International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 6, November-December 2025||

DOI:10.15662/IJRPETM.2025.0806021

From Attribution to Action: Causal Incrementality and Bandit-Based Optimization for Omnichannel Customer Acquisition in Retail Media Networks

Saurabh Mittal

North Carolina State University, USA

ABSTRACT: Retail Media Networks (RMNs) have become a central marketplace for performance marketing, yet the measurement and optimization loop remains fragile when decisions are driven by correlational signals such as last-click attribution or conventional Return on Ad Spend (ROAS). Two problems appear repeatedly in practice: (i) attribution systems that mis-allocate conversion credit across multi-touch journeys, and (ii) bidding and budgeting policies that over-invest in placements that look profitable only because organic demand is misattributed to ads. This research article proposes an integrated framework that unifies machine learning attribution, causal incrementality measurement (incremental ROAS, iROAS), and sequential decision-making via multi-armed bandits (MABs). The contribution is a closed-loop architecture in which a causal estimator produces uncertainty-aware incremental value signals at campaign and placement levels, and a contextual bandit policy allocates budget to maximize cumulative incremental profit under realistic constraints (delayed conversions, inventory coupling, advertiser objectives, and privacy). The article specifies a deployable pipeline, catalogs experimental and quasi-experimental designs suitable for RMNs and omnichannel acquisition, and provides a simulation study illustrating how iROAS-aligned bandits can outperform ROAS-driven heuristics when organic baseline and confounding are present. The discussion emphasizes validity threats (interference, spillover, novelty, and measurement leakage), governance requirements (auditability, explainability, and privacy-by-design), and a research agenda for real-time causal estimation and multi-objective bandit learning in high-dimensional retail settings.

KEYWORDS: Retail Media Networks, customer acquisition, multi-touch attribution, causal inference, incrementality, iROAS, multi-armed bandits, contextual bandits, offline policy evaluation, privacy-preserving optimization

I. INTRODUCTION

Customer acquisition has shifted from a channel-centric activity to a journey-centric optimisation problem. Brands now orchestrate search, sponsored listings, social, email, affiliates, and offline touchpoints to move customers through discovery, consideration, and purchase. Retail Media Networks (RMNs) add a distinctive layer to this journey because the retailer owns both demand signals and the transaction outcome. This creates an opportunity for precise measurement, but it also amplifies common failure modes: attribution can be distorted by organic demand and by the selective exposure of ads, and bidding systems can become optimisers of biased labels.

Two applied strands motivate this article. First, intelligent attribution models using machine learning provide richer representations of journeys and can move beyond rigid rules such as last-touch crediting, while still requiring disciplined handling of bias, interpretability, and privacy [1]. Second, incrementality metrics such as iROAS reframe ad effectiveness as a causal question: how much net-new value was created by advertising, relative to what would have happened anyway [2], [3]. When iROAS is used as the reward signal in a multi-armed bandit, bidding becomes a sequential decision problem grounded in incremental value rather than correlated outcomes [2], [10], [12].

The goal of this research article is to connect these components into a single, publishable blueprint. The proposed framework aligns multi-touch attribution with causal identification strategies from econometrics and causal machine learning [6]-[9], then operationalises the resulting incremental signals through uncertainty-aware contextual bandits [10]-[12]. The result is an attribution-to-action loop in which measurement quality is not an afterthought, but a first-class input to optimisation.

Contributions.

- A unified problem formulation that connects multi-touch attribution, causal incrementality, and sequential allocation under operational constraints.
- A modular closed-loop architecture separating ingestion, privacy, causal estimation, and policy optimisation, enabling independent validation and auditability.
- A methodological catalogue of experimental and quasi-experimental designs for iROAS estimation in RMNs, including variance reduction and interference diagnostics.
- A policy layer design for iROAS-based contextual bandits addressing delayed feedback, non-stationarity, and multi-objective or constrained optimisation.
- A simulation study demonstrating how iROAS-aligned bandits dominate ROAS-aligned heuristics under organic baseline and selection bias.

II. BACKGROUND AND RELATED WORK

The proposed framework sits at the intersection of attribution modelling, causal inference for advertising, incrementality metrics for budget allocation, and bandit learning for adaptive decision-making. This section clarifies what is conceptually established, where practical constraints dominate, and which research gaps limit production-grade adoption in RMNs.

2.1. Attribution modelling in multi-touch journeys

Rule-based attribution assigns conversion credit using fixed heuristics such as first-touch, last-touch, or linear splits. These rules are interpretable and low-cost but cannot represent interaction effects across channels, saturation, or diminishing returns. Probabilistic approaches model journeys as stochastic processes, for example using Markov chains to estimate removal effects. They improve realism but depend on structural assumptions and may struggle with sparse or high-dimensional paths.

