**The E[X] Group**
Ad spend, revenue & margin optimization

# Ad Spend & Margin Review

## Technical Report

Marketing Mix Modeling · Budget Reallocation Analysis



| | |
|---|---|
| **Prepared for:** | Northwind Outdoor Co. *(illustrative client)* |
| **Engagement:** | Full Spend & Margin Review |
| **Data window:** | Nov 2015 – Nov 2019 (weekly) |
| **Prepared by:** | The E[X] Group |
| **Report date:** | [DATE] |
| **Status:** | **Confidential** |

## Contents

## 1. Executive Summary

This review examines whether Northwind Outdoor Co.'s advertising budget is allocated efficiently across its five paid channels, and what a reallocation of the *existing* budget — with no additional spend — would be worth.

### 1.1 Headline finding

The budget is appropriately sized but unevenly allocated. Two channels (out-of-home and print) absorb a large share of spend while returning close to break-even, and two channels (paid search and television) show materially higher returns with room to absorb more budget. Reallocating within the current total is estimated to increase media-driven revenue by approximately **10%** (plausible range **4%–16%**), at **no additional cost**.

### 1.2 What we recommend

1. Shift budget from out-of-home (−20%) and print (−50%) toward paid search (+50%) and television (+50%), holding total spend fixed.
2. Before committing budget, validate the single largest assumption — that this spend *causes* the associated revenue — with a two-week geographic holdout test on paid search.
3. Supply full spend history, including any retail-media or creator spend not in the current data, and refit.

### 1.3 How much confidence to place in this

We are confident in the *direction* of the recommendation: across every alternative model specification we tested, paid search and television remained the highest-return channels and out-of-home the lowest. We are less certain of the exact *magnitude*: the estimated revenue gain ranges from 4% to 16% depending on modeling assumptions detailed in Sections 7 and 8, and one channel (television) is estimated with wide uncertainty. This is a correlational model, not a causal one; the holdout test in our recommendation is what would convert the estimate into a verified result.



**Read this report as a prioritized, testable hypothesis — not a settled fact.** It identifies where the budget is most likely being wasted and where the next dollar most likely earns more, and it specifies the test that would confirm it. Used that way, it is a reliable guide to the next budget decision. Treated as a guarantee, it would be overread.

## 2. Data Provenance & Preparation

### 2.1 Sources and window

The analysis uses weekly observations spanning November 2015 to November 2019 (208 weeks), comprising:

| Category | Variables | Role in model |
|---|---|---|
| Revenue | Weekly total revenue | Dependent variable |
| Paid media (spend) | TV, out-of-home, print, Facebook, search | Modeled with carryover & saturation |
| Paid media (exposure) | Facebook impressions, search clicks | Exposure proxies |
| Organic | Newsletter sends | Non-paid marketing |
| Context | Competitor sales, event flags | Controls |

### 2.2 Preparation steps

Data was cleaned and reconciled before modeling: channel spend was verified as numeric and consistently denominated; dates were aligned to a common weekly boundary; and per-channel spend totals were checked against source exports. No observations were dropped.

**Data limitations to note up front.**

- **Window length.** 208 weekly observations is a reasonable but not large sample for a five-channel model with carryover and saturation. Channel-level parameters carry meaningful uncertainty, quantified in Section 7.
- **Spend reconciliation.** In a real engagement, modeled spend would be reconciled against the client's finance ledger. Any marketing cost sitting outside the ad-platform exports — retail-media fees, creator/influencer payments, agency retainers — would be absent from this model, and its effect would be absorbed into the baseline or into correlated channels. *For this illustrative dataset no such reconciliation is possible.*
- **Margin data.** This sample models revenue. Profit-on-ad-spend (POAS) recommendations require per-SKU cost data, which would be requested in a real engagement and is not present here.

## 3. Model Specification

Revenue was modeled using a marketing mix model implemented in Meta's Robyn framework. Each paid channel's spend is transformed by two mechanisms that reflect real advertising behavior, then combined in a regularized regression. The full mathematical treatment — optimization conditions, alternative transforms, and the seasonality decomposition — is provided in Appendix A; this section states the model at the level needed to interpret the findings.

