A Unified Approach to Interpretable Causal Root Cause Attribution

Jing Zhou
Amazon
Tokyo, Japan
zhoujj@amazon.co.jp

Dominik Janzing
Amazon
Tuebingen, Germany
janzind@amazon.de

Sepp Tsang
Amazon
Tokyo, Japan
tsepp@amazon.co.jp

Patrick Blöbaum
Amazon
Santa Clara, United States
bloebp@amazon.com

Marco Visentini Scarzanella
Amazon
Tokyo, Japan
marcovs@amazon.co.jp

Abstract

Understanding why marketplace metrics change is a central problem in two-sided marketplace optimization. We study root cause attribution for metric changes in complex e-commerce systems, focusing on trade-offs between interpretability, efficiency, and causal validity. As a starting point, we extend a metric-decomposition method into a recursive metric-tree framework for multi-level root cause analysis, but this relies on independence and decomposability assumptions that miss complex causal dependencies. In contrast, graphical causal models (GCMs) relax these assumptions and improve causal validity, at the cost of interpretability, higher computational and data demands, and potential attribution target misalignment. Through real-world applications, mathematical proofs, and simulations, we characterize the fundamental sources of these trade-offs. Guided by these insights, we propose a unified, causally informed attribution approach that integrates structural causal information into the metric-tree decomposition framework and corrects key sources of misalignment in GCM-based causal attributions, substantially improving causal validity while preserving interpretability and fast computation. Analytical proofs and simulations demonstrate that the proposed approach produces more accurate root cause attributions, and we also present a real-world application.

CCS Concepts

• **Information systems → Data mining**; • **Computing methodologies → Causal reasoning and diagnostics**; • **Mathematics of computing →** *Distribution functions.*

Keywords

root cause attribution, causal inference, graphical causal models, distribution change, metric-tree change decomposition (MTCD), interpretability, unbiased estimator

1 Introduction

Two-sided marketplaces are operated through tightly coupled metrics spanning both the demand side (e.g., traffic, conversion, engagement) and the supply side (e.g., selection, price, availability). In practice, a recurring and high-stakes task on these platforms is understanding why key performance metrics change: for example, why did revenue drop this month, was the change driven mainly by traffic, pricing, or selection, and which levers should be prioritized next? These are not rare anomaly cases, but routine operational questions that directly affect growth, pricing, and seller-side strategy in large-scale online platforms.

A natural way to formalize such questions is as root cause attribution for metric changes across interacting marketplace variables. Prior work on root cause analysis focuses primarily on anomaly or outlier detection [1, 13, 17, 20, 24, 27]. However, a more relevant work [5] framed root causing as a distribution change problem and used graphical causal models (GCMs, [19]) to attribute distribution changes to local causal mechanisms. They achieved this by factorizing the joint distribution into causal mechanisms and using Shapley values [14, 22] to quantify each mechanism's contribution to the overall shift. But this approach suffers from interpretability, computational cost, and data demands, making it impractical for real-time marketplace monitoring. There is also a long-standing decomposition literature that attributes differences or changes in aggregate quantities to underlying factors [2, 3, 18, 23]. However, these methods were not designed for high-frequency metrics monitoring in modern online systems, and hence, not as interpretable. In parallel, metric decomposition methods, rarely documented in academia but widely used in industry, offer fast and highly interpretable decompositions of metric changes, yet still fall short for complex metric systems. Additionally, because they are grounded in analytics and metric definitions rather than explicit modeling, they typically ignore causal dependencies. In this work, we drew on these two representative lines of work (causal distribution-change methods and industry-style metric change decomposition) and developed an attribution method that is interpretable to platform operators, scalable to large marketplace metric systems (130+ metrics in our deployment), and grounded in a principled causal framework. In summary, we make the following contributions:

- We formalize an industry-style metric change decomposition method and extend it into a recursive metric-tree framework for multi-level root cause analysis in complex marketplace metric systems (Sec. 3).
- We provide a theory- and simulation-based comparison with the GCM-based distribution-change method that clarifies their trade-offs in interpretability, scalability, causal validity, and attribution target misalignment (Sec. 5).

KDD '26, Workshop on Two-sided Marketplace Optimization (TSMO): Search, Discovery, Matching, Pricing & Growth, Jeju, Korea

- Guided by these insights, we propose a unified, causally informed attribution approach that injects structural causal information into the metric-tree framework and introduces an unbiased estimator for mean-change attributions, improving upon GCM-based estimates while preserving interpretability and fast computation (Sec. 6). We support this with mathematical proofs and validation on simulations in Sec. 7, and a real-world marketplace application is provided in Sec. 8.

2 Background: Graphical Causal Model-Based Distribution Change Method (GCM-DC)

To identify root causes, [5] introduced a GCM-based approach by framing the root causing task as "which mechanisms are responsible for the change in the marginal distribution of a target variable". Given a correctly assumed causal graph G, a.k.a. causal DAG (causal Directed Acyclic Graph, [16]), over variables $X_1, \ldots, X_m$, the joint distribution factorizes into causal mechanisms (conditional distributions) as

$$P_X(x_1, \ldots, x_m) \; = \; \prod_{j=1}^{m} P_{X_j|\mathrm{PA}_j}(x_j \mid \mathrm{pa}_j),$$

where PA_j are the parents of X_j. A mechanism change means replacing some conditionals $P_{X_j|\mathrm{PA}_j}$ with new ones $\tilde{P}_{X_j|\mathrm{PA}_j}$, yielding a new joint distribution

$$P_X^T(x) \; = \; \prod_{j \in T} \tilde{P}_{X_j|\mathrm{PA}_j}(x_j \mid \mathrm{pa}_j) \prod_{j \notin T} P_{X_j|\mathrm{PA}_j}(x_j \mid \mathrm{pa}_j),$$

for a change set $T \subseteq \{1, \ldots, m\}$. The new marginal of a target X_k is then obtained by summing all variables but x_k:

$$\tilde{P}_{X_k}(x_k) \equiv P^T_{X_k}(x_k) \; = \; \sum_{\mathbf{x}_{\setminus k}} P^T_X(x_1, \ldots, x_n),$$

so any change in the marginal P_{X_k} to $\tilde{P}_{X_k}$ is explained by the subset of mechanisms that changed. Since attribution depends on the order in which mechanisms are changed, the method uses Shapley values to attribute the marginal change to individual mechanisms by averaging over all possible orderings of the change set T.

For root causing changes in the mean of the target variable X_k, let $\mathbb{E}_{X_k \sim P_{X_k}}[X_k]$ denote the mean of the variable X_k, writing the mean change as $\mathbb{E}_{\tilde{P}_{X_k}}[X_k] - \mathbb{E}_{P_{X_k}}[X_k]$. The Shapley value contribution of a node X_j to the mean change of target X_k is

$$\phi_j(\mathbb{E}) \; = \; \sum_{T \subseteq M \setminus \{j\}} \frac{1}{m\binom{m-1}{|T|}} C_{\mathbb{E}}(j \mid T), \tag{1}$$

where M denotes a set of m variables, and the marginal contribution $C_{\mathbb{E}}(j \mid T)$ of X_j given a change set T is defined as

$$C_{\mathbb{E}}(j \mid T) \; = \; \mathbb{E}_{X_k \sim P^{T \cup \{j\}}_{X_k}}[X_k] - \mathbb{E}_{X_k \sim P^T_{X_k}}[X_k]. \tag{2}$$

This method interprets the change in the target's mean as arising from changes in the local causal mechanisms and uses Shapley values to allocate the mean shift to individual variables in the causal graph in a principled, order-invariant way. Based on the ordering and magnitudes of the Shapley value contribution of each variable, root causes are identified.

3 Metric-Tree Change Decomposition (MTCD)

Metric decomposition is widely used across industries to explain metric changes, but existing treatments appear mostly in technical blogs [9, 10, 15, 21] and have two key limitations. First, metric and attribution types are not organized into a systematic framework. Second, existing practical treatments either operate only at a single parent–child level or cover only the simplest attribution type in a simple tree setup. These limitations make it difficult for practitioners to derive a general algorithm that propagates attributions across common metric types in a multi-level tree while preserving exact additivity to the root.

We address these limitations. We first distinguish *non-rate* metrics y (e.g., revenue, page views) from *rate* metrics $\bar{y}$ (e.g., average unit price (AUP), conversion rate(CVR)). For two periods, $t = 0$ and $t = 1$, let z_0, z_1, and $\Delta z := z_1 - z_0$ denote the baseline value, new-period value, and change of any scalar quantity z.

A *metric decomposition* writes a target metric as a deterministic function of its components,

$$y = f(\text{components}) \qquad \text{or} \qquad \bar{y} = f(\text{components}),$$

typically through sums, products, or weighted averages defined by business logic. Its corresponding *change decomposition* expresses the target metric change as

$$\Delta y = \sum_j \text{contrib}_j \qquad \text{or} \qquad \Delta \bar{y} = \sum_j \text{contrib}_j,$$

where the component contributions are interpretable and sum exactly to the observed change. Table 1 summarizes the main metric decompositions and their associated change decompositions. In Type 2 (Fig. 1), the allocation attributes the volume effect using the new-period price rather than the old-period price, reflecting a natural business intuition; in Type 4, $(\bar{y}_{k,0} - \bar{y}_0)$ centers subgroup rates around the baseline overall rate so that the share effect reflects whether higher- or lower-rate groups gain share [21].

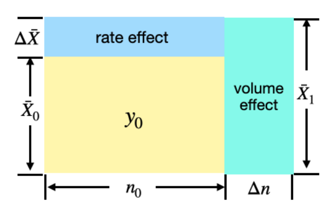

Figure 1: Type 2 metric allocation strategy

These four patterns define the local rules of our metric-tree change decomposition (MTCD) framework. We first note that the parent–child direction and root/leaf terminology in a metric tree is reversed relative to the GCM formulation. Let $\mathcal{T}$ be a metric tree with root r, where each node $v \in \mathcal{T}$ denotes a metric with baseline and new-period values $y_{v,0}$ and $y_{v,1}$, and $\Delta y_v := y_{v,1} - y_{v,0}$. Each child $c \in \mathrm{Ch}(v)$ has a decomposition type $\mathtt{type}(c) \in \{\text{Type1}, \text{Type2}, \text{Type3}, \text{Type4}\}$ linking to its parent v from Table 1 and has a dimension label $\dim(c)$. We construct the metric tree under the MECE (mutually exclusive, collectively exhaustive) principle [12, 25] to systematically link metrics and decompose target changes.

Table 1: Basic metric decompositions and their change decompositions.