Machine learning attribution extends the feature space to include context, user intent proxies, and campaign metadata, enabling non-linear response surfaces and cohort-specific effects [1]. In practice, supervised learning predicts conversion probability from journey representations, while unsupervised learning segments journeys into behavioural cohorts. However, predictive attribution scores are not automatically causal. If exposure is correlated with latent intent, a model can be an accurate predictor while still assigning misleading causal credit. This difference between prediction and causation motivates explicit causal identification strategies for any attribution score used in budget allocation [6]-[9].

2.2. Causal inference and incrementality in advertising

Advertising measurement is difficult because counterfactual outcomes are unobserved. Randomised experiments address this by assigning exposure independently of latent purchase propensity, but experiments can be expensive, can suffer from interference, and cannot be run for every micro-decision. Applied systems therefore combine design-based identification with model-based adjustment, using tools such as propensity scores and doubly robust estimators [15], [33], [35].

Incremental ROAS (iROAS) operationalises the causal effect as incremental revenue divided by incremental cost. Compared with traditional ROAS, it explicitly removes organic baseline and therefore corrects the common inflation seen in brand-heavy or high-intent settings [2], [3]. iROAS is attractive in RMNs because the retailer observes both exposure and purchase, enabling platform-native holdouts and quasi-experiments. Yet iROAS is noisy and sensitive to design choices, making uncertainty estimation and variance reduction central to trustworthy optimisation.

2.3. Bandits and sequential decision-making for bidding

Multi-armed bandits formalise exploration-exploitation trade-offs in repeated allocation problems [10], [12]. In retail media, arms can correspond to placements, keywords, products, creatives, or bid multipliers. Thompson sampling and upper confidence bound methods incorporate uncertainty and are therefore commonly used in noisy reward settings [10], [11]. Contextual bandits learn action values as a function of context, which is important when incremental lift varies by query intent, seasonality, and inventory.

Bandits are adaptive, but they are sensitive to reward specification. If the reward embeds organic baseline, a bandit can converge on actions that maximise correlation, not incremental value. This motivates using iROAS, or other causal lift signals, as rewards, and maintaining an explicit separation between estimation and policy learning to avoid reward hacking.

2.4. Privacy, governance, and operational constraints

RMNs operate under increasing privacy constraints, including restrictions on identifiers and regulations governing consent and automated decision-making [24]. Practical systems must be designed for aggregation, minimisation, and auditability. Privacy-by-design influences measurement choices, often favouring aggregate experiments and cohort-level estimators over user-level tracking. Governance requirements, including transparency to advertisers and internal audit trails, shape which optimisation policies are deployable.

Explainability methods such as SHAP and LIME can help translate complex models into human-interpretable drivers [22], [23]. For RMNs, explainability must go beyond generic feature importance and connect directly to incrementality and constraints, for example explaining why a policy reduced bids due to low incremental lift with narrow confidence bounds.

Table 1. Comparative overview of attribution approaches in customer acquisition

Approach	Typical assumption	Strengths	Common limitations
Rule-based (first-touch, last-touch, linear)	Credit is fixed by position in the journey	Simple, easy to explain, low compute	Systematically mis-credits assist channels, ignores interactions and saturation
Probabilistic (Markov, hazard/survival)	Journeys follow a structured stochastic process	Captures sequence effects, more data-driven than rules	Relies on structural assumptions, limited with non-linear context and sparse paths
Predictive ML attribution	Conversion propensity is learnable from journey and context features	Flexible non-linear modelling, cohort effects, scalable scoring	Can be non-causal under selection bias, often opaque, needs large labelled data
Causal ML and experimentation	Counterfactual lift is identifiable via design or assumptions	Supports incrementality metrics, consistent with decision-making	Design complexity, variance and interference, requires governance and careful rollout

III. PROBLEM FORMULATION

Consider an RMN where, at each decision opportunity t, the platform chooses an action a_t from an action set A (for example a placement-bid pair or a creative eligibility configuration). The decision is made under context x_t summarising environment state, including product and inventory information, campaign constraints, and coarse customer features available under privacy constraints.

Let Y_t denote downstream value such as revenue, profit, or customer lifetime value. Let C_t denote the incremental cost incurred by taking action a_t. Conventional optimisation uses observed proxies such as last-click revenue or total attributed sales. The proposed framework instead defines reward using incremental value:

$$r_t = \mathrm{E}[Y_t(1) - Y_t(0) \mid a_t, x_t] - C_t$$

where $Y_t(1)$ and $Y_t(0)$) represent potential outcomes with and without the advertising exposure induced by action a_t [7]-[9]. Incremental ROAS is then

$iROAS_t = \mathrm{E}[Y_t(1) - Y_t(0) \mid a_t, x_t] / C_t$ when $C_t > 0$.

