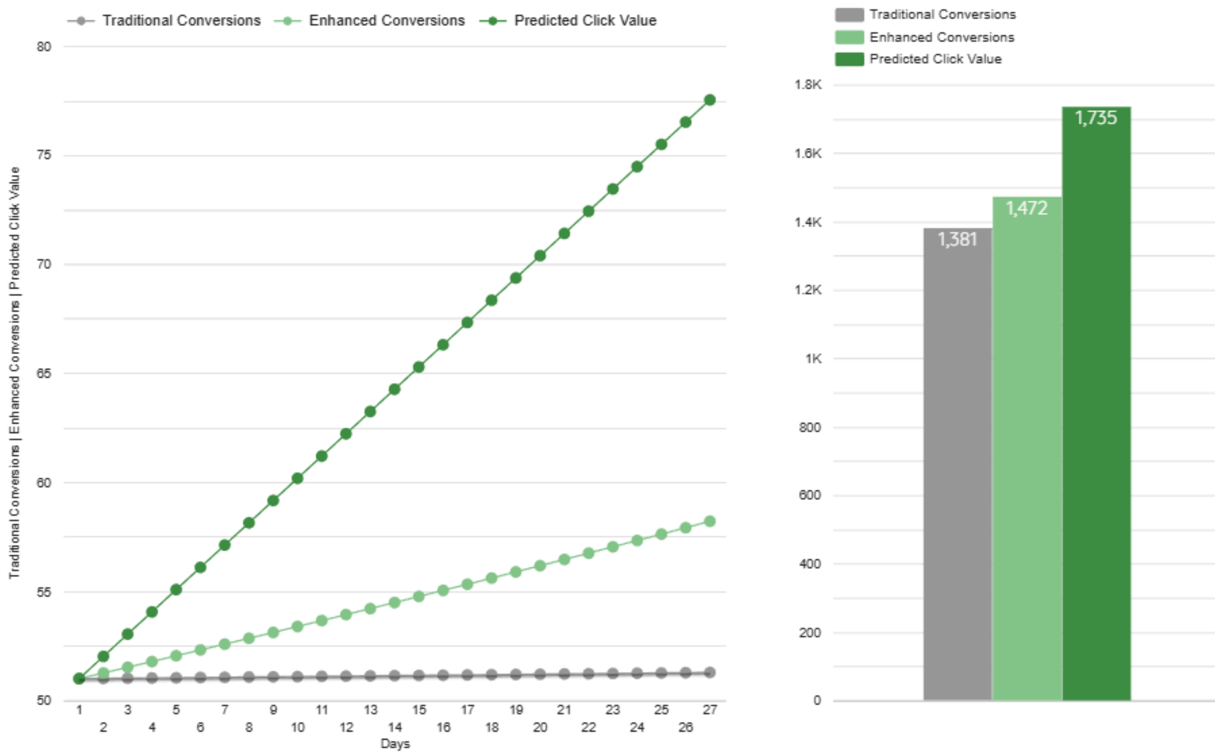

Figure 3: Incremental Business Outcomes over time: pCV *versus* $1P$ *(via API) and standard ad conversion data* $3P$ *(pixel or API)*

4. Related Works

The optimization of digital advertising auctions is increasingly governed by machine learning models that struggle with two primary challenges: **Data Sparsity** and **Sample Selection Bias**. The Predicted Click Value pCV framework addresses these by shifting the training target from sparse outcomes to dense predictions of incrementality with modest accuracy.

4.1. CVR Sparsity and Entire Space Modeling

In digital advertising, the Conversion Rate (CVR) is measured only on the subset of users who both clicked on an ad, N_t, and converted subsequent to that ad click, C_t, a phenomenon known in academic literature as *Sample Selection Bias*. Because C_t is typically a small fraction (e.g., <1%) of N_t, models frequently suffer from "Data Sparsity".