Category	Metric decomposition	Change decomposition
Type 1: Non-rate(or rate), (sum)	$y = \sum_{k=1}^{K} y_k$ components are $y_1, \ldots, y_K$	$\Delta y = \sum_{k=1}^{K} \Delta y_k$, where $\Delta y_k = y_{k,1} - y_{k,0}$. Effect of changes of y_k is given by Δy_k.
Type 2: Non-rate, (product)	$y = n \cdot \bar{X}$ components are $n, \bar{X}$ e.g., revenue = units sold·AUP	Δy = vol + rate (Fig. 1), vol (volume effect of n) = $\Delta n \cdot \bar{X}_1$, rate (rate effect of $\bar{X}$) $= \Delta \bar{X} \cdot n_0$.
Type 3:Non-rate, (sum of products)	$y = \sum_{k=1}^{K} n_k \cdot \bar{X}_k$ components are $n_k, \bar{X}_k$	$\Delta y = \sum_{k=1}^{K} (\text{vol}_k + \text{rate}_k)$, vol_k (volume effect of n_k) $= \Delta n_k \cdot \bar{X}_{k,1}$, rate_k (rate effect of $\bar{X}_k$) $= \Delta \bar{X}_k \cdot n_{k,0}$.
Type 4: Rate, (weighted average)	$\bar{y} = \sum_{k=1}^{K} p_k \cdot \bar{y}_k$, where $\sum_{k=1}^{K} p_k = 1$ components are $p_k, \bar{y}_k$	$\Delta \bar{y} = \sum_{k=1}^{K} (\text{share}_k + \text{rate}_k)$, share_k (share effect of p_k) $= \Delta p_k \cdot (\bar{y}_{k,0} - \bar{y}_0)$, rate_k (rate effect of $\bar{y}_k$) $= \Delta \bar{y}_k \cdot p_{k,1}$.

The root contribution is initialized as $C_r = \Delta y_r$, and we can recursively apply the corresponding local change decomposition to each internal node from Table 1 by propagating

$$C_c = \begin{cases} \text{contrib}_c \cdot \frac{C_v}{\Delta y_v}, & \Delta y_v \neq 0, \\ 0, & \Delta y_v = 0, \end{cases} \qquad c \in \text{Ch}(v) \text{ for all } \dim(c).$$

This recursion preserves exact additivity within each decomposition dimension and allocates the target change to all descendant nodes. Non-root nodes with the largest absolute contributions are reported as root causes. Appendix A provides the pseudo-code.

For illustration, revenue can be decomposed into units sold (n) and average unit price (AUP) by Type 2, while AUP can be further decomposed by dimensions such as deal status through Type 4. Using this construction, we built a metric-tree system covering over 130 metrics, a subset of which is in Fig. 2.

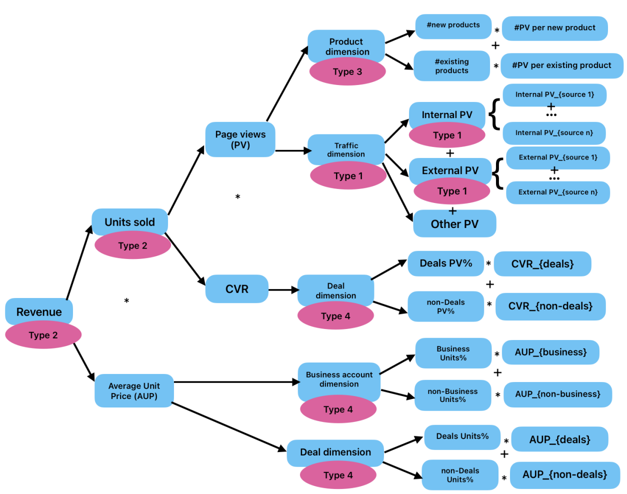

Figure 2: Complex metric tree example

4 Problem Statement

In a complex metrics system, both the GCM-DC method and the MTCD method offer clear benefits but also important limitations for interpretable and reliable root cause attribution. In particular, MTCD is simple to construct, highly interpretable, and essentially instantaneous to compute at scale. However, it has structural limitations:

(1) **Cross-branch dependencies not modeled.** The tree decomposition structure does not allow dependence among siblings or nodes of the same layer, hence not able to incorporate all causal relations: e.g., AUP may have impact on units sold.

(2) **Incomplete coverage of non-decomposable metrics.** The framework cannot cover metrics that are not decomposable. Although much of the time, we can define metrics following the MECE rule and decompose consequently, but there are exceptions.

The GCM-DC approach can fix both restrictions, but it also comes with some limitations of its own compared to MTCD:

Known limitations of GCM modeling:

(1) **Lower interpretability and lack of exact additivity in implementation.** Attributions are produced by a collection of conditional models and numerical procedures, making the method effectively a black box. Contributions typically do not sum exactly to the observed change in the target metric due to residual noise in each conditional model;

(2) **Dependence on the causal graph.** Validity hinges on the correctness of the assumed causal graph. Mis-specified edges, directions, or unmeasured confounders can make the inference unreliable, and the number of possible graphs grows exponentially (up to $2^{m(m-1)}$ edge configurations for m nodes), making misspecification hard to avoid in practice, even with falsification tests [7].

Limitations discovered in our later analyses and simulations:

(3) **Computational cost.** For realistic graphs (e.g., a 34-node system with daily data for 1 year), computing a point estimate

using GCM-DC can take minutes, and confidence intervals via resampling can take over ten minutes, making the real-time root causing impossible. For comparison, MTCD is on the order of a second;

(4) **Sample size requirements.** Reliable estimation requires substantial data as each conditional distribution needs to be estimated. Month-over-month comparisons with roughly 30 daily observations often yield high-variance and unstable attributions;

(5) **Estimand mismatch under unique causal ordering.** Even with the correct graph, GCM-DC's Shapley-based attributions diverge from the true contribution target when a unique causal ordering exists, because Shapley allocation averages over impossible orders instead;

(6) **Sensitivity to outliers.** The use of residual resampling to approximate marginal distributions [4] makes estimates sensitive to heavy-tailed residuals and outliers, which can further destabilize the results;

(7) **Limited granularity of structural changes.** Structural / Mechanism changes at a node are often summarized as a single term (e.g., overall CVR distribution is changed), whereas MTCD can expose finer-grained subgroup effects (e.g., which segment of CVR is changed).

5 Formal Comparisons of GCM-DC and Decomposition Approach

5.1 Real-world Applications

First, we illustrate the first two layers (Fig. 3(a)) for revenue change in a real use case. Based on domain knowledge, we define the causal graph in Fig. 3(b). We collected two years of daily data from a global e-commerce site and applied the GCM-DC method (using Fig. 3(b)) and MTCD (using Fig. 3(a)) to identify the root causes of year-over-year revenue changes for two vendors. Contribution values were rescaled to protect confidentiality. Because high variability made month-to-month GCM-DC results statistically insignificant, revealing its sample-size limitations, we restricted all real-world analyses to yearly comparisons.

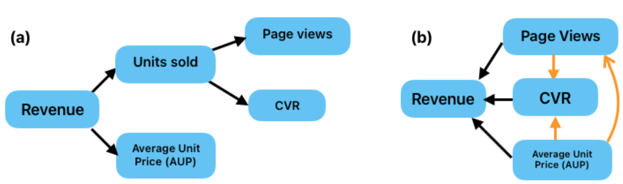

Figure 3: (a) Top-level decomposition tree; (b) Simple causal graph for revenue

For the first vendor, we see from Fig. 4 that contributions are swapped between CVR and page views under two approaches. For the second vendor (Fig. 5), MTCD identified CVR as the second root cause, while GCM-DC did not. In summary, the results between GCM-DC and MTCD differ considerably, and it remains unclear which method is more trustworthy and why. This ambiguity necessitates further investigation of: (1) theoretical difference and linkage; and (2) simulations where the ground truth is known.

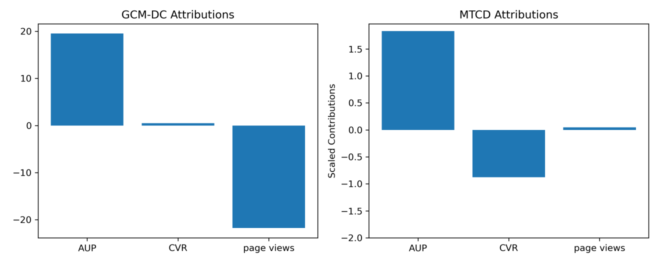

Figure 4: Vendor 1 contributions for year-over-year revenue change

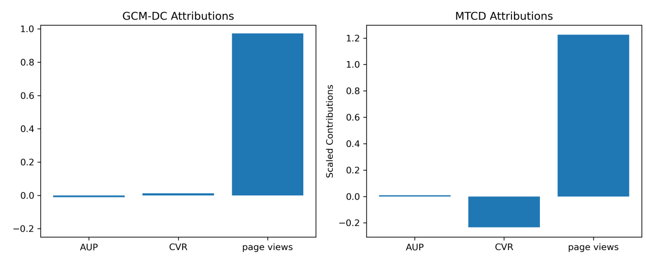

Figure 5: Vendor 2 contributions for year-over-year revenue change

5.2 Theoretical Proofs

These two methods appear fundamentally different and stem from distinct methodological traditions, yet a detailed mathematical examination reveals their underlying connections and key differences. Without loss of generality, we will separate the findings into non-rate metrics and rate metrics. As for the non-rate metric, since Type 1 change decomposition is rather trivial, and Type 3 is a combination of Type 1 and 2, we will primarily focus on Type 2 decomposition.

(I). Non-rate metric decomposition. e.g., Revenue (Rev) = units sold (n) × AUP

Differences:

(i) Different target scale (mean vs total change). Taking the same revenue example, the GCM-DC approach studies the mean change in revenue, while the decomposition-based approach allocates "overall revenue change" rather than the "mean change". However, either results can be adjusted to be on the same scale by GCM-DC's contribution × sample size or Decomposition's contribution / sample size;

(ii) GCM-DC uses Shapley allocation: calculate each node's average impact across all possible combinations of other nodes' change status, showing how much it contributes to the target metric change.