The objective is to learn a policy pi(a|x) that maximises cumulative expected incremental profit over a horizon T, subject to constraints such as budget caps, pacing, inventory coupling, share-of-voice requirements, and delayed feedback. The key complication is that incremental reward is not observed directly and must be estimated with uncertainty. The policy layer should therefore consume a distribution over incremental value, not only a point estimate.

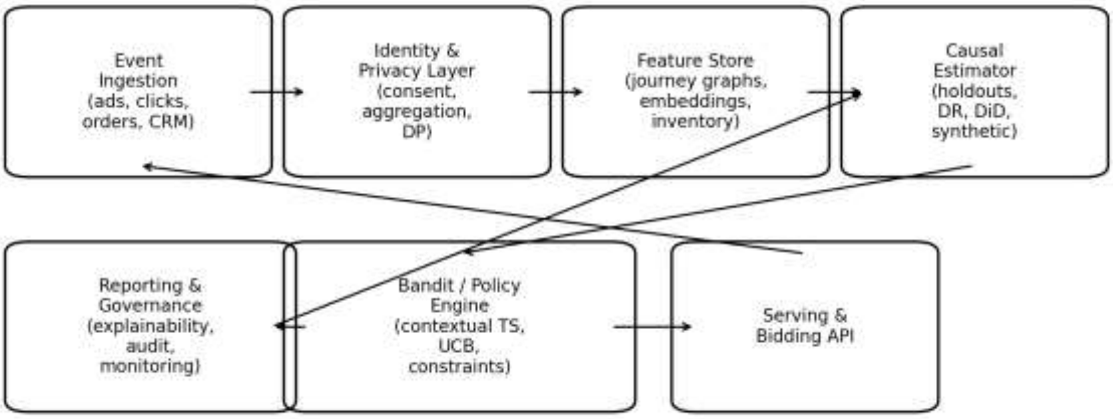
Event Ingestion (ads, clicks, orders, CRM)
Identity & Privacy Layer (consent, aggregation, DP)
Feature Store (journey graphs, embeddings, inventory)
Causal Estimator (holdouts, DR, DiD, synthetic)
Reporting & Governance (explainability, audit, monitoring)
Bandit / Policy Engine (contextual TS, UCB, constraints)
Serving & Bidding API

Figure 1. Closed-loop architecture linking attribution, causal incrementality, and bandit-based optimisation in RMNs.

IV. METHODOLOGY: A CLOSED-LOOP ATTRIBUTION-TO-ACTION FRAMEWORK

The framework is organised into four modules: (i) journey representation and feature engineering, (ii) causal incrementality estimation, (iii) policy optimisation via bandits, and (iv) governance and monitoring. The modules are separated by explicit interfaces, which supports independent validation and reduces the risk that a single modelling error propagates silently through the full system.

4.1. Journey representation and feature engineering

Omnichannel journeys can be represented at multiple resolutions. A low-resolution representation aggregates events into channel-level counts, recency, and frequency features. A higher-resolution representation uses sequence models over time-stamped touchpoints, including dwell time, query intent proxies, and campaign metadata. A graph representation treats channels and touchpoints as nodes with edges capturing transitions and time gaps.

Feature engineering should be aligned with the causal question. Variables affected by treatment can induce post-treatment bias if included in estimators. For example, ad-driven page views can be outcomes of exposure and should not be used as pre-treatment covariates. In RMNs, inventory, pricing, and delivery time signals act as confounders and must be included as context for both estimation and policy learning.

4.2. Causal incrementality estimation

Incrementality estimation should prioritise design-based identification where feasible. RMNs can enable platform-native holdouts by randomly withholding eligible traffic from ads at the user, request, or geo level. Where randomisation is limited, quasi-experimental strategies can be used, including difference-in-differences with matched controls [32], synthetic control methods [31], and doubly robust estimators combining propensity scores with outcome models [15], [33].

A practical estimator stack often starts with a design layer and then chooses the lightest estimator consistent with the design. In a user-level holdout, a simple difference in means may be sufficient, augmented with regression adjustment to reduce variance. In geo-level tests where clusters differ, synthetic control or matched DiD provides stronger counterfactual construction. In fully observational settings, doubly robust estimators provide protection against misspecification because they are consistent if either the propensity model or the outcome model is correctly specified [15].

Variance reduction is critical because iROAS is noisier than click-based metrics. Covariate adjustment, pre-period controls, and hierarchical pooling can stabilise estimates for sparse campaigns. Bayesian partial pooling can share statistical strength across products within a category, reducing overreaction to short-term noise. For decision making, the output should include uncertainty, such as credible intervals or bootstrap confidence intervals, enabling conservative optimisation and formal risk bounds.