### 3.1 The two transformations

**Carryover (adstock).** Advertising seen in one week continues to influence purchases in later weeks. Each channel's spend is spread forward geometrically:

$$A_t = x_t + \theta\, A_{t-1}, \qquad \theta \in [0, 1),$$

where $\theta$ is the channel's carryover rate; a higher $\theta$ means a longer memory. (Robyn also supports two Weibull carryover forms that allow a *delayed* peak effect; these are described in Appendix A.4 and were used in the robustness checks of Section 7.)

**Diminishing returns (saturation).** Each additional dollar returns less than the last. Adstocked spend passes through a Hill curve:

$$S_t = \frac{A_t^{\alpha}}{A_t^{\alpha} + \kappa^{\alpha}}, \qquad \kappa = \gamma \max_t(A_t),$$

producing an S- or C-shaped response that flattens as spend grows. The shape parameter $\alpha$ and the inflection proportion $\gamma$ are estimated per channel.



**Response.** The transformed channels enter a ridge regression with non-negative coefficients on media, alongside baseline controls for trend, seasonality, and competitor activity:

$$\hat{y}_t = \beta_0 + \sum_c \beta_c S_{c,t} + \sum_j \delta_j Z_{j,t}, \qquad \beta_c \geq 0.$$

### 3.2 How the model is estimated

Three distinct estimation problems are solved, each detailed formally in the appendix:

1. **Baseline decomposition (Prophet).** Trend, seasonality, and holiday effects are separated from the revenue series by a decomposable additive model before media effects are attributed. This prevents the model from crediting to advertising what is really seasonal demand. Appendix A.5.

2. **Coefficient fitting (constrained ridge).** For any fixed set of carryover and saturation parameters, the channel coefficients are found by a ridge regression with a non-negativity constraint on media — a convex optimization with a Lagrangian and Karush–Kuhn–Tucker (KKT) conditions given in Appendix A.1.

3. **Hyperparameter search (Nevergrad).** The carryover and saturation parameters themselves are chosen by an evolutionary algorithm that optimizes two objectives simultaneously — predictive accuracy and business plausibility — producing a set of Pareto-optimal models rather than a single fit. Appendix A.3.

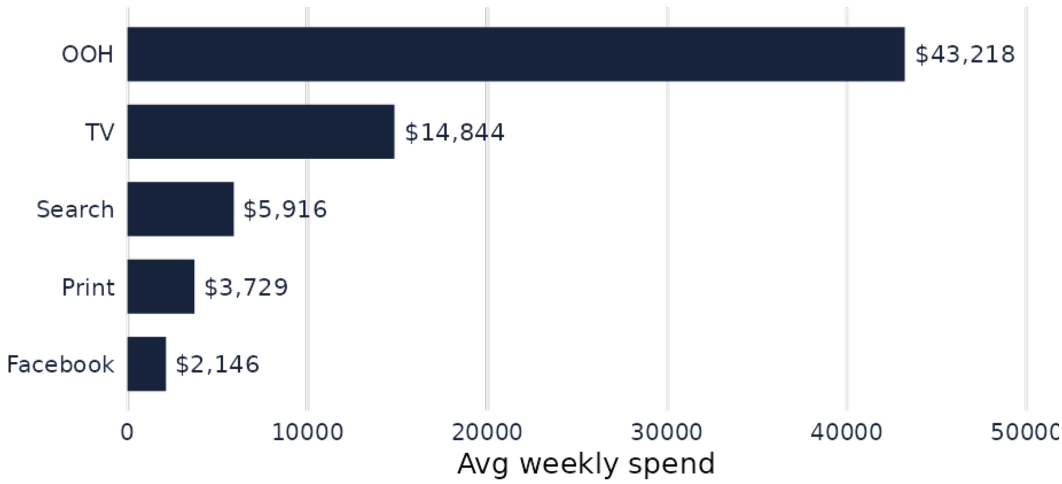
### 3.3 Model selection

The two-objective search returns a Pareto front: models for which neither objective can improve without the other worsening. From this front, one model was selected for the reported findings on the basis of stable channel estimates and plausible response curves. The consequence of this selection — and the business prior embedded in the second objective — is discussed in Section 7 and Appendix A.3.