For example, contribution of units sold (n), using Eq. (1)-(2),

$$= \frac{1}{2} \cdot \text{impact of n on revenue given AUP not changed}$$
$$+ \frac{1}{2} \cdot \text{impact of n on revenue given AUP changed}$$
$$= \frac{1}{2}\left(\mathbb{E}^{\phi\cup n}_{\text{Rev}}[\text{Rev}] - \mathbb{E}^{\phi}_{\text{Rev}}[\text{Rev}]\right) + \frac{1}{2}\left(\mathbb{E}^{\text{AUP}\cup n}_{\text{Rev}}[\text{Rev}] - \mathbb{E}^{\text{AUP}}_{\text{Rev}}[\text{Rev}]\right) \tag{3}$$

vs. Decomposition uses ordered allocation: pre-specified order of change impact. In the revenue case, AUP is assumed to be changed first, units sold is changed consequently;

(iii) GCM-DC can incorporate additional causal relations among causes (which are sibling nodes in the metric tree): e.g., AUP impacts units sold by modeling units sold with a function of AUP and noise.

Underlying Connection:

Proposition 5.1. *The scale adjusted GCM-DC and decomposition approach is equivalent in expectation when causes are independent and ordered allocation (instead of Shapley) is used.*

Appendix B.1 provides the proof.

(II). Rate metric decomposition. e.g., AUP = business units%$*$AUP$_b$+ non-business units% $*$ AUP$_{nb}$, where subscripts b and nb are abbreviated for the business and non-business subgroup.

Differences:

(i) Slightly different target scale. Unlike non-rate decomposition, here GCM-DC analyzes the change of mean AUP between two periods, which is not mathematically equivalent to overall AUP change between two periods;

(ii) Shapley allocation (used by GCM-DC) differs from ordered allocation (used by Decomposition), similar to the discussion in the non-rate decomposition above.

Underlying Connection:

Proposition 5.2. *In a linear setting, GCM-DC $\approx$ decomposition for sum of share nodes when ordered allocation (instead of Shapley) is used, or the rate's distribution is not changed.*

This is proved in Appendix B.2 and simulation studies in the next section (detail in Appendix C.2 Scenario 4a). In summary, the main differences between the two methods are (a) the potential dependence among causes (i.e., sibling nodes in the metric tree) and (b) Shapley allocation vs ordered allocation.

5.3 Simulation Studies

This section verifies through simulations whether either method is able to select the correct root causes when ground truth is available for various scenarios, reveals pros and cons of each method summarized in the problem statement section, and confirms the findings from the theoretical proofs section above.

I. Non-rate metric simulation. From Appendix C.1, in summary, when there is dependence among causes, MTCD would fail as expected. Under the correct DAG, GCM-DC can successfully identify root causes in the right order even with weak causal relationships, though contribution magnitudes diverge from the true contribution (Appendix Figs. 14 & 16) under the known causal ordering, revealing the estimand misalignment problem in Sec. 4. GCM-DC, as a gold-standard causal root-causing approach, still relies on correct causal DAG assumption, and fails when the causal DAG is mis-specified (Appendix Figs. 17-18) or edges are missed (Appendix Fig. 19), underscoring the importance of correct graph assumptions. Additionally, a small sample size (e.g., 30) will make GCM-DC unreliable even under the correct DAG (Appendix Fig. 15).

II. Rate metric simulation. From Appendix C.2, it confirmed the two differences and the underlying connection stated in the above theoretical section. Moreover, given that GCM-DC models all parent nodes together with their joint child node vs. MTCD which decomposes rate metric into multiple dimensions separately in parallel (e.g., AUP in Fig. 2), we tested the hypothesis whether MTCD inflates the contribution of each node compared to GCM-DC. Under the assumption that multiple nodes are independent (e.g., subscription is independent of business account, Appendix C.2 Scenario 5), we verified that MTCD, which computes contributions in separate dimensions, can still estimate the true contributions fairly well similar to GCM-DC, hence proving that MTCD is a valid approach.

In our business case, it is usually reasonable to assume such independence among causes (i.e., sibling nodes in MTCD) under Type 4 metrics. Our focus is thus to primarily tackle the problem observed in non-rate metric, specially the Type 2 metric. However, when such independence assumption is violated, a similar strategy proposed in the next section can be applied to Type 4 decomposition.

6 Unified Root Causing Approach

Given the limitations of both approaches listed in Section 4 together with the findings of previous section, we propose a unified approach that addresses most of those challenges. It formalizes root cause attribution as causal distribution change and provides a clean bridge between metric decomposition and full causal inference. This can be directly applicable to industrial metrics monitoring systems.

Let us start with the original decomposition approach, non-rate metrics Type 2 uses the allocation in Fig. 1. The GCM-DC method incorporates both the causal relationship and the Shapley allocation. Based on the theoretical proofs section for the non-rate metric, we can conceptually represent its root causing mechanism of Eq. (3) geometrically in Fig. 6(a), where $n_1^* = \hat{f}_1(\bar{X}_0)$, $n_0^* = \hat{f}_0(\bar{X}_1)$ and $\hat{f}_0(\cdot)$

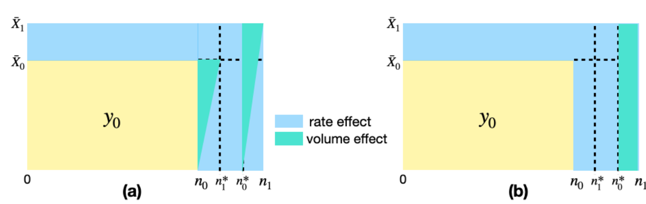

Figure 6: (a) Conceptual GCM-DC Allocation vs. (b) Proposed Allocation

and $\hat{f}_1(\cdot)$ are to be learned from old-period and new-period data respectively. This is because in the case of AUP and units sold (n), the total contribution of n comes from Eq. (3):

$$= \frac{1}{2}\left(\mathbb{E}^{\phi\cup n}_{\text{Rev}}[\text{Rev}] - \mathbb{E}^{\phi}_{\text{Rev}}[\text{Rev}]\right) + \frac{1}{2}\left(\mathbb{E}^{\text{AUP}\cup n}_{\text{Rev}}[\text{Rev}] - \mathbb{E}^{\text{AUP}}_{\text{Rev}}[\text{Rev}]\right)$$

$$\approx \frac{1}{2}\Big[\frac{1}{J}\sum_{j=1}^{J}\left(n^*_{1,j}\cdot AUP_{0,j} - n_{0,j}\cdot AUP_{0,j}\right)\Big]$$

$$+ \frac{1}{2}\Big[\frac{1}{J}\sum_{j=1}^{J}\left(n_{1,j}\cdot AUP_{1,j} - n^*_{0,j}\cdot AUP_{1,j}\right)\Big], \tag{4}$$

where j is the data granularity index, j^{th} day, for example. The two green triangles in Fig. 6(a) correspond to two bracketed terms respectively in Eq. (4) on a daily basis. Depending on the fitted two functions $\hat{f}(\cdot)$, n^*_1 and n^*_0 do not need to lie between n_1 and n_0. As a result, the estimated rate and volume effects can deviate from the original blue and green areas in Fig. 1.

Inspired by GCM-DC, we build on MTCD by incorporating causal structure but with ordered allocation inferred by the causal path, akin to asymmetric Shapley methods [8, 11]. Unlike prior work that restricts Shapley permutation weights for individual prediction explanations, our setting targets aggregate metric changes across two distributions and removes Shapley averaging entirely. When the added causal links have a unique causal order, a node's contribution is determined by its parents' status. Symmetric Shapley, by averaging over causally infeasible permutations, therefore targets a different attribution quantity than the ordered causal contribution, as confirmed in Sec. 7. We instead follow the causal path and compute each node's contribution conditional on its causes under the new-period distribution, yielding a simpler attribution aligned with the intended causal estimand.

Theorem 6.1. *For three metrics $\{y, n, \bar{X}\}$ and $y = n\cdot\bar{X}$, under the correct local causal graph of $\bar{X}$ impacts n, and both $\bar{X}$ and n impact y, if we let*

$$\hat{RC}(n|\bar{X}) = \frac{1}{J}\sum_{j=1}^{J}\left[n_{1,j} - \hat{f}_0(\bar{X}_{1,j})\right]\cdot\bar{X}_{1,j}, \tag{5}$$

where the subscript $\{1, j\}$ denotes data from the new period, $\hat{f}_0(\cdot)$ is a function estimated from the baseline period between n and $\bar{X}$, then $\hat{RC}(n|\bar{X})$ is an unbiased estimator of the true root cause contribution of n on the mean change of y.

Remark: Keeping or removing the residual term within $\hat{f}_0(\bar{X}_{1,j})$ does not change the result under the assumption of the residual has mean of 0.

The proof is provided in Appendix B.3. In the running example, the contribution of n to revenue can now be estimated by $\frac{1}{J}\sum_{j=1}^{J}\left(n_{1,j} - n^*_{0,j}\right)\cdot AUP_{1,j}$, where $n^*_{0,j} = \hat{f}_0(AUP_{1,j})$, same as the 2^{nd} bracket in Eq. (4). Conceptually, this bracketed term corresponds to the green area in Fig. 6(b) at the daily level. Thus, the contribution of n is its impact on revenue conditional on AUP's new status, isolating the pure effect of n's own mechanism change after removing AUP's influence on both n and revenue. The remainder blue area in Fig. 6(b) is attributed to AUP, through its direct effect $n_0\cdot\Delta\bar{X}$ and indirect effect $(n^*_0 - n_0)\cdot\bar{X}_1$. Unlike GCM-DC, this approach does not require n^*_1 or Shapley averaging, yielding a simpler estimator that more accurately recovers root causes, as confirmed in the simulation section below.

Under a certain assumption, we can further simplify the calculation of the proposed estimator by

Proposition 6.2. *under the assumption when there is no relational change between $\bar{X}$ and n in the new period (i.e., change is only from the intercept), if we estimate*

$$\hat{RC}(n|\bar{X}) = \frac{1}{J}\Big[n_1 - \sum_{j=1}^{J}\hat{f}_0(\bar{X}_{1,j})\Big]\cdot\bar{X}_1, \tag{6}$$

where n_1 and $\bar{X}_1$ are the aggregated metrics at the period level. We can show that the simplified $\hat{RC}(n|\bar{X})$ is also an unbiased estimator of the true root cause contribution of n on the mean change of y.

The proof is provided in Appendix B.4. Although aggregation may lead to an ill-defined estimator [28], we show that under the stated assumption, the simplified estimator in Eq. (6) remains unbiased and is computationally cheaper than Eq. (5). Its form closely matches the original decomposition term $\Delta n\cdot\bar{X}_1$, with n_0 replaced by $\sum_{j=1}^{J}\hat{f}_0(\bar{X}_{1,j})$. This allows the remaining nodes be computed directly at the aggregate level, reducing complexity from $O(Jm)$ to $O(kJ + m)$ when k causal dependence is added to an m-node tree. Regarding the proposition assumption, it is plausible when relationships among causes (i.e., sibling nodes in MTCD) are stable over limited time horizons or under relatively stable market conditions [6]. In practice, it can be checked by fitting a pooled model, e.g.,$n = \beta_0 + \beta_1\bar{X} + \beta_2 I(\text{period}) + \beta_3 I(\text{period})\bar{X} + \epsilon$, and testing $\beta_3 = 0$. If violated, we recommend the full estimator in Eq. (5), which does not require this slope stability assumption. We note that both estimators target the mean change of y, and multiplying by the sample size J yields the contribution to the total change.