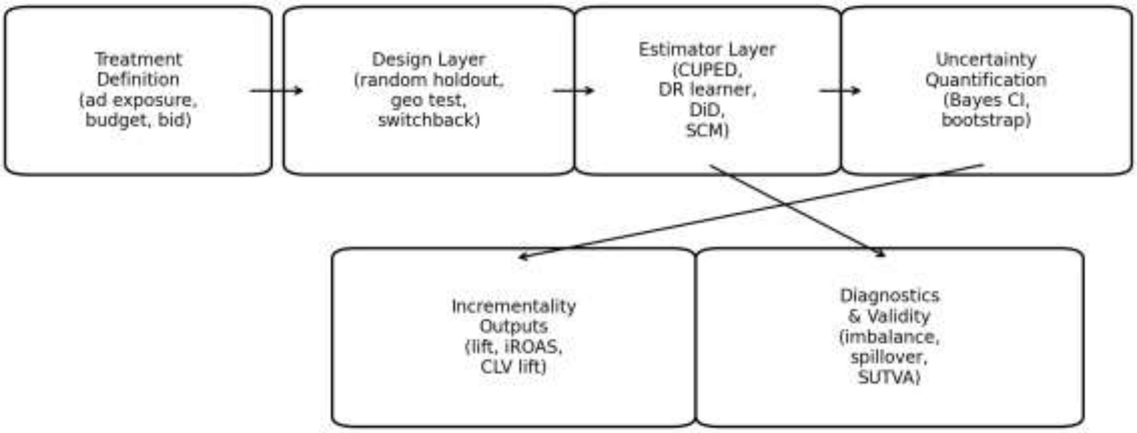
Treatment Definition (ad exposure, budget, bid)
Design Layer (random holdout, geo test, switchback)
Estimator Layer (CUPED, DR learner, DiD, SCM)
Uncertainty Quantification (Bayes CI, bootstrap)
Incrementality Outputs (lift, iROAS, CLV lift)
Diagnostics & Validity (imbalance, spillover, SUTVA)

Figure 2. Incrementality module for estimating lift and iROAS with uncertainty and validity diagnostics.

4.3. Attribution models as causal enablers

Attribution models remain valuable in a causal system, but their role shifts. Instead of treating predictive attribution scores as ground truth, they guide experimentation and define strata for more efficient estimation. First, journey embeddings and segmentations can define homogeneous strata, enabling more precise uplift estimation within cohorts. Second, predictive propensity models can support doubly robust estimation, reducing bias when randomisation is incomplete [15], [33]. Third, attribution outputs can be used as diagnostic signals, highlighting channels where predictive contributions disagree with experimental lift, which can indicate measurement leakage, targeting confounding, or interference.

To avoid circularity, causal estimators should treat attribution features as inputs only when they are pre-treatment and not derived from the current policy outcome. Separating training windows for attribution and lift estimation reduces feedback-induced bias. In addition, when attribution methods such as Shapley-based decompositions are used, they should be framed as allocation tools over a predictive model, rather than as causal proof of channel contribution [30].

4.4. Bandit policy optimisation with iROAS rewards

The policy layer consumes incremental reward estimates and allocates budget across arms. In low-dimensional settings, epsilon-greedy, UCB, and Thompson sampling can be effective [10]-[12]. In RMNs, context matters: the incremental effect of a placement depends on query intent, product category, seasonality, and inventory. Contextual bandits address this by learning a mapping from x_t to expected incremental reward.

Delayed feedback is a defining challenge. Purchases can occur hours or days after exposure, creating biased learning if the policy updates too early. Practical systems use delay models to produce provisional rewards that are corrected when outcomes mature, or use switchback designs to observe stable incremental effects over defined windows. Non-stationarity arises from promotions, price changes, and competitor behaviour. Discounted updates, sliding windows, and change-point detection can prevent over-commitment to stale estimates.

Constraints and multi-objective goals are central. Advertisers seek a balance across incremental profit, new-to-brand acquisition, pacing, and fairness constraints. Constrained bandits incorporate these requirements via Lagrangian multipliers or constrained optimisation, and risk-aware variants optimise conservative objectives such as lower confidence bounds of iROAS. This aligns with satisficing and risk-sensitive formulations that avoid chasing high-variance arms [28].

Table 2. ROAS vs iROAS in retail advertising decision-making

Dimension	Traditional ROAS	Incremental ROAS (iROAS)
Target quantity	Total attributed revenue / spend	Causal lift in revenue / spend
Organic baseline	Included, can inflate signal	Explicitly removed via counterfactual
Bias sensitivity	High under targeting and intent confounding	Lower when identified with experiments or robust estimators
Use for optimisation	Can reward correlated placements	Aligns optimisation with net-new value
Operational complexity	Low	Higher (design, estimation, uncertainty)

4.5. Offline policy evaluation and counterfactual safety

Before online deployment, candidate policies should be evaluated offline using logged feedback. Inverse propensity scoring estimates the expected value of a target policy by re-weighting outcomes by the probability that the logging policy chose the same action [14]. Doubly robust methods reduce variance and improve robustness by combining a direct reward model with propensity weighting [15]. Offline evaluation is not a substitute for online experiments, but it provides a safety layer to eliminate policies that are clearly inferior or unstable.