# 4. Current State of Spend & Efficiency

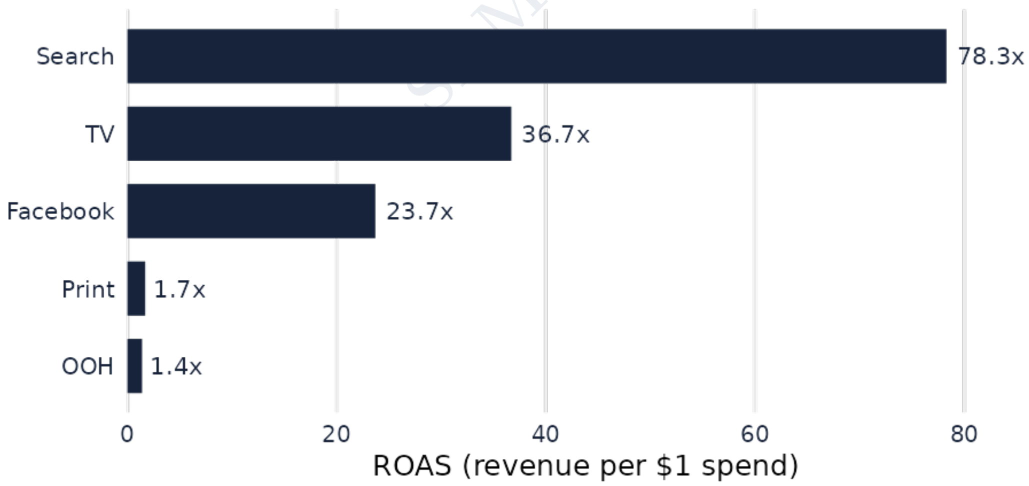
## 4.1 Where the budget goes today

Out-of-home absorbs the majority of weekly media spend — substantially more than any other channel.




## 4.2 What each channel returns

Efficiency tells a different story. Paid search and television return many times their spend, while out-of-home and print sit close to break-even — each dollar returns barely more than a dollar of revenue, before costs are even considered.




**The core inefficiency.** The channel receiving the most budget (out-of-home) is among the least efficient. This is the central finding the reallocation addresses: money is concentrated where returns are lowest.

## 5. Recommended Reallocation

### 5.1 The proposed change

Holding total budget fixed, we recommend moving spend from the two low-return channels toward the two high-return channels, within a cap of ±50% per channel (a guardrail keeping recommendations inside the range the data supports).

| Channel | Now | Recommended | Change | ROAS now | ROAS new |
|---|---|---|---|---|---|
| Paid search | $5,916 | $8,873 | +50% | 78.3× | 63.8× |
| Television | $14,844 | $22,266 | +50% | 36.7× | 26.1× |
| Facebook | $2,146 | $2,320 | +8% | 23.7× | 22.1× |
| Out-of-home | $43,218 | $34,528 | −20% | 1.4× | 1.0× |
| Print | $3,729 | $1,864 | −50% | 1.7× | 1.7× |
| **Total** | **$69,853** | **$69,851** | **fixed** | | |

### 5.2 Why raising spend lowers a channel's ROAS

Note that scaling up the strong channels slightly *lowers* their ROAS (search from 78.3× to 63.8×). This is expected and correct: diminishing returns mean each added dollar earns a little less than the average dollar before it. The objective is not maximum ROAS on any single channel — it is maximum *total* return across the mix. The response curve below shows why: at current search spend, the curve is still steep, so additional budget is still productive.


The recommended point sits where paid search's next dollar earns roughly the same as the next dollar into the other scaled channels — the mathematical condition for an optimal allocation. Moving further would push search past the point where it beats the alternatives.

## 6. Validation & Model Diagnostics

A model's fit on the data used to build it is not evidence. This section reports how the model behaves under scrutiny.

### 6.1 Convergence

The hyperparameter search was assessed against Robyn's two convergence criteria (stability of the objective distribution across late iterations). The prediction-error objective (NRMSE) converged. The business-plausibility objective (DECOMP.RSSD) did *not* fully converge within the iterations run.

**What non-convergence means here.** The selected model should be read as provisional on the business-plausibility axis. In a production engagement, the search would be re-run with additional iterations until both objectives converge before findings were finalized. For this illustrative report, the estimate is presented with this caveat attached.