The same strategy extends to other metric types. For Type 1 (sum), Δy_k can be reassigned to its causes when dependencies exist. Type 3 (sum of products) combines Types 1 and 2, using the Type 2 rule within each product term and the Type 1 rule across groups. For Type 4 (weighted average), dependent share terms across dimensions (e.g., $\Delta p_{\text{business}}$ and Δp_{deals}) can be reallocated along the assumed causal path, and if a rate affects its own share within a dimension, part of Δp_k can likewise be reassigned to the rate contribution using a Type 2-style adjustment.

In general, Table 2 lists the comparison between the proposed approach and the GCM-DC approach. In order for a causal root causing while maintaining other properties, the proposed approach clearly outperforms the existing go-to causal approach.

7 Simulations

In this section, we use the same motivating example of AUP, units sold(n) and revenue to simulate data, and allow potential interaction between AUP and n using the causal graph as in Fig. 7. In the baseline period, we start by simulating AUP from a normal distribution, simulate n using a function of AUP with a normal noise distribution, and let revenue = $n\cdot$ AUP. In the new period, we increase AUP and/or n on different scales to test how the proposed unified approach performs compared to the original decomposition approach, the GCM-DC approach, and the true root causes attributions. Simulation parameters are calibrated from real-world data to make the synthetic data realistic in scale, while retaining

Table 2: Comparison between GCM-DC and Proposed Approach

Evaluation Criteria	GCM-DC	Proposed Approach
Cross-branch dependence	Yes	Yes
Non-decomposable metrics coverage	Yes	No (however, > 95% are decomposable in practice)
Interpretability	Low	High
Causal graph correctness	entire DAG	Only added edges
Computational cost	Relatively high (minutes for 30-40 nodes)	Low (~ 1 second)
Sample size requirement	High (quarterly+)	Lower (monthly is okay)
Attribution target alignment	Misaligned in some cases	Aligned & unbiased with true target attribution
Outlier sensitivity	Higher (residual resampling)	As usual
Granularity of structural changes	Not available (e.g., $\text{CVR}_{\text{deals}}$)	Yes

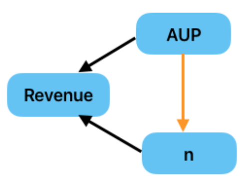

Figure 7: 3 nodes causal DAG

full control over the causal mechanism changes and ground-truth attributions. Both periods have a sample size of 100.

To compute the contribution using the proposed approach, we first need to learn the relationship $f_0(AUP)$ from the simulated data in the baseline period, then instead of using the original decomposition formula of $\Delta n \cdot AUP_1$, we will use $(n_1 - n_0^*) \cdot AUP_1$ where n_0^* can be estimated by $\sum f_0(AUP_{1j})$. This corresponds to the simplified estimator in Eq. (6) before scaling. We repeat the simulations 100 times, calculate the average across 100 runs, and compute 95% confidence intervals for each method. As stated earlier, for both the original decomposition and the proposed approach, we also scaled the results by sample size to make them comparable with the GCM-DC results.

Table 3: Simulation Cases

	Causal relationship for n&AUP	True root causes
case 1a (Fig. 8)	linear	AUP
case 1b (Fig. 9)		AUP & n
case 2a (Fig. 10)	quadratic	AUP
case 2b (Fig. 11)		AUP & n

We performed simulations in 4 different cases (Table 3). From the results of Figs. 8 and 10, the original decomposition incorrectly identified n as the root cause because it did not take into account the causal dependence between AUP and n. Both the unified approach and the GCM-DC are very similar to the truth. This is expected when only one factor changes, Shapley allocation equals ordered allocation. From the results of Figs. 9 and 11 where both variables change, this is where the Shapley value may suffer from reduced accuracy, as it does not utilize the information of the causal path. In these two cases, the proposed unified approach is better than the GCM-DC approach and is closest to the true root causes. Again, if the original decomposition approach is used, the root causes are far from the truth, or in the wrong order. We also used Eq. (5) instead of its simplified form to estimate the contributions and obtained nearly identical results, so we omit the plots. We purposely used a low-dimensional case to illustrate the essence of the problem and improvement, and because they are MTCD's recursive building blocks, validity extends recursively to larger structures.

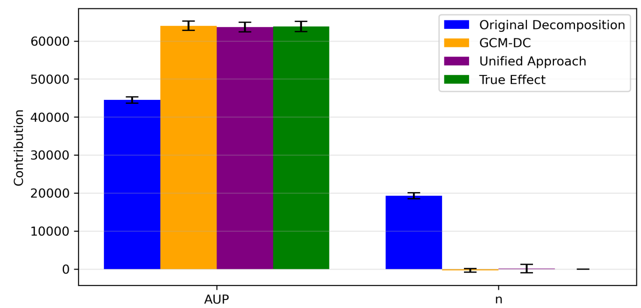

Figure 8: Case 1a. Linear, only AUP changed

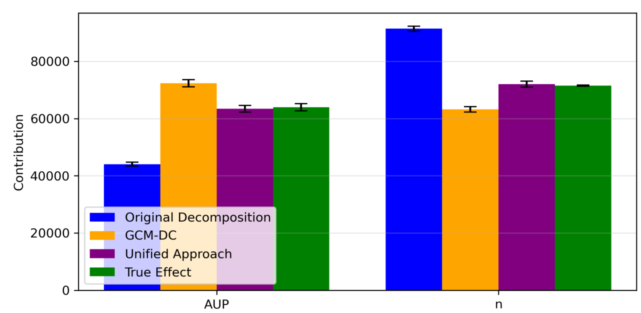

Figure 9: Case 1b. Linear, both AUP and n changed

8 Real-World Marketplace Application

Our proposed method is currently deployed as a real-time root cause engine with flexible time and selection scopes, covering >130 metrics. As an illustration, we conduct a year-over-year comparison for a global e-commerce vendor using two years of daily data. We add only causal edges strongly supported by domain knowledge: AUP $\rightarrow$ units sold and page views $\rightarrow$ conversion rate. The runtime is 1.6 seconds on a standard laptop, approximately 400 times faster than GCM-DC under a comparable configuration. In Fig. 12, less important dimensions/leaf node drivers are omitted for clarity.

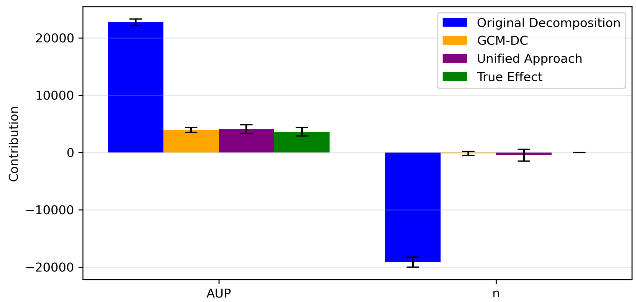

Figure 10: Case 2a. Quadratic, only AUP changed

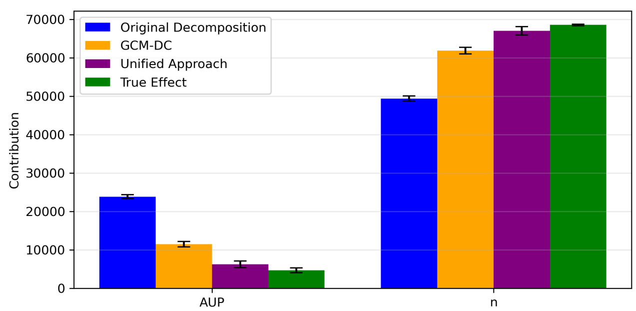

Figure 11: Case 2b. Quadratic, both AUP and n changed

We further rescaled by a constant to preserve data confidentiality. Applying the proposed approach, the top three positive drivers are page view (PV) with deal only, #internal PV, and #PV per new product, while the top three negative drivers are PV without deal or coupon, #other PV, and PV with coupon only, with deal, traffic, and selection as the three main driver segments. This more accurate attribution helps vendors/sellers quantify contribution sizes and prioritize growth actions.

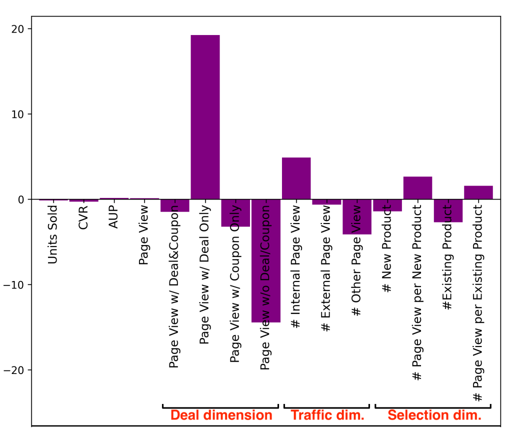

Figure 12: Real-world application

9 Concluding Remarks

In this paper, we proposed a unified method for root cause analysis in complex metric systems that is causally informed, business interpretable, computationally efficient, and unbiased when the relevant causal ordering is unique. The method preserves the operational advantages of MTCD while incorporating sparse local causal structure to correct cross-branch attribution errors and prevents the estimand misalignment in GCM-DC. In production, modeling and maintenance costs remain negligible because modeling is required only for the sparse set of added causal edges, while the rest of the tree remains a model-free MTCD. This is substantially lighter than GCM-DC, which fits conditional models across the entire DAG.

The approach, although much less restrictive than the GCM-DC approach, relies on correct assumptions for the added *local* dependent relationships. In practice, we rely on domain knowledge, temporal ordering of metrics, and conditional falsification checks [7] to validate those added causal links. Because this approach is still subject to "standard" sensitivity to outliers, it suggests future directions including outlier handling methods for root causing, as well as extending the framework to handle temporal dependence in production monitoring systems.

References

[1] Burak Aksar, Efe Sencan, Benjamin Schwaller, Omar Aaziz, Vitus J Leung, Jim Brandt, Brian Kulis, Manuel Egele, and Ayse K Coskun. 2024. Runtime performance anomaly diagnosis in production hpc systems using active learning. *IEEE Transactions on Parallel and Distributed Systems* 35, 4 (2024), 693–706.