For RMNs, offline evaluation must address support and exploration. If the logging policy never explores certain placements or bid levels, the target policy value is not identifiable. This motivates purposeful exploration in logging policies, or the use of policy mixing, so that offline estimators have adequate coverage. Offline evaluation should also include stress tests for reward noise, delayed conversions, and inventory coupling, because these factors can cause online divergence even when offline estimates appear favourable.

4.6. Data schema, metrics, and instrumentation

Publishing-grade claims require clear definitions of data and metrics. RMN data typically includes impression logs (timestamp, placement, eligibility, bid, winning price), interaction logs (clicks, add-to-cart, detail views), and transaction logs (orders, returns, cancellations). For omnichannel measurement, offline purchase and loyalty signals can be integrated at aggregated or consented levels. A consistent identity key is not always available, so the system should support multiple join strategies: deterministic joins where consented identifiers exist, probabilistic matching where allowed, and cohort-level aggregation when user linkage is restricted.

Metric definitions should distinguish the measurement layer from the optimisation layer. The measurement layer estimates incremental lift in outcomes such as revenue, profit, or new-to-brand conversion. The optimisation layer consumes these estimates to make decisions, but should also track guardrails: cost per incremental acquisition, user experience measures (for example, ad load), retailer constraints such as inventory availability, and advertiser constraints such as daily pacing. Instrumentation should also include policy logging fields (policy version, exploration probability, constraint status) because these are prerequisites for offline policy evaluation and auditability.

Finally, deduplication logic must be explicit. In omnichannel settings, conversions can be counted multiple times across tracking systems. Publishing-ready systems define a single source of truth for outcomes, with reconciliation rules, and quantify how sensitive iROAS is to alternative outcome definitions.

4.7. Governance, explainability, and audit trails

Because policies autonomously reallocate budgets, governance is a requirement. A production system should record which policy version made each decision, which incremental estimate and uncertainty interval were used, and which constraints were active. Advertiser-facing reporting should distinguish incremental value from attributed value and present uncertainty.

Explainability should focus on actionable drivers. A policy explanation can report that a placement was deprioritised because incremental lift was statistically indistinguishable from zero in a holdout, or because uncertainty exceeded a risk threshold. Feature-based explanations such as SHAP and LIME can support these narratives, but should not replace causal diagnostics [22], [23].

Privacy controls should be embedded into each layer, including data minimisation, aggregation thresholds, and differential privacy where appropriate [25]. Where user-level data is used for estimation, outputs can still be aggregated at campaign level to reduce re-identification risk, and privacy-preserving training mechanisms such as federated learning can further reduce exposure [26].

V. EXPERIMENTAL DESIGN AND EVALUATION PROTOCOL

A publishing-ready optimisation framework must specify how performance is evaluated, not only how it is optimised. Evaluation is divided into offline policy evaluation, controlled online experimentation, and ongoing monitoring.

Offline evaluation provides rapid iteration and safety checks, using logged data and counterfactual estimators [14], [15]. However, credibility requires sufficient exploration and stable logging propensities. Online evaluation remains the standard for validating decision impact. Platform-native holdouts can compare an iROAS-based bandit policy against baselines such as rule-based bidding or ROAS-optimised heuristics. Because bandits are adaptive, evaluation should use designs that account for time variation and interference, including switchback experiments, cluster randomisation, or geo-level tests.

Monitoring bridges evaluation and operations. Key metrics include incremental profit, iROAS stability, policy entropy (degree of exploration), budget pacing, and guardrails such as user experience and advertiser fairness.

Table 3. Practical evaluation checklist for iROAS-driven bandit systems

Stage	Checks	Failure modes if ignored
Data quality	Latency, missingness, deduplication, join integrity	Spurious lift, unstable policy updates
Design validity	Randomisation integrity, spillover, SUTVA risks	Biased iROAS, policy converges on artefacts
Estimator diagnostics	Balance, overlap, sensitivity to model choice	Hidden confounding, overconfident intervals
Policy safety	Budget caps, risk bounds, exploration limits	Overspend, advertiser harm, runaway feedback loops
Reporting	Incremental vs attributed separation, uncertainty, audit trail	Loss of trust, disputes, compliance risk

5.1. Statistical power, sequential monitoring, and guardrails

RMN experiments often face low signal-to-noise ratios, especially when incremental effects are small relative to baseline conversion. Power analysis should therefore be treated as part of system design, not as a one-off statistical step. At the campaign level, power depends on traffic volume, baseline conversion rate, expected lift, variance of revenue, and the holdout fraction. In practice, stratification and covariate adjustment can materially reduce variance by accounting for stable pre-treatment signals.