### 6.2 Out-of-sample performance

**Disclosure.** This illustrative model reports in-sample fit only. A production engagement would enable time-series validation (a chronological train/validation/test split) and report out-of-sample error, which is the honest measure of predictive accuracy. In-sample fit alone overstates performance, and the estimates in this sample should be read with that in mind.

### 6.3 Calibration

No incrementality experiment was available to calibrate the model. Without a holdout or lift test to anchor the absolute effect sizes, the *relative ranking* of channels is more reliable than the *absolute* ROI levels. This is the single most important reason the recommendation specifies a holdout test as its first validation step.



## 7. Sensitivity & Robustness

We tested how much the conclusions depend on the specific modeling choices made.

### 7.1 Parameter stability

Carryover estimates were checked for stability across resampled fits. Most channels were stably estimated. **Television was not**: its carryover estimate varied widely across resamples, and because television carries a large share of attributed revenue, its contribution should be read as a broad range rather than a point estimate.

### 7.2 Alternative specifications

The model was refit under alternative assumptions (different carryover forms, inclusion/exclusion of controls). The estimated *magnitude* of the revenue gain moved — from roughly 4% to 16% — but the *direction* of the recommendation did not: paid search and television remained the highest-return channels and out-of-home the lowest under every specification tested.

**The distinction that matters.** The exact numbers are uncertain. The decision is not. Whichever reasonable model is chosen, the recommendation is to move budget out of out-of-home and print and into search and television. That directional robustness is the basis for acting, while the magnitude range is the basis for testing before committing fully.

### 7.3 Marginal-return convention

Marginal-return figures in this report follow Robyn's allocator convention (the value of the next dollar given carryover already in the system). This is stated for consistency: an alternative convention modeling sustained future spend would produce larger figures, and the two must not be mixed.

## 8. Limitations

Consolidated in plain language. None of these invalidate the analysis; together they define exactly how far it can be trusted.

1. **Correlational, not causal.** The model identifies channels *associated* with incremental revenue. It does not prove causation. Where budgets are set in anticipation of demand, estimated effects are biased upward. *Remedy: the recommended holdout test.*
2. **In-sample fit only (this sample).** Predictive accuracy is overstated without out-of-sample validation, which a production engagement would include.
3. **Incomplete convergence.** The business-plausibility objective did not fully converge; the selected model is provisional on that axis.
4. **Uncalibrated.** No lift experiment anchors the absolute effect sizes; trust the channel ranking more than the absolute ROI levels.
5. **One channel poorly identified.** Television's carryover, and therefore its contribution, carries wide uncertainty.
6. **Revenue, not profit.** Without per-SKU cost data, recommendations optimize revenue, not margin. A channel selling low-margin products could look attractive on revenue and be less so on profit.
7. **Unmodeled channels.** Any spend outside the supplied data (retail media, creator partnerships) is invisible to the model and its effect misattributed.
8. **Stable-relationship assumption.** The model assumes a dollar on a channel meant the same thing throughout the window.



### 8.1 What would strengthen a future engagement

- A geographic or conversion-lift holdout on paid search, supplied back to the model as calibration.
- Full spend history including all channels and any off-platform marketing cost.
- Per-SKU cost data to convert revenue recommendations into profit (POAS) recommendations.
- Additional history, which narrows the uncertainty on every channel-level parameter.

## 9. Recommended Next Steps

1. **Approve the reallocation** in principle — it requires no additional budget and its direction is robust.
2. **Run a two-week holdout test** on paid search to confirm the modeled effect is causal before committing the full shift.
3. **Share complete data** — full spend history and per-SKU costs — to enable a calibrated, margin-aware refit.
4. **Refit and review** on a regular cadence as new data accumulates, tracking blended ROAS and profit margin over time.

Prepared by The E[X] Group · hello@theexgroup.com

## A. Technical Appendix: Mathematics & Methods

This appendix gives the full mathematical specification behind the model. Notation: $x_{c,t}$ is raw spend for channel $c$ in week $t$; $A_{c,t}$ its adstocked value; $S_{c,t}$ its saturated value; $\theta_c, \alpha_c, \gamma_c$ its carryover and saturation parameters; $\beta_c$ its response coefficient; $Z_{j,t}$ a control (baseline) regressor with coefficient $\delta_j$; $y_t$ observed revenue and $\hat{y}_t$ fitted revenue. All formulas are as implemented in Robyn (v3.12.1): `transformation.R`, `model.R`, `allocator.R`.