[2] B. W. Ang. 2005. The LMDI approach to decomposition analysis: a practical guide. *Energy Policy* 33, 7 (2005), 867–871.

[3] Alan S. Blinder. 1973. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources* 8, 4 (1973), 436–455.

[4] Patrick Blöbaum, Peter Götz, Kailash Budhathoki, Atalanti A. Mastakouri, and Dominik Janzing. 2024. DoWhy-GCM: An Extension of DoWhy for Causal Inference in Graphical Causal Models. *Journal of Machine Learning Research* 25, 147 (2024), 1–7. http://jmlr.org/papers/v25/22-1258.html

[5] Kailash Budhathoki, Dominik Janzing, Patrick Bloebaum, and Hoiyi Ng. 2021. Why did the distribution change?. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 1666–1674.

[6] Kailash Budhathoki, Lenon Minorics, Patrick Blöbaum, and Dominik Janzing. 2022. Causal structure-based root cause analysis of outliers. In *International conference on machine learning*. PMLR, 2357–2369.

[7] Elias Eulig, Atalanti A Mastakouri, Patrick Blöbaum, Michaela Hardt, and Dominik Janzing. 2025. Toward falsifying causal graphs using a permutation-based test. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 26778–26786.

[8] Christopher Frye, Colin Rowat, and Ilya Feige. 2020. Asymmetric shapley values: incorporating causal knowledge into model-agnostic explainability. *Advances in neural information processing systems* 33 (2020), 1229–1239.

[9] Max Halford. 2023. Answering "Why did the KPI change?" using decomposition. https://maxhalford.github.io/blog/kpi-evolution-decomposition/

[10] Marshall Hargrave. 2026. DuPont Analysis: Definition, Uses, Formulas, and Examples. https://www.investopedia.com/terms/d/dupontanalysis.asp. Investopedia. Accessed: 2026-05-01.

[11] Tom Heskes, Evi Sijben, Ioan Gabriel Bucur, and Tom Claassen. 2020. Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models. In *Advances in Neural Information Processing Systems*, Vol. 33. 9983–9993.

[12] Chia-Yen Lee and Bo-Syun Chen. 2018. Mutually-exclusive-and-collectively-exhaustive feature selection scheme. *Applied Soft Computing* 68 (2018), 961–971.

[13] Mingjie Li, Zeyan Li, Kanglin Yin, Xiaohui Nie, Wenchi Zhang, Kaixin Sui, and Dan Pei. 2022. Causal inference-based root cause analysis for online service systems with intervention recognition. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3230–3240.

[14] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*.

[15] Microsoft. 2025. Create and View Decomposition Tree Visuals in Power BI. https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-decomposition-tree. Microsoft Learn. Accessed: 2026-05-01.

[16] Brady Neal. 2020. Introduction to causal inference. *Course lecture notes (draft)* 132 (2020).

[17] Phuoc Nguyen, Truyen Tran, Sunil Gupta, Thin Nguyen, and Svetha Venkatesh. 2024. Root cause explanation of outliers under noisy mechanisms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 20508–20515.

[18] Ronald Oaxaca. 1973. Male-female wage differentials in urban labor markets. *International Economic Review* 14, 3 (1973), 693–709.

[19] Judea Pearl. 2009. *Causality.* Cambridge university press.

[20] Manjula Peiris, James H Hill, Jorgen Thelin, Sergey Bykov, Gabriel Kliot, and Christian Konig. 2014. Pad: Performance anomaly detection in multi-server distributed systems. In *2014 IEEE 7th International Conference on Cloud Computing*. IEEE, 769–776.

[21] Zhifei Shao. 2023. Metric Decomposition Formula to Understand Metric Trend. *Medium* blog post. https://medium.com/@shaozhifei/metric-decomposition-formula-to-understand-metric-trend-e693b7a4c8cf

[22] Lloyd S Shapley et al. 1953. A value for n-person games. (1953).

[23] Anthony F. Shorrocks. 1982. Inequality decomposition by factor components. *Econometrica* 50, 1 (1982), 193–211.

[24] Yue Sun, Rick S Blum, and Parv Venkitasubramaniam. 2025. Unifying Explainable Anomaly Detection and Root Cause Analysis in Dynamical Systems. *arXiv preprint arXiv:2502.12086* (2025).

[25] Tran Minh Tung, Trinh Le Tan, Hoang Thanh Hien, Duong Hoai Lan, Vo Thi Kim Oanh, and Tran Thi Kim Cuc. 2024. Algebraic Method of Problem Analysis in Business Case by Mece Principles. *International Journal of Multiphysics* 18, 3 (2024).

[26] Sanford Weisberg. 2005. *Applied linear regression.* Vol. 528. John Wiley & Sons.

[27] Shifu Yan, Caihua Shan, Wenyi Yang, Bixiong Xu, Dongsheng Li, Lili Qiu, Jie Tong, and Qi Zhang. 2022. Cmmd: Cross-metric multi-dimensional root cause analysis. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4310–4320.

[28] Yuchen Zhu, Kailash Budhathoki, Jonas M Kübler, and Dominik Janzing. 2024. Meaningful causal aggregation and paradoxical confounding. In *Causal Learning and Reasoning*. PMLR, 1192–1217.

Appendices

A Metric-Tree Change Decomposition (MTCD): Pseudo-code

Algorithm 1 implements the recursive MTCD procedure introduced in Sec. 3. We use the same notation as in the main text. In the implementation, the call to BaseDecomp would be computed once per dimension k and reused for all children c with $dim(c) = k$.

Algorithm 1 Metric-Tree Change Decomposition (MTCD)

Require: Tree $\mathcal{T}$ with root r; node values $y_{v,0}, y_{v,1}$; decomposition types $\texttt{type}(c)$; dimension label $dim(c)$

Ensure: Contribution C_v for each node $v \in \mathcal{T}$, such that $\sum_{u\in \mathrm{Ch}(v),\, \dim(u)=k} C_u = C_v$ for all dimensions within node v.

procedure Propagate(v)
  ▷ A function taking a node v, calculating contribution of its children nodes c and looping through the children nodes.
  if v is leaf **then**
    return
  end if
  $\Delta y_v \leftarrow y_{v,1} - y_{v,0}$
  for all $c \in \mathrm{Ch}(v)$ **do**
    $k \leftarrow \dim(c)$
    $\left(\mathrm{contrib}_c\right)_{u\in \mathrm{Ch}(v),\, \dim(u)=k}$
    $\leftarrow \textsc{BaseDecomp}\Big(\texttt{type}(c),\ (y_{u,0}, y_{u,1})_{u\in\{v\}\cup \mathrm{Ch}(v),\, \dim(u)=k}\Big)$
    ▷ Establish parent-children local decomp. per dimension
    if $\Delta y_v \neq 0$ **then** ▷ Avoid dividing by zero
      $C_c \leftarrow \mathrm{contrib}_c \cdot \dfrac{C_v}{\Delta y_v}$
      ▷ Scale by the parent's importance
    else
      $C_c \leftarrow 0$
    end if
    Propagate(c)
  end for
end procedure

$\Delta y_r \leftarrow y_{r,1} - y_{r,0}, \quad C_r \leftarrow \Delta y_r$
▷ Define root's contribution as its Δ
Propagate(r) ▷ Start by calling the function on r

B Proofs

B.1 Proof of Proposition 5.1

Using the same example of Revenue (Rev) = units sold (n) × AUP, assuming the causal graph in Fig. 7, if we use ordered allocation (AUP first and n later) rather than Shapley value allocation, Eq. (3) reduces to

$$\mathbb{E}^{\mathrm{AUP}\cup n}_{\mathrm{Rev}}[\mathrm{Rev}] - \mathbb{E}^{\mathrm{AUP}}_{\mathrm{Rev}}[\mathrm{Rev}]. \tag{7}$$

(7) can then be estimated by

$$\frac{1}{J}\sum_{j=1}^{J}\left(n_{1,j}\cdot AUP_{1,j} - n^*_{0,j}\cdot AUP_{1,j}\right), \tag{8}$$

where $n^*_{0,j}$ is the units sold that would have changed due to the change of AUP for record j. In the special case where AUP is independent of units sold, all $n^*_{0,j} = n_{0,j}$, (8) reduces to

$$\frac{1}{J}\sum_{j=1}^{J}\left(n_{1,j} - n_{0,j}\right)\cdot AUP_{1,j}. \tag{9}$$

Taking the expectation of (9) results in

$$\frac{1}{J}\sum_{j=1}^{J}\mathbb{E}\big[\left(n_{1,j}-n_{0,j}\right)\cdot AUP_{1,j}\big] = \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}[n_{1,j}-n_{0,j}]\cdot\mathbb{E}[AUP_{1,j}], \tag{10}$$

where the last equation holds due to the special assumption that AUP is independent of units sold. Likewise, taking the expectation of the scaled decomposition formula of $\Delta n \cdot AUP_1/J$ gives

$$\begin{aligned}\mathbb{E}\Big[\frac{1}{J}\sum_{j=1}^{J}\left(n_{1,j} - n_{0,j}\right)\cdot AUP_1\Big] &= \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}[n_{1,j} - n_{0,j}]\cdot\mathbb{E}[AUP_1]\\ &= \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}[n_{1,j} - n_{0,j}]\cdot\mathbb{E}[AUP_{1,j}].\end{aligned} \tag{11}$$

The second last equation holds due to the independence assumption, and the last equation holds because $\mathbb{E}[AUP_1] = \mathbb{E}[AUP_{1,j}] = \mathbb{E}\big[\frac{\sum_j \mathrm{Revenue}_{1,j}}{\sum_j n_{1,j}}\big]$ (formal proof can be found in the proof in B.4 Equations (27) and (28)). This completes the proof.