Because bandit policies adapt over time, naive frequentist inference can be invalid if stopping or monitoring decisions depend on observed outcomes. A practical approach is to separate monitoring for safety from final inference for decision making. Safety monitoring can use sequential boundaries to halt policies that violate guardrails (for example, rapid spend increases with collapsing iROAS). Final inference can be performed on pre-specified windows using robust estimators and correction for multiple looks.

Guardrails should be multi-layered. At the policy level, enforce hard spend caps, pacing constraints, and minimum exploration entropy. At the marketplace level, monitor distributional effects such as exposure concentration and advertiser equity. At the customer level, track experience constraints such as ad load and relevance. Publishing-grade reporting should document each guardrail and report whether it was active during the experiment.

5.2. Practical threats to validity

Interference and spillover are common in RMNs. Ads can shift demand across products, brands, and channels, violating independence assumptions. Cluster designs, inventory-aware randomisation, and interference diagnostics should be used when spillovers are material.

Novelty and learning effects can inflate early lift for new formats or placements. Policies that react to novelty without modelling decay can overspend. Time-aware estimators, holdout refresh strategies, and change-point detection help mitigate this.

Measurement leakage occurs when conversion tracking is influenced by the policy itself, for example through changes in event logging or eligibility. Independent logging validation and counterfactual checks are required before concluding that iROAS improvements reflect true lift.

VI. SIMULATION STUDY

To illustrate the consequences of reward specification, a simulation compares two policies in a simplified RMN environment. The environment contains three placements with different organic baseline demand, different true incremental lift, and different costs. A ROAS-driven greedy policy selects the placement with the largest estimated total-revenue-per-cost, which is the analogue of optimising attributed sales. An iROAS-driven Thompson sampling policy instead treats incremental lift per cost as the decision criterion and updates its posterior using noisy lift observations [10], [11].

The simulation is intentionally simple but captures a pervasive condition: organic baseline can be large relative to incremental lift and can differ across placements. In such settings, optimising total revenue selects placements with high baseline demand even if they generate little incremental lift. This mechanism is consistent with empirical observations that ROAS can be inflated in brand-dense contexts and can mislead budget allocation [3], [5].

Figure 3 reports cumulative incremental profit over 2,000 decision steps. The iROAS-aligned policy converges toward the placement with the highest incremental lift per cost and achieves substantially higher cumulative incremental profit. The ROAS-aligned policy over-allocates to the placement with the highest baseline demand and therefore leaves incremental value on the table.

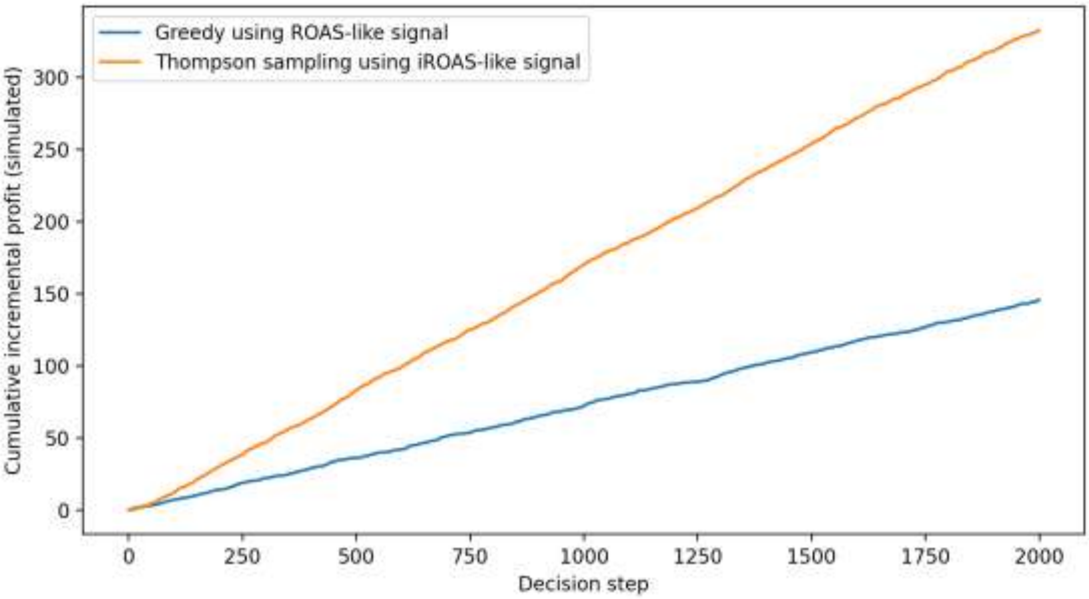

Figure 3. Simulation showing cumulative incremental profit: iROAS-aligned Thompson sampling vs ROAS-aligned greedy selection.