### A.1 Coefficient fitting: constrained ridge regression

For a *fixed* set of transform hyperparameters, the design matrix $X = [\, S_1, \ldots, S_C,\ Z_1, \ldots, Z_J \,]$ is fully determined, and the coefficients solve a penalized least-squares problem. Robyn calls `glmnet` with $\alpha = 0$ (pure ridge), a non-negativity constraint on media coefficients, and per-coefficient penalty factors $p_k$:

$$\hat{\beta} \;=\; \arg\min_{\beta_0,\beta} \underbrace{\frac{1}{2n}\sum_{t=1}^{n}\Big(y_t - \beta_0 - \textstyle\sum_k X_{tk}\beta_k\Big)^2}_{\text{mean squared error}} + \underbrace{\frac{\lambda}{2}\sum_k p_k\beta_k^2}_{\text{ridge penalty}} \quad \text{s.t.} \quad \beta_c \ge 0 \;\; \forall c \in \text{media}. \tag{1}$$

The $\frac{1}{2n}$ scaling means the effective penalty is $n\lambda$; the intercept $\beta_0$ is unpenalized; predictors are standardized internally.

**Lagrangian and KKT conditions.** The penalty is itself the Lagrangian form of a constraint on coefficient magnitude: minimizing (1) is equivalent to minimizing the squared error subject to $\sum_k p_k\beta_k^2 \le s(\lambda)$ for a corresponding budget $s$. Writing the media non-negativity constraints with multipliers $\mu_c \ge 0$, the Lagrangian is

$$\mathcal{L}(\beta,\mu) = \frac{1}{2n}\|y - \beta_0 - X\beta\|_2^2 + \frac{\lambda}{2}\sum_k p_k\beta_k^2 - \sum_c \mu_c\beta_c. \tag{2}$$

The KKT conditions characterizing the optimum are

$$\begin{aligned} \text{stationarity:}\quad & -\tfrac{1}{n}X_k^\top(y - \beta_0 - X\beta) + \lambda p_k\beta_k - \mu_k\mathbf{1}[k \in \text{media}] = 0, & (3)\\ \text{primal feasibility:}\quad & \beta_c \ge 0, & (4)\\ \text{dual feasibility:}\quad & \mu_c \ge 0, & (5)\\ \text{complementary slackness:}\quad & \mu_c\,\beta_c = 0. & (6) \end{aligned}$$

Complementary slackness is the operative condition: for each media channel, either the constraint is slack ($\beta_c > 0$, $\mu_c = 0$, and the channel's coefficient is determined by the regression) or it binds ($\beta_c = 0$, $\mu_c \ge 0$, and the channel is pinned to zero). A channel whose unconstrained estimate would be negative is set exactly to zero — which is why the model cannot report a harmful media effect, only a null one.

**Why ridge rather than OLS.** Media channels are frequently collinear (budgets move together). Ordinary least squares then yields unstable, high-variance coefficients that can flip sign. The ridge penalty trades a small bias for a large reduction in variance, stabilizing the attribution. Because pure ridge ($\alpha = 0$) shrinks but does not zero coefficients, exact zeros arise only from the non-negativity constraint above, not from the penalty.

### A.2 Baseline / media identity

The fitted revenue decomposes additively into a baseline and a media contribution:

$$\hat{y}_t = \underbrace{\beta_0 + \sum_j \delta_j Z_{j,t}}_{\text{baseline}_t} + \underbrace{\sum_c \beta_c S_{c,t}}_{\text{media}_t}, \tag{7}$$

and each channel's *effect share* — central to the model-selection objective in A.3 — is

$$\mathrm{eff}_c = \frac{\beta_c \sum_t S_{c,t}}{\sum_{c'} \beta_{c'} \sum_t S_{c',t}}. \tag{8}$$

### A.3 Hyperparameter search: two-objective evolutionary optimization

The transform hyperparameters $\Theta = \{\theta_c, \alpha_c, \gamma_c\}_c \cup \{\lambda, \text{train\_size}\}$ cannot be fit in closed form: the adstock recursion, the Hill nonlinearity, and the constrained ridge solve compose into an objective with no usable analytic gradient. Robyn searches $\Theta$ with **Nevergrad** (default algorithm `TwoPointsDE`, a two-points differential-evolution variant) against *two* objectives evaluated for each candidate model.