B.2 Proof of Proposition 5.2

Let p_b and p_{nb} denote business units% and non-business units% respectively, and use the AUP example in Sec. 5.2(II): AUP = $p_b \cdot \mathrm{AUP}_b + p_{nb} \cdot \mathrm{AUP}_{nb}$. In the GCM realm, understanding the impact of the business account on overall AUP is equivalent to modeling p_b directly with overall AUP. If we assume that the rate node AUP is linearly impacted by its cause p_b, to compute p_b's impact on

AUP, GCM-DC followed the Shapley allocation's strategy by first computing two scenarios: (1) when nothing is changed:

$$\mathbb{E}^{p_b}_{AUP}[AUP]-\mathbb{E}^{\emptyset}_{AUP}[AUP] = (\beta_0+\beta_1 P_{b,1})-(\beta_0+\beta_1 P_{b,0}) = \beta_1 \Delta P_b; \tag{12}$$

(2) when AUP's mechanism is changed:

$$\begin{aligned}\mathbb{E}^{p_b\cup AUP}_{AUP}[AUP] - \mathbb{E}^{AUP}_{AUP}[AUP] &= (\beta_0^*+\beta_1^* P_{b,1}) - (\beta_0^*+\beta_1^* P_{b,0}) \\ &= \beta_1^*\Delta P_b. \end{aligned}\tag{13}$$

Thus the Shapley value contribution of p_b equals the average of (12) and (13), which is

$$\begin{aligned} &= \tfrac{1}{2}\,\Delta P_b\,(\beta_1+\beta_1^*) \\ &\approx \tfrac{1}{2}\,\Delta P_b\left[(AUP_{b,0}-AUP_{nb,0})+(AUP_{b,1}-AUP_{nb,1}))\right], \end{aligned}\tag{14}$$

where the last approximation in (14) holds true because

$$\begin{aligned}\mathrm{AUP} &= p_b\,\mathrm{AUP}_b + p_{nb}\,\mathrm{AUP}_{nb} \\ &= \mathrm{AUP}_{nb} + p_b\,(\mathrm{AUP}_b - \mathrm{AUP}_{nb}), \end{aligned}\tag{15}$$

and when a linear regression is estimated using granular observations, $\beta_0 = \mathbb{E}[\mathrm{AUP}_{nb}]$ and $\beta_1 = \mathbb{E}[\mathrm{AUP}_b - \mathrm{AUP}_{nb}]$ holds true ([26]).

On the other hand, in the decomposition paradigm,

(i) contribution of $p_b = \Delta p_b\,(AUP_{b,0} - AUP_0)$,

(ii) contribution of $p_{nb} = \Delta p_{nb}\,(AUP_{nb,0} - AUP_0)$,

because $\Delta p_{nb} = -\Delta p_b$, adding (i) and (ii) equals

$$\begin{aligned} &= \Delta p_b\,(AUP_{b,0}-AUP_0) - \Delta p_b\,(AUP_{nb,0}-AUP_0) \\ &= \Delta p_b\,(AUP_{b,0}-AUP_{nb,0}). \end{aligned}\tag{16}$$

When AUP's distribution is not changed (i.e., $\beta_1 = \beta_1^*$), (14) and (16) are equivalent. When ordered allocation is used (i.e., use the first scenario in (12) instead of both scenarios), (12) $\approx$ (16). The proof is complete.

B.3 Proof of Theorem 6.1

Given the rational stated in Section 6, when a reasonable causal graph is available, or the sequence of metrics ("players") is fixed, one variable's contribution on the target variable depends only on its parent node(s)' (i.e., causes) status. Hence, in the case of the Type 2 non-rate metric, given the provided causal graph, it is natural to define the true root cause as $\mathbb{E}^{\bar{X}\cup n}_y[y] - \mathbb{E}^{\bar{X}}_y[y]$. We are ready to prove the unbiasedness. Take the expectation of our proposed estimator,

$$\begin{aligned}\mathbb{E}[\hat{RC}(n|\bar{X})] &= \mathbb{E}\Big[\frac{1}{J}\sum_{j=1}^{J}\big(n_{1,j}\cdot\bar{X}_{1,j} - \hat{f}_0(\bar{X}_{1,j})\cdot\bar{X}_{1,j}\big)\Big] \\ &= \mathbb{E}\big[n_{1,j}\cdot\bar{X}_{1,j}\big] - \mathbb{E}\big[\hat{f}_0(\bar{X}_{1,j})\cdot\bar{X}_{1,j}\big]. \end{aligned}\tag{17}$$

The first term in (17) is equivalent to $\mathbb{E}^{\bar{X}\cup n}_y[y]$ by definition as both n and $\bar{X}$ are values under the new distribution, and $y = n\cdot\bar{X}$ holds true at any given point. Now, focusing on the second term, we start by working backward from the final term of $\mathbb{E}^{\bar{X}}_y[y]$, let us define $f_0(\bar{X}_{1,j},\epsilon_{0,j})$ as the true n that would have changed under the new distribution of $\bar{X}$, so

$$\begin{aligned}\mathbb{E}^{\bar{X}}_y[y] &\equiv \mathbb{E}^{\bar{X}}_y\big[f_0(\bar{X}_{1,j},\epsilon_{0,j})\cdot\bar{X}_{1,j}\big] \\ &= \mathbb{E}^{\bar{X}}_y\big[[f_0(\bar{X}_{1,j})+\epsilon_{0,j}]\cdot\bar{X}_{1,j}\big] \qquad (18)\\ &= \mathbb{E}^{\bar{X}}_y\big[f_0(\bar{X}_{1,j})\cdot\bar{X}_{1,j}\big] + \mathbb{E}^{\bar{X}}_y[\epsilon_{0,j}\cdot\bar{X}_{1,j}] \\ &= \mathbb{E}^{\bar{X}}_y\big[f_0(\bar{X}_{1,j})\cdot\bar{X}_{1,j}\big] + \mathbb{E}^{\bar{X}}_y[\epsilon_{0,j}]\cdot\mathbb{E}^{\bar{X}}_y[\bar{X}_{1,j}] \qquad (19)\\ &= \mathbb{E}^{\bar{X}}_y\big[f_0(\bar{X}_{1,j})\cdot\bar{X}_{1,j}\big]. \qquad (20)\end{aligned}$$

Eq. (18) is true under the typical additive noise model, and (19) is true because each noise term ϵ is commonly assumed to be independent of $\bar{X}$ variable. And under the assumption of $\mathbb{E}[\epsilon_{0,j}] = 0$, (20) is true. Regarding the second term of (17), under the assumption of modeling estimation of $\mathbb{E}[\hat{f}_0(\bar{X}_{1,j})|\bar{X}_{1,j}] = f_0(\bar{X}_{1,j})$,

$$\begin{aligned}\mathbb{E}\big[\hat{f}_0(\bar{X}_{1,j})\cdot\bar{X}_{1,j}\big] &= \mathbb{E}\big[\mathbb{E}[\hat{f}_0(\bar{X}_{1,j})\cdot\bar{X}_{1,j}|\bar{X}_{1,j}]\big] \\ &= \mathbb{E}\big[\mathbb{E}[\hat{f}_0(\bar{X}_{1,j})|\bar{X}_{1,j}]\cdot\bar{X}_{1,j}\big] \\ &= \mathbb{E}\big[f_0(\bar{X}_{1,j})\cdot\bar{X}_{1,j}\big]. \end{aligned}\tag{21}$$

Because (21) and (20) are equivalent, we proved the unbiasedness that $\mathbb{E}[\hat{RC}(n|\bar{X})] = \mathbb{E}^{\bar{X}\cup n}_y[y] - \mathbb{E}^{\bar{X}}_y[y]$, and that the noise terms do not need to be in the estimator to be unbiased.

B.4 Proof of Proposition 6.2

First, the estimator in (6) can be re-written as

$$\begin{aligned} &= \frac{1}{J}\Big[\sum_{j=1}^{J} n_{1,j} - \sum_{j=1}^{J}\hat{f}_0(\bar{X}_{1,j})\Big]\cdot\bar{X}_1 \\ &= \frac{1}{J}\Big[\sum_{j=1}^{J}\big[\hat{f}_1(\bar{X}_{1,j},\epsilon_{1,j}) - \hat{f}_0(\bar{X}_{1,j})\big]\Big]\cdot\bar{X}_1. \end{aligned}\tag{22}$$

Similar to the proof in B.3, under the assumption of (i) an additive noise model, (ii) $\mathbb{E}[\hat{f}_0(\bar{X}_{1,j})|\bar{X}_{1,j}] = f_0(\bar{X}_{1,j})$, and (iii) $\mathbb{E}[\epsilon_{1,j}] = 0$, taking the expectation and applying the law of total expectation leads to

$$\mathbb{E}[(22)] = \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}\big[\big[f_1(\bar{X}_{1,j}) - f_0(\bar{X}_{1,j})\big]\cdot\bar{X}_1\big]. \tag{23}$$

Under the proposition assumption that the change distribution between two periods is only from an intercept, denoted as C, (23) becomes

$$\frac{1}{J}\sum_{j=1}^{J}\mathbb{E}[C\cdot\bar{X}_1] = C\cdot\mathbb{E}[\bar{X}_1]. \tag{24}$$

Following a similar strategy, Eq.(5) can be re-written as

$$\begin{aligned} &= \frac{1}{J}\sum_{j=1}^{J}\big[(n_{1,j} - \hat{f}_0(\bar{X}_{1,j}))\cdot\bar{X}_{1,j}\big] \\ &= \frac{1}{J}\sum_{j=1}^{J}\big[(\hat{f}_1(\bar{X}_{1,j},\epsilon_{1,j}) - \hat{f}_0(\bar{X}_{1,j}))\cdot\bar{X}_{1,j}\big]. \end{aligned}\tag{25}$$

Similar to (23) (or B.3), taking the expectation and applying the law of total expectation along with assumptions (i)-(iii) above,

$$\begin{aligned}\mathbb{E}[(25)] &= \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}\big[\big(f_1(\bar{X}_{1,j}) - f_0(\bar{X}_{1,j})\big)\cdot \bar{X}_{1,j}\big] \\ &= \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}[C\cdot \bar{X}_{1,j}] \\ &= C\cdot \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}[\bar{X}_{1,j}]. \end{aligned} \tag{26}$$

Comparing (26) with (24), now let us prove $\mathbb{E}[\bar{X}_1] = \mathbb{E}[\bar{X}_{1,j}]$. Assume $\mathbb{E}[\bar{X}_{1,j}] = \mu$, because

$$\bar{X}_1 = \frac{\sum_j y_{1,j}}{\sum_j n_{1,j}} = \frac{\sum_j n_{1,j}\cdot \bar{X}_{1,j}}{\sum_j n_{1,j}} \equiv \sum_j w_j \bar{X}_{1,j} \tag{27}$$

where $w_j = \frac{n_{1,j}}{\sum_j n_{1,j}}$ and $\sum_j w_j = 1$. So

$$\begin{aligned}\mathbb{E}[\bar{X}_1] = \mathbb{E}[\sum_j w_j \bar{X}_{1,j}] &= \mathbb{E}[\mathbb{E}[\sum_j w_j \bar{X}_{1,j} | n_{1,1}, \ldots, n_{1,J}]] \\ &= \mathbb{E}[\sum_j w_j \mathbb{E}[\bar{X}_{1,j}]] \\ &= \mathbb{E}[\sum_j w_j \mu] = \mathbb{E}[\mu] = \mu \end{aligned} \tag{28}$$

by proving $\mathbb{E}[\bar{X}_1] = \mathbb{E}[\bar{X}_{1,j}]$, the proof is complete.