6.1. Interpretation and sensitivity

The simulation outcome is robust to moderate parameter changes because it reflects a structural mechanism: when observed total revenue mixes organic baseline with incremental lift, an optimiser can rationally chase baseline. In practice, this appears when high-intent queries, brand terms, or navigational traffic are heavily advertised. Such contexts often yield strong attributed ROAS but limited incremental value.

Sensitivity analysis in real systems should include: (i) varying holdout sizes to quantify the variance-lift trade-off, (ii) testing alternative quasi-experimental estimators to identify model dependence, and (iii) stress-testing delayed reward handling and non-stationarity. When policy decisions change substantially under minor estimator variations, governance should require more conservative bidding until stability improves.

VII. DISCUSSION AND PRACTICAL IMPLICATIONS

Optimisation is only as good as the reward. If a system is rewarded for attributed value, it will learn to maximise attribution, including any systematic bias in the measurement layer. Using iROAS shifts the optimisation target to a causal quantity, but it increases reliance on design validity and estimator quality.

For practitioners, the most important operational implication is separation of concerns. The causal module must be treated as measurement infrastructure, with its own quality gates and diagnostics, rather than as an internal signal generator for the policy. Similarly, the policy module must be treated as a controlled product with rollout procedures, audit trails, and well-defined failure modes.

For advertisers, the framework implies a reporting shift. Incremental metrics should be presented alongside attributed metrics, with uncertainty. Where incremental and attributed metrics disagree, the system should explicitly explain likely causes, such as baseline demand, targeting confounding, spillover, or delay. This improves trust and reduces disputes.

7.1. Organisational adoption and human-AI collaboration

Adoption depends on interpretability and decision rights. RMN teams often operate with legacy KPIs anchored in attributed ROAS. Moving to iROAS-based bidding requires training, dashboards that separate baseline from lift, and governance processes defining when automated decisions can override manual controls. Human-in-the-loop workflows are particularly valuable during cold-start phases, when uncertainty is high.

Explainability is most effective when framed around incremental decision drivers: which cohorts drove lift, which placements were deprioritised due to low or uncertain lift, and which constraints shaped the final allocation. Post-hoc feature explanations should support, not replace, these causal narratives.

VIII. FUTURE RESEARCH DIRECTIONS

Several directions are high leverage for moving from batch incrementality to real-time causal optimisation.

Real-time causal estimation. RMNs increasingly demand low-latency lift estimates. Hybrid designs that combine sparse experiments with continuous observational updates, and streaming-friendly doubly robust estimators, are promising.

High-dimensional action spaces. Decisions span placements, keywords, products, and creatives. Deep contextual bandits and Bayesian neural bandits can model non-linear reward surfaces, but they require uncertainty calibration and stability constraints.

Delayed and multi-touch credit assignment. Combining bandits with sequential delay models, and integrating multi-touch incrementality across journeys, remains challenging.

Privacy-preserving optimisation. Differential privacy, federated learning, and secure aggregation can reduce exposure risk, but their impact on bandit learning dynamics and causal estimation accuracy remains under-studied [25], [26].

Multi-objective optimisation. Real campaigns optimise not only incremental profit but also new customer acquisition, category growth, and brand equity. Multi-objective bandits and constrained optimisation are required to move beyond single-metric tuning.

IX. CONCLUSION

This article presented a publishing-ready framework connecting omnichannel attribution modelling with causal incrementality measurement and bandit-based decision making in Retail Media Networks. Attribution becomes operationally valuable when it supports causal estimation, and sequential optimisation becomes trustworthy when it uses incremental signals as rewards and is continuously validated through experimentation.

By aligning optimisation with iROAS rather than attributed ROAS, RMNs can reduce wasted spend on placements that harvest organic demand and move toward accountable, value-based automation. The remaining challenges, especially real-time estimation, interference, and privacy-preserving learning, define a clear research agenda for next-generation retail media measurement systems.