**Objective 1 — prediction error (NRMSE).**

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_t (y_t - \hat{y}_t)^2}}{\max_t(y_t) - \min_t(y_t)}. \tag{9}$$

When time-series validation is enabled this is computed on the validation split ($\mathrm{NRMSE}_{\mathrm{val}}$); otherwise on the training data.

**Objective 2 — business plausibility (DECOMP.RSSD).**

$$\mathrm{DECOMP.RSSD} = \sqrt{\sum_c \left(\mathrm{eff}_c - \mathrm{spend}_c\right)^2} \times \left(1 + \frac{\#\{c : \mathrm{eff}_c = 0\}}{C}\right), \tag{10}$$

where $\mathrm{spend}_c$ is channel $c$'s share of total spend and $\mathrm{eff}_c$ its effect share (8). This term is the squared distance between how budget is *spent* and how effect is *attributed*; the trailing factor additionally penalizes models that zero out channels. It encodes a prior that a channel's contribution should not stray too far from its spend share — a guard against over-attributing to small channels, at the cost of biasing the search against finding a large channel ineffective.

**Ask–tell loop.** Nevergrad proceeds by (i) *asking* for a candidate $\Theta$, (ii) letting Robyn apply the transforms, solve the constrained ridge (1), and compute the objective pair, and (iii) *telling* Nevergrad the tuple $\big(\mathrm{NRMSE}(\Theta), \mathrm{DECOMP.RSSD}(\Theta)\big)$, which updates the population it samples from. With `trials` $\times$ `iterations` runs (e.g. $5 \times 2000 = 10{,}000$), the search explores the $\sim 3C + 2$ dimensional space.

**Pareto optimality.** With two competing objectives there is no single minimizer. A model $\Theta$ is *Pareto-optimal* if no other model improves one objective without worsening the other:

$$\mathcal{P} = \Big\{\Theta : \nexists\, \Theta' \text{ with } f_i(\Theta') \le f_i(\Theta)\ \forall i \text{ and } f_j(\Theta') < f_j(\Theta) \text{ for some } j\Big\}, \tag{11}$$

with $f_1 = \mathrm{NRMSE}$, $f_2 = \mathrm{DECOMP.RSSD}$. Robyn accumulates successive Pareto fronts until it has a minimum number of candidate models, clusters them by response profile, and presents the front for a final human selection. If a calibration experiment is supplied, a third objective — the mean absolute percentage error between modeled and experimentally measured lift — is added, and optimization proceeds over the three-dimensional objective space identically.

### A.4 Adstock forms: geometric and Weibull

Robyn offers three carryover forms. The reported model uses geometric; the Weibull forms were used in the robustness checks and are included here for completeness.

**Geometric (one parameter).**

$$A_t = x_t + \theta A_{t-1}, \qquad A_t = \sum_{j=0}^{t-1} \theta^j x_{t-j}, \qquad \theta \in [0,1). \tag{12}$$

The peak effect is fixed at lag 0; memory decays at a constant rate. Half-life $t_{1/2} = \ln(0.5)/\ln\theta$.

**Weibull decay.** Both Weibull forms weight spend by a time-varying kernel $w(\cdot)$ with a *shape* parameter $k$ and a *scale* parameter $\lambda_s$, allowing the peak effect to occur *after* the spend period — realistic for brand media whose impact builds:

$$A_t = \sum_{\tau=0}^{t-1} w(\tau)\, x_{t-\tau}, \qquad w(\tau) \text{ normalized to } w(0) = 1 \text{ over the window.} \tag{13}$$

*Weibull CDF form* — the kernel is one minus the Weibull cumulative distribution function, giving a flexible monotonic decay whose steepness and timing are controlled by $(k, \lambda_s)$:

$$w_{\text{CDF}}(\tau) = 1 - F(\tau; k, \lambda_s), \qquad F(\tau; k, \lambda_s) = 1 - \exp\Big(-\big(\tau/\lambda_s\big)^k\Big). \tag{14}$$

So $w_{\text{CDF}}(\tau) = \exp\big(-(\tau/\lambda_s)^k\big)$. For $k = 1$ this reduces to geometric-like exponential decay; $k > 1$ produces a slower initial decay then a sharper fall; $k < 1$ an immediate steep drop.