C Simulation Studies Details for Section 5.3

(note: for all the figures in this section, "GCM-based" denotes the GCM-DC approach, whereas "Decomposition-based" denotes the MTCD approach.)

C.1 Non-rate metric simulations

Scenario 1. Assuming causes (i.e., sibling nodes in MTCD) are independent & causal DAG is correctly assumed. We use revenue = AUP *units sold as the simplest case. We want to test the special case when there is no relationship between units sold and AUP in simulated data, and the DAG is correctly specified with only revenue being a function of units sold and AUP, and AUP being independent of units sold. We simulated data for these three variables: AUP, units sold, revenue for the base period with sample size 100, and assumed that both AUP and units sold increased in mean separately in the new period with the same sample size of 100. As discovered in the theoretical proofs section, we also scaled the contribution by sample size to make the two methods comparable and the same for all the following simulations. In this case, as observed in Fig. 13, both approaches correctly identified the right root causes with the correct order, but with slightly different magnitudes. This difference originates exactly from the fact that GCM-DC uses Shapley, while the decomposition approach uses ordered allocation, aligned with the theoretical proofs section.

Scenario 2a. Assuming causes (i.e., sibling nodes in MTCD) are dependent & causal DAG is correctly assumed. Now we include one more layer of nodes under units sold and use the same 4-node causal graph shown in the real-world application in Fig. 3(b). We used

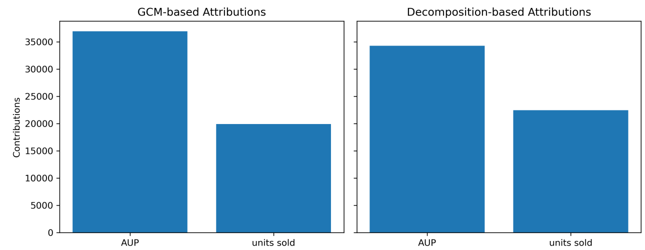

Figure 13: Scenario 1. when causal DAG is correctly assumed with independent causes

the same DAG structure to simulate the data for one period and increase the AUP and page views in the mean for the next period. GCM-DC and MTCD (using Fig. 3(a)) are applied separately to find the root causes of the change in revenue. Compared with the ground truth of AUP and the page views in the leftmost panel of Fig. 14, GCM-DC correctly identified the true root causes with the correct order, while MTCD failed to detect the correct root causes both in existence and magnitude. This highlights the limitation of MTCD mentioned earlier that, when there is a true dependence among causes, this approach is likely to fail. It can also be seen that the magnitude of the GCM-DC contributions is slightly off from the true magnitude. We further increased the sample size from 100 to 1000 and this does not change the findings. This revealed the fact that using the Shapley value may deviate from the true contribution allocation, which will be addressed in our proposed unified approach section.

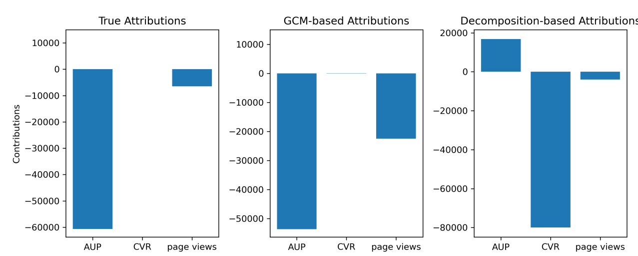

Figure 14: Scenario 2a. when causal DAG is correctly assumed (n=100): true AUP and page view effects

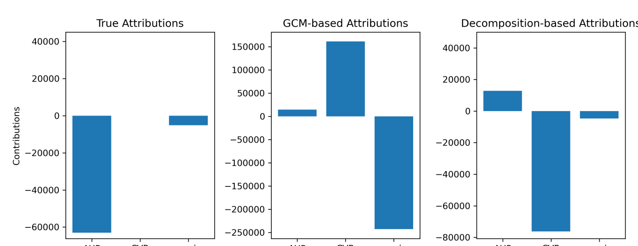

Figure 15: Scenario 2a. when causal DAG is correctly assumed (n=30): true AUP and page view effects

Separately, as noted in the GCM-DC limitations that a reasonable sample size is required to learn the conditional distribution for each children-parent pair, thus, we tested when the sample size is reduced to 30. As shown in Fig. 15, GCM-DC no longer dectects the correct root causes, and MTCD still has similar findings to those under sample size of 100. To this end, sample size is another important factor in determining the right root causes. All following simulations are conducted with sample size 100, unless otherwise noted.

We also examined other cases where the true root cause is AUP only and page views only, and the findings are similar (plots omitted). On the other hand, when there is only the CVR effect, it is observed and expected that both methods are able to only pick up CVR as the root cause and with a similar magnitude since there are no other factors causing the CVR to change.

Scenario 2b. weak relationship among nodes & small signal-to-noise ratio. Similar to Scenario 2a, we assumed that causes are dependent & and the causal DAG is correctly assumed; however, we manually weakened the nodes relationship by simulating page views = f(AUP) with person correlation of 0.2, and increased the variance of the noise term in CVR = f(AUP,page views) such that R^2 is approximately 0.2. This is designed to test whether a weak relationship between dependent nodes, and a small signal-to-noise ratio would lead to a wrong inference. But from the simulation results in Fig. 16, GCM-DC continued to recognize the correct root causes with the correct order, but the magnitude is slightly off. This will be tackled in the proposed unified approach section. On the other hand, MTCD, as expected, was unable to identify the correct root causes due to the missing dependence between the sibling nodes (i.e., causes).

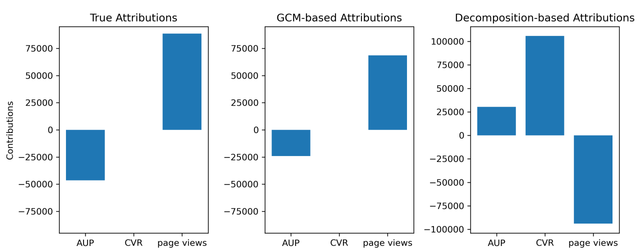

Figure 16: Scenario 2b. when causal DAG is correctly assumed with weak relations: true AUP and page view effects

Scenario 3a. Assuming causes (i.e., sibling nodes in MTCD) are dependent but causal DAG is incorrectly assumed (i.e., same graph but wrong edge directions). We want to examine how these two methods behave if, unlike Scenario 2a, the truth causal graph is different from the assumed one. As a hypothetical example, the assumed DAG is CVR = f(AUP,page views); page views = f(AUP), but we generated data by AUP = f(CVR,page views) and page views = f(CVR) for the baseline period. In the new period, we set the true root cause to be AUP by only increasing the mean of AUP, and other nodes are changed correspondingly using the same data generating process in the baseline period. It is noted that in this setting, the causal graph cannot be falsified as there are edges between every possible pair, making the falsification test [7] impossible. From Fig. 17, neither method can successfully identify the correct root causes, and the GCM-DC approach detected more incorrect root causes than MTCD.

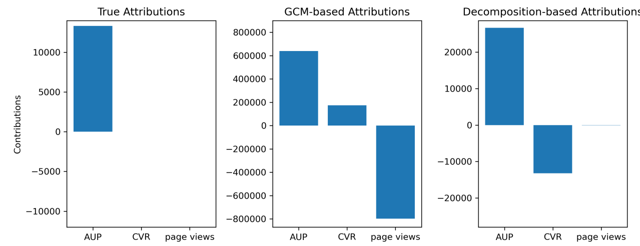

Figure 17: Scenario 3a. when causal DAG is incorrectly assumed: true AUP effect only

When CVR is set as the only true root cause in the simulation, similarly, MTCD seemed to select fewer incorrect root causes compared to GCM-DC (Fig. 18). Both cases underscore the importance of correct causal DAG assumption in order for GCM-DC to have a correct inference.

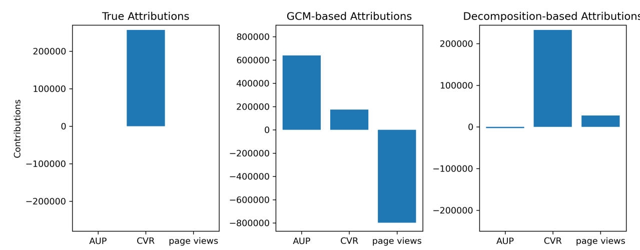

Figure 18: Scenario 3a. when causal DAG is incorrectly assumed: true CVR effect only

Scenario 3b. Assuming causes (i.e., sibling nodes in MTCD) nodes are dependent but causal DAG is incorrectly assumed (i.e., fewer edges). In this scenario, we generated the data from CVR = f(AUP,page views); page views = f(AUP), but assumed a simple causal graph to be used in GCM-DC. That is, revenue is affected by all three nodes of AUP, page views, and CVR, but these three nodes are independent of each other. Then, we increased the mean values of AUP and page views in the process of generating the data for the new period. In summary, both methods are somewhat similar because both have the same underlying causal structure: no dependence among causes. And the difference is mostly due to Shapley and ordered allocation. However, both failed to select the correct root causes (Fig. 19). In addition, the built-in falsification test in the GCM package (dowhy.gcm.falsify_graph(·), [4]) failed to reject the incorrect over-simplified DAG assumption. In such a case, the GCM-DC results become untenable. Testing in simpler true root cause cases of: (1) AUP only; and (2) page views only resulted in similar findings. When the true root cause is only CVR instead, both methods can actually recognize CVR correctly as the root cause (figure omitted here). This is simply because the other two nodes that affect CVR did not change from the previous period, and hence ignoring the dependence relationship becomes trivial.

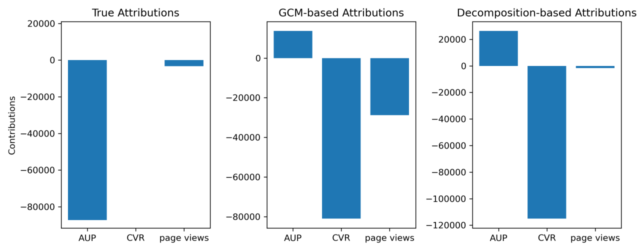

Figure 19: Scenario 3b. when causal DAG is incorrectly assumed with independence among causes

C.2 Rate metric simulations

We use Fig. 20 as an illustration of the rate metric to conduct simulation studies. Namely, AUP $=$ business units% $*$ AUP$_b$ $+$ non-business units% $*$ AUP$_{nb}$.