The pCV approach aligns with the Entire Space Multi-Task Model (ESMM)[8] framework discussed by Ma et. al (2018), which suggests that training on the entire interaction space N_t—rather than just the conversion space C_t—alleviates the information bottleneck. By assigning a score to every click, pCV ensures that the bidding algorithm receives a continuous training signal S_t, even in the absence of a

[8] Ma, Xiao, et al. "Entire space multi-task model: An effective approach for estimating post-click conversion rate." *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 2018.

terminal conversion C_t.

4.2 Incrementality and Uplift Modeling

A significant limitation of standard conversion data C_t is its inability to distinguish between "Sure Things" (users who would convert without an ad) and "Persuadables" (I_t, or incremental outcomes).

As discussed in Uplift Modeling[9] by Rzepakowski (2011), relatively limited machine learning research exists which addresses maximizing the *Individual Treatment Effect* (ITE). Current conversion-based models optimize for C_t, which include non-incremental value, leading the bidding algorithm to waste resources on users with a high baseline probability of action. The pCV method explicitly targets γ (Signal Precision) relative to I_t, ensuring the model optimizes for the "true lift" rather than mere outcome correlation.

5. Discussion & Implications

5.1 The Inefficiency of Data Purity Efforts:

So what do we make of the massive capital and human investment currently deployed today to "enhance" sparse, sometimes inaccurate conversion data using MTA and API procedures? Our simulation shows improving conversion data quality provides **minimal lift** to advertiser outcomes. Other simulations could generate higher numbers, but the range of outcomes doesn't scale exponentially.

We postulate that efforts to improve conversion data should cease, so that engineering and data science resources across the marketing industry can focus on eliminating the vast quantity of 'unscored' ad interactions sent to platform bid algorithms. Models trained to predict incremental click value offer substantially more bang for the advertiser buck.

5.2 Generalizability & Practical Application:

pCV theory generalizes to *all* click-based, second-price auction systems: Google Ads, Microsoft Ads, Meta (Facebook) Ads, TikTok Ads, and more coming down the road.

Though a broad market of pCV training data products haven't yet come to market, production-ready predicted click value algorithms powering Google Smart Bidding and Meta Bid Strategies already exist from companies like Bonsai[10], and will likely be developed by custom AI companies like Chalice[11].

5.3 Broader Implication for Consumer Privacy:

$1P$'s limited incremental outcome benefit incentivizes advertisers to share *more* sensitive customer data with third-parties than they would without it. pCV *eliminates this economic incentive, because it outperforms* $1P$. Because pCV is trained on click-level behavior and predicts *future* value, it shares no user behavior data, user-identifiers, or customer information at all.

Therefore, adopting a pCV optimization framework can simultaneously improve advertiser outcomes *and* enhance consumer privacy.

---

[9] Rzepakowski, P., Jaroszewicz, S. Decision trees for uplift modeling with single and multiple treatments. Knowl Inf Syst 32, 303–327 (2012). https://doi.org/10.1007/s10115-011-0434-0
[10] bonsaidata.io
[11] chalice.ai

5. Conclusion

This simulation proves that training bidding algorithms on high-density, directionally-correct predicted click-value (pCV) data is materially superior to training on low-density, high-purity conversion data. pCV outperformed even a "theoretically perfect" incremental conversion model by 6% over 27 periods, demonstrating the power of signal volume over signal purity.

The industry's focus on refining sparse conversion data is inefficient. A paradigm shift to pCV not only unlocks significant economic gains for all advertisers but also offers a path to a more private-by-default digital advertising ecosystem.

6. Resources

The simulation can be replicated with parameters of your own choosing. Here's a link to two tools you can use to create your own analysis:

- https://docs.google.com/spreadsheets/d/1yGXwvtfXx99Ky5Dp6s3-2x3KbeiKMH4g-X-tYM1TRWM/edit?gid=384577474#gid=384577474 This template Google Sheet can be used to build your own simulation using Google BigQuery.
- The visualizations and raw data output results from our simulation are available here: https://lookerstudio.google.com/u/0/reporting/736dca65-adae-40a6-aa62-f62381d0b7d3/page/p_akgpdqi0xd