*Weibull PDF form* — the kernel is the Weibull probability density, which *can rise before it falls*, placing the peak influence at a lag $> 0$:

$$w_{\text{PDF}}(\tau) \propto f(\tau; k, \lambda_s), \qquad f(\tau; k, \lambda_s) = \frac{k}{\lambda_s}\Big(\frac{\tau}{\lambda_s}\Big)^{k-1} \exp\Big(-\big(\tau/\lambda_s\big)^k\Big). \tag{15}$$

Behavior by shape: $k \le 1$ gives a peak at $\tau = 0$ (monotonic decay); $1 < k < 2$ a delayed peak with an infinite initial slope; $k > 2$ a delayed, bell-shaped carryover with zero initial slope — the most flexible option, and the one Robyn recommends when calibrating to experiments. The cost is a second hyperparameter per channel, roughly doubling the search dimension for those channels and requiring more iterations to converge. The mode (peak lag) of the PDF form is $\lambda_s\big(\frac{k-1}{k}\big)^{1/k}$ for $k > 1$.



## A.5 Saturation: the Hill function

$$S_t = \frac{A_t^\alpha}{A_t^\alpha + \kappa^\alpha}, \qquad \kappa = \gamma \max_t(A_t), \qquad \alpha \in [0.5, 3],\ \gamma \in [0.3, 1]. \tag{16}$$

$\alpha$ controls curvature: $\alpha > 1$ gives an S-shape (a spend threshold before traction), $\alpha \le 1$ a C-shape (diminishing returns from the first dollar). $\gamma$ places the half-saturation point $\kappa$ as a fraction of the largest adstocked value observed. The inflection (steepest) point of the curve, where it exists ($\alpha > 1$), sits at $A^* = \kappa\big(\frac{\alpha-1}{\alpha+1}\big)^{1/\alpha}$, strictly below $\kappa$.

## A.6 Budget allocation: constrained nonlinear optimization

Given a fitted model, the optimal reallocation of a fixed budget $\bar{B}$ solves a constrained nonlinear program. Writing each channel's steady response in the form Robyn's allocator uses — adstock entering as an additive historical-carryover offset $\bar{h}_c$, so $A_c = s_c + \bar{h}_c$ and $dA_c/ds_c = 1$:

$$\max_{s_1,\dots,s_C} \sum_c \beta_c \Big(1 + \frac{\kappa_c^{\alpha_c}}{(s_c + \bar{h}_c)^{\alpha_c}}\Big)^{-1} \quad \text{s.t.} \quad \sum_c s_c = \bar{B}, \quad \ell_c \bar{s}_c \le s_c \le u_c \bar{s}_c, \tag{17}$$

where $(\ell_c, u_c)$ are the per-channel bounds (e.g. 0.5 and 1.5 times current spend $\bar{s}_c$). The bracket is the Hill function rearranged.

**Marginal ROI and the Lagrangian condition.** Differentiating each channel's response gives the marginal return of the next dollar:

$$\mathrm{mROI}_c(s_c) = \frac{\partial R_c}{\partial s_c} = \beta_c \, \frac{\alpha_c \, \kappa_c^{\alpha_c} \, (s_c + \bar{h}_c)^{\alpha_c - 1}}{\big((s_c + \bar{h}_c)^{\alpha_c} + \kappa_c^{\alpha_c}\big)^2}. \tag{18}$$

Forming the Lagrangian $\mathcal{L} = \sum_c R_c(s_c) - \eta\big(\sum_c s_c - \bar{B}\big)$ with budget multiplier $\eta$, the first-order (KKT) condition is

$$\boxed{\mathrm{mROI}_c(s_c) = \eta \quad \text{for every channel at an interior optimum.}} \tag{19}$$

*Every channel receiving budget must earn the same marginal return.* If one channel's next dollar earns more than another's, shifting budget toward it raises total revenue — so at the optimum all interior channels are equalized, and any channel at a bound has marginal return on the far side of $\eta$. This single condition is the entire economic content of the reallocation. Robyn solves (17) numerically with a sequential least-squares quadratic programming routine (`NLOPT_LD_SLSQP`), using the analytic gradient (18).