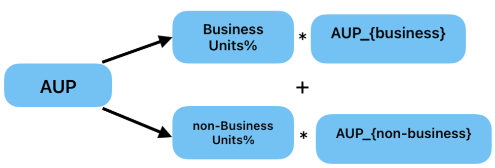

Figure 20: Rate metric decomposition

We want to understand how each of the 4 nodes contributes to the overall AUP change. This is not a typical causal graph, which usually involves outcome, predictor, and confounder nodes instead of including the subgroup metrics of the outcome and share nodes (i.e., business units% and non-business units%) given that they add up to 1. In order to produce a proper causal DAG set up for a GCM-based approach, we reframed this decomposition relationship to business units% (denoted as p_b) impacting AUP as in Fig. 21. We use an analogy example to illustrate why only one node is needed instead of four. For example, if the outcome is average income and

Figure 21: causal DAG for GCM

gender percentage per region is an observed predictor, instead of decomposing the average income in the following decomposition framework (Fig. 22), we can just model the average income against male% instead. That is, it becomes natural that we only need to model the ones with red circles, which is typical in a regular machine learning setting. In the linear regression case, the subgroup income for the female is equivalent to the intercept in the regression, and {avg income$_{\text{male}}$ − avg income$_{\text{female}}$}is equivalent to the linear slope term.

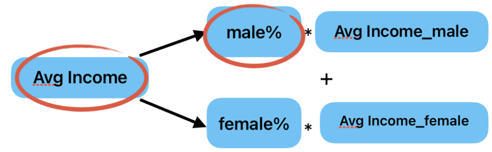

Figure 22: Analog example

Scenario 4a. only p_b changes. Returning to the AUP example with business account dimension breakdown, we simulated AUP, and units sold for business and non-business subgroups from separate normal distributions, and derived the corresponding business and non-business share percentages and the overall AUP. Then, in the new period, we simulated a case where only business units share percent (p_b) is the root cause of the change for the overall AUP while keeping other factors with the same distribution. Observing the results in Fig. 23, firstly, according to the theoretical proofs section for the rate metric in Sec. 5.2, the target scale for GCM-DC and the decomposition-based approach is slightly different, leading to the difference in the true contributions not being identical in the left and right panels in Fig. 23. However, both GCM-DC and the decomposition-based methods can identify the correct root cause. Secondly, while the GCM-DC uses Shapley allocation and the decomposition-based approach uses ordered allocation, but in this case of AUP distribution mechanism did not change, both allocations become equivalent. Hence, bearing the subtle difference of the target scale, and comparing the two orange bars in the left and right plots, this simulation study also verifies the special case in Sec. 5.2 that the business node contribution of the GCM-DC is approximately equal to the sum of share nodes contribution of the decomposition-based approach. Both methods are comparable in this case.

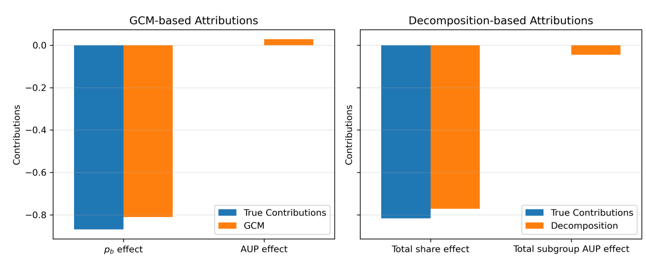

Figure 23: Scenario 4a. only p_b changed

Scenario 4b. both AUP_b and p_b change (assuming AUP_b and p_b are independent). In this scenario, in the new period, we not only increased the mean of units sold for business to indirectly increase the business share percent like scenario 4a, but also increased the mean of business subgroup AUP. This makes Shapley and ordered allocation no longer equivalent. In this case, even though the true increase of each variable is known, when nodes are independent, there is no unique causal ordering, hence there is no ground truth contribution of subgroup AUP and units share percentages. From Fig. 24, both methods can correctly recognize the root causes with the right order, but the magnitude is different, and this aligns with the theoretical proofs section that the differences of the two approaches are mainly due to: (1) target scales are slightly different; and (2) the choice of Shapley vs ordered allocation. Neither allocation approach can be determined better when there is no ground truth. However, the decomposition-based approach also provides a more granular contribution of each AUP subgroup (p_b : $-0.701, p_{nb}$: $-0.071, AUP_b$: $1.337, AUP_{nb}$: -0.037). The decomposition-based approach also successfully selected AUP_b as the root cause, while GCM-DC can only know that the distribution of AUP has changed but does not know which part.

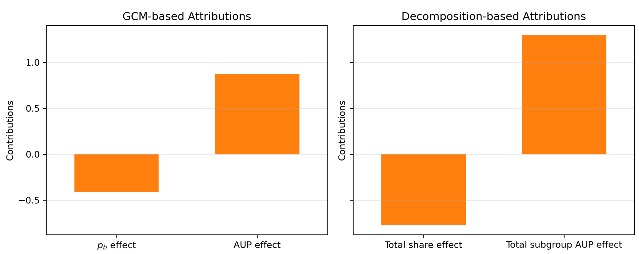

Figure 24: Scenario 4b. both p_b and AUP$_b$ changed

Scenario 5: multivariate dimensions. Like the MTCD approach we developed in Section 3, it is often reasonable to break down a rate metric into multiple dimensions. Given that GCM-DC models all parent nodes together with their joint child node in each conditional distribution, while in contrast, the MTCD approach decomposes a node into multiple dimensions separately in parallel, we want to test the hypothesis whether MTCD inflates each node's contribution compared to the GCM-DC approach. Without loss of generality, we will simulate a 2-dimensional scenario and hypothetically name them as business and subscription for easier demonstration (Fig. 25). Similarly, we add the p_s node to represent subscription's impact in the causal graph for GCM-DC (Fig. 26). We assume that the business account factor is independent from the subscription factor in this simulation. Given the model mechanism difference between GCM-DC and MTCD, the data generating process becomes more complex and was conducted as follows: Because we model $AUP = \beta_0 + \beta_1 p_b + \beta_2 p_s + \epsilon$ for GCM-DC, and $AUP = p_b AUP_b + (1 - p_b) AUP_{nb}$ for MTCD, solving these two equations leads to:

$$\mathbb{E}(AUP_b) = \beta_0 + \beta_1 + \beta_2 \mathbb{E}(p_s) \tag{29}$$

$$\mathbb{E}(AUP_{nb}) = \beta_0 + \beta_2 \mathbb{E}(p_s)$$

$$AUP_{nb} = (AUP - p_b AUP_b)/(1 - p_b). \tag{30}$$

With these relationships, we can simulate p_b and p_s from two beta distributions, and pre-specify β coefficients to generate the AUP distribution, and consequently simulate AUP_b with the mean from (29) above with a fixed variance, and obtain the corresponding AUP_{nb} by (30). We can also solve $\mathbb{E}(AUP_s)$ similarly as a function of p_b, and the corresponding subgroup AUP distributions for subscription can be simulated. This consitutes the data for the base period.

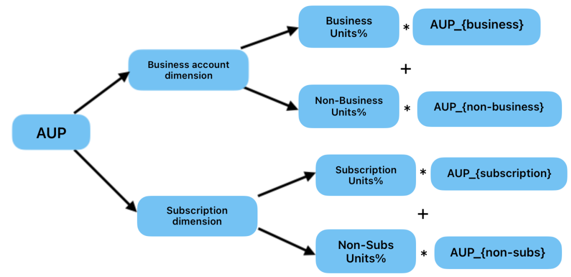

Figure 25: 2-dimensional MTCD

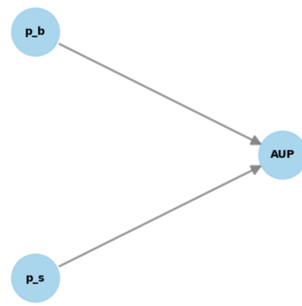

Figure 26: 2-dimensional example causal DAG

Scenario 5a: simulate only p_b change. In the new period, if only p_b is the root cause, according to the data generating process, this will consequently change the overall AUP, AUP_s, and AUP_{ns}. We plotted the p_b effect from GCM-DC, the total share effect from MTCD, and the true p_b effect on the left, and similarly for subscription effect on the right within the left panel of Fig. 27. Both GCM-DC and MTCD effectively identified root causes and provided satisfactory contribution estimates. This also confirms p_b from GCM-DC $\approx p_{nb} + p_b$ from MTCD, which aligns with the theoretical proofs section. Additionally, in MTCD, the contribution of $AUP_s + AUP_{ns}$ also approximately equals the contribution of the total share effect for business (i.e., $p_{nb} + p_b$).

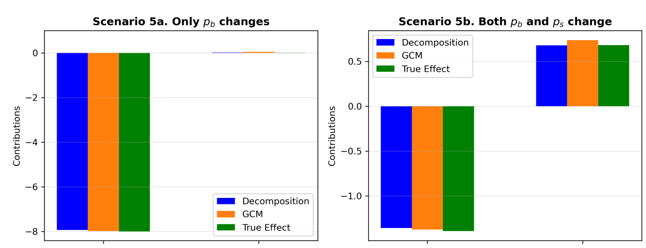

Figure 27: Scenario 5a-b. multivariate dimensions for rate metrics

Scenario 5b: simulate p_b and p_s change. When both p_b and p_s are the root causes, based on the data generating process, the general AUP and all subgroup AUP (AUP_s, AUP_{ns}, AUP_b, and AUP_{nb}) will be consequently changed. The right plot of Fig. 27 indicates that both approaches accurately estimated the root causes in both identification and magnitude.

In summary, under the assumption that multiple nodes are independent (e.g., p_s is independent of p_b), computing the contributions in separate dimensions does not exaggerate the contribution to its parent node in the tree (e.g., overall AUP), proving MTCD is a valid approach.

Below are some additional data generating process details for preparing inputs needed for both GCM-DC and decomposition-based approaches throughout the simulations:

(1) Non-rate simulations: We assume true mean and variance parameters to simulate the daily records, then aggregate the daily data to period-level metric through units sold, page views, and revenue in order to derive corresponding AUP, CVR on the aggregate level. These will be inputs for the decomposition-based approach.

(2) Rate metric simulations: We generate daily sub-group units sold and sub-group AUP under certain distributions, and then derive period-aggregated sub-group revenue, sub-group units%, overall AUP, which are needed for the decomposition-based approach.