REFERENCES

[1] Intelligent Customer Acquisition Modeling via Campaign Channel Attribution: A Machine Learning Approach, International Journal of Computer Engineering and Science in Emerging Technologies (IJCESEN), Vol. 11 No. 4 (2025)
[2] iROAS-Based Dynamic Bidding Strategy Using Multi-Armed Bandits in Retail Media Networks, Contemporary Advances in Network Applications (CANA), Vol. 32 No. 1s (2025)
[3] R. A. Lewis and J. M. Rao, "The unfavorable economics of measuring the returns to advertising," The Quarterly Journal of Economics, vol. 130, no. 4, pp. 1941-1973, 2015.
[4] B. Dalessandro, C. Perlich, O. Stitelman, and F. Provost, "Causally motivated attribution for online advertising," in Proc. 6th Int. Workshop on Data Mining for Online Advertising and Internet Economy, 2012, pp. 1-9.
[5] L. Bottou et al., "Counterfactual reasoning and learning systems: The example of computational advertising," Journal of Machine Learning Research, vol. 14, no. 1, pp. 3207-3260, 2013.
[6] S. Athey and G. W. Imbens, "The state of applied econometrics: Causality and policy evaluation," Journal of Economic Perspectives, vol. 31, no. 2, pp. 3-32, 2017.
[7] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd ed. Cambridge, UK: Cambridge University Press, 2009.
[8] M. A. Hernán and J. M. Robins, Causal Inference: What If. Boca Raton, FL, USA: Chapman & Hall/CRC, 2020.
[9] G. W. Imbens and D. B. Rubin, Causal Inference for Statistics, Social, and Biomedical Sciences. Cambridge, UK: Cambridge University Press, 2015.
[10] S. L. Scott, "A modern Bayesian look at the multi-armed bandit," Applied Stochastic Models in Business and Industry, vol. 26, no. 6, pp. 639-658, 2010.
[11] O. Chapelle and L. Li, "An empirical evaluation of Thompson sampling," in Advances in Neural Information Processing Systems, vol. 24, 2011.
[12] T. Lattimore and C. Szepesvári, Bandit Algorithms. Cambridge, UK: Cambridge University Press, 2020.
[13] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
[14] A. Swaminathan and T. Joachims, "Counterfactual risk minimization: Learning from logged bandit feedback," in Proc. 32nd Int. Conf. on Machine Learning (ICML), 2015.
[15] M. Dudík, J. Langford, and L. Li, "Doubly robust policy evaluation and learning," in Proc. 28th Int. Conf. on Machine Learning (ICML), 2011.
[16] S. Wager and S. Athey, "Estimation and inference of heterogeneous treatment effects using random forests," Journal of the American Statistical Association, vol. 113, no. 523, pp. 1228-1242, 2018.
[17] S. Athey, J. Tibshirani, and S. Wager, "Generalized random forests," Annals of Statistics, vol. 47, no. 2, pp. 1148-1178, 2019.
[18] F. Johansson, U. Shalit, and D. Sontag, "Learning representations for counterfactual inference," in Proc. 33rd Int. Conf. on Machine Learning (ICML), 2016, pp. 3020-3029.
[19] S. Rojas-Carulla, B. Schölkopf, R. Turner, and J. Peters, "Invariant models for causal transfer learning," Journal of Machine Learning Research, vol. 19, no. 36, pp. 1-34, 2018.
[20] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in Proc. 21st Int. Conf. on World Wide Web (WWW), 2012, pp. 519-528.
[21] H. B. McMahan et al., "Ad click prediction: A view from the trenches," in Proc. 19th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD), 2013, pp. 1222-1230.
[22] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. 31st Conf. on Neural Information Processing Systems (NeurIPS), 2017.

[23] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD), 2016, pp. 1135-1144.
[24] Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation), 2016.
[25] C. Dwork and A. Roth, The Algorithmic Foundations of Differential Privacy. Hanover, MA, USA: Now Publishers, 2014.
[26] Z. Li, V. Sharma, and S. P. Mohanty, "Preserving data privacy via federated learning: Challenges and solutions," IEEE Consumer Electronics Magazine, vol. 9, no. 3, pp. 8-16, 2020.
[27] L. Li, K. Jamieson, A. Rostamizadeh, E. Gonina, E. Ben-Tzur, M. Hardt, B. Recht, and A. Talwalkar, "A system for massively parallel hyperparameter tuning," Proc. of Machine Learning and Systems, vol. 2, pp. 230-246, 2020.
[28] P. Reverdy, V. Srivastava, and N. E. Leonard, "Satisficing in multi-armed bandit problems," IEEE Transactions on Automatic Control, vol. 62, no. 8, pp. 3788-3803, 2016.
[29] M. Mierswa and M. Wurst, "Efficient case-based feature construction," in Proc. European Conf. on Machine Learning, 2005, pp. 641-648.
[30] L. S. Shapley, "A value for n-person games," in Contributions to the Theory of Games, vol. 2, H. W. Kuhn and A. W. Tucker, Eds. Princeton, NJ, USA: Princeton University Press, 1953, pp. 307-317.
[31] A. Abadie, A. Diamond, and J. Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," Journal of the American Statistical Association, vol. 105, no. 490, pp. 493-505, 2010.
[32] J. D. Angrist and J.-S. Pischke, Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ, USA: Princeton University Press, 2009.
[33] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," Biometrika, vol. 70, no. 1, pp. 41-55, 1983.
[34] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, pp. 529-533, 2015.
[35] K. Jamieson and A. Talwalkar, "Non-stochastic best arm identification and hyperparameter optimization," in Proc. 19th Int. Conf. on Artificial Intelligence and Statistics (AISTATS), 2016.