**Note on convention.** Equation (18) uses $dA/ds = 1$ (next-dollar-given-carryover). An alternative convention modeling *sustained* weekly spend gives $A = s/(1-\theta)$, $dA/ds = 1/(1-\theta)$, and hence a marginal return larger by a factor $1/(1-\theta)$ — up to tenfold at $\theta = 0.9$. All figures in this report use Robyn's allocator convention; the two must never be mixed.

### A.7 Baseline decomposition: Prophet

Before media effects are attributed, Robyn uses Prophet (Taylor & Letham, 2018) to represent the non-media structure of revenue as a decomposable additive model:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t, \tag{20}$$

with trend $g(t)$, seasonality $s(t)$, holidays $h(t)$, and noise $\varepsilon_t$. These enter the media model as the control regressors $Z_{j,t}$, so that variation explained by season or trend is not misattributed to advertising.

**Trend $g(t)$ — piecewise linear with changepoints.** The default trend is piecewise linear. A set of $S$ changepoints at times $s_1, \ldots, s_S$ each carry a rate adjustment $\delta_j$, collected in $\boldsymbol{\delta}$, with a vector $\mathbf{a}(t) \in \{0,1\}^S$ indicating which changepoints have occurred by $t$:

$$g(t) = \big(k + \mathbf{a}(t)^\top \boldsymbol{\delta}\big)\, t + \big(m + \mathbf{a}(t)^\top \boldsymbol{\gamma}\big), \tag{21}$$

where $k$ is the base growth rate, $m$ the offset, and $\gamma_j = -s_j \delta_j$ keeps the trend continuous at each changepoint. A sparse (Laplace) prior $\delta_j \sim \mathrm{Laplace}(0, \varsigma)$ regularizes the rate changes: most stay near zero, and a changepoint activates only where the data demands it, with $\varsigma$ (the `changepoint_prior_scale`) controlling flexibility. A logistic-growth alternative $g(t) = C/\big(1 + \exp(-k(t-m))\big)$ is available when the series saturates.

**Seasonality $s(t)$ — Fourier series.** Each periodic pattern of period $P$ (e.g. $P = 52.18$ weeks for yearly seasonality) is approximated by a Fourier sum of order $N$:

$$s(t) = \sum_{n=1}^{N} \Big[ a_n \cos\Big(\tfrac{2\pi n t}{P}\Big) + b_n \sin\Big(\tfrac{2\pi n t}{P}\Big) \Big], \tag{22}$$

whose $2N$ coefficients $\{a_n, b_n\}$ are estimated with the rest of the model. A larger $N$ captures more intricate seasonal shape at the risk of overfitting; multiple seasonalities (weekly, yearly) are summed.

**Holidays $h(t)$.** Holidays and events enter as indicator regressors over their dates $D_\ell$ with fitted effects $\kappa_\ell$:

$$h(t) = \sum_{\ell} \kappa_\ell \, \mathbf{1}\big(t \in D_\ell\big), \tag{23}$$

optionally with windows around each date. Prophet is estimated in a Bayesian framework, which selects changepoints and quantifies component uncertainty; the fitted components are what Robyn passes to the media regression as trend, season, and holiday controls.

### A.8 End-to-end summary

The complete pipeline, in sequence:

1. **Prophet** decomposes revenue into trend + seasonality + holidays → baseline controls $Z_{j,t}$ (A.7).
2. For each candidate $\Theta$ proposed by **Nevergrad**: apply **adstock** (geometric or Weibull, A.4) then **Hill saturation** (A.5) to every channel's spend.
3. Fit channel coefficients by **constrained ridge** (1), solving the KKT system A.1.
4. Score the model on **NRMSE** and **DECOMP.RSSD**; **tell** Nevergrad; iterate (A.3).
5. Collect the **Pareto front**, cluster, and select one model.
6. Solve the **budget allocation** program (17) for the reallocation, equalizing marginal ROI (18) (A.6).