J-INTECH (Journal of Information and Technology)
Accredited Sinta 4 Ministry of Higher Education, Science and Technology
Republic of Indonesia SK No. 10/C/C3/DT.05.00/2025
E-ISSN: 2580-720X || P-ISSN: 2303-1425

From Small Data to MMM-Style Budget Decisions: Reproducible Conjugate Bayesian Log-Response Regression with ElasticNet Benchmarks and Explainable Budget Scenario Rollouts

Yifei Lu[1*], Jinyi Mu[2], Arthur Lefebvre[3]

[1,2,3] *Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA*

Keywords
Bayesian linear regression ; budget allocation; ElasticNet; Marketing Mix Modeling; sensitiviy analysis

Corresponding Author.
yifei.lu0816@outlook.com

Abstract

Marketing mix modeling (MMM) is often treated as a large-sample, time-series task, but many organizations must make near-term budget decisions with limited historical data. This study develops a reproducible small-sample MMM-style workflow using the public ISLR Advertising.csv dataset (n=200 markets; TV, radio, newspaper spend; sales). Reproducibility is ensured by using a public dataset, fixed random seeds, explicitly stated preprocessing, analysis scripts supplied with the manuscript, and specified Python libraries. The workflow combines predictive benchmarking and prescriptive decision support. First, OLS, Ridge, Lasso, and ElasticNet are evaluated under 5-fold cross-validation and a fixed 80/20 hold-out split. Second, a conjugate Bayesian regression on log(1+spend) features models diminishing returns and yields closed-form posterior and Student-t predictive distributions. Third, for each total budget B, the allocation that maximizes the posterior mean prediction is solved under non-negativity, budget-balance, and observed-maximum channel caps; posterior samples are then evaluated at each optimized allocation to form 90% credible bands for the budget-sales curve. On the fixed split, test RMSE ranges from 1.829 to 1.871, while 100 repeated splits indicate that raw-feature OLS is most accurate on average (mean RMSE 1.671). At B=200, the bounded optimum allocates \$147.25k to TV, \$49.60k to radio, and \$3.15k to newspaper, predicting 18.149 sales. The results suggest that, in small samples, regularization and log-response modeling mainly support stable, interpretable budget recommendations rather than improving point prediction alone.

1. Introduction

A practical marketing question appears deceptively simple: "How should we split the budget across channels?" Yet answering it credibly requires more than intuition. It requires a model that links spend to outcomes, a method for translating model parameters into allocations under constraints, and a communication layer that explains why the recommended plan makes sense. MMM is the umbrella term for these activities: estimating

the effect of different marketing actions on sales and using the estimates to inform future budgets, consistent with market response and media mix modeling literature that links marketing effects measurement to allocation decisions [1], [2], [3]. Traditionally, MMM systems use time-series data with many periods and covariates, enabling adstock/carryover, seasonality controls, competitive activity, and nonlinear saturation curves.

Many real deployments do not have that luxury. Small and midsize businesses frequently have only a cross-sectional snapshot, short-lived campaigns, or data collected at low frequency. Even large organizations often start with small pilot datasets to validate measurement pipelines before scaling. In these settings, full-scale MMM is infeasible, but the decision problem still exists. This motivates a "small-data MMM-style" workflow. In this study, "small-data" is defined operationally as a setting with limited sample size and no long time-series history; specifically, the empirical demonstration uses n=200 cross-sectional observations and three media-spend predictors rather than a production-scale weekly or monthly MMM panel. The workflow is (i) rigorous about empirical evaluation, (ii) transparent about assumptions and constraints, and (iii) capable of producing actionable sensitivity analysis rather than only regression coefficients [1], [2], [3].

This paper uses the Advertising dataset from the statistical learning literature as a controlled testbed [4], [5]. The dataset contains 200 markets with spending in three channels (TV, radio, newspaper) and observed sales. Although it is not a time series, it has become a canonical example for multiple regression, variable selection, and prediction in marketing-like contexts [4]. Our choice is deliberate: the dataset has a limited sample size (n=200), is widely used, and is easy to reproduce, making it an appropriate benchmark for demonstrating that "budget allocation can be written like MMM" even when only a small cross-sectional dataset is available.

We design the workflow around a separation of concerns. The predictive layer aims to answer: "How well can we predict sales from spends?" We evaluate OLS and regularized linear models, including ElasticNet. The correlation matrix in Table 2 shows that radio and newspaper have a moderate correlation (0.354), indicating that some campaign variables covary; ElasticNet is relevant in such settings because its combined L1-L2 penalty can stabilize estimation and encourage grouped behavior among correlated predictors [6][7][8][9]. Because small datasets are sensitive to split randomness, we report both a fixed hold-out split and a repeated-split robustness study [10],[11][12][13].

The prescriptive layer aims to answer: "Given a total budget B, what allocation maximizes predicted sales, and how does expected sales change when B changes?" This requires solving an optimization problem, not merely fitting a regression. Purely linear response functions yield brittle corner solutions: the optimizer assigns all budget to the channel with the largest coefficient. MMM practice therefore employs diminishing returns through concave or saturating response curves, such as log transforms or Hill functions, and further bounds spends by operational or empirical-support constraints [2], [3], [14], [15]. We adopt a transparent concave mapping $\varphi(s)=\log(1+s)$, which is simple, interpretable, and yields a concave objective suitable for convex optimization [16].

Uncertainty is unavoidable in small-data decision making. A point estimate of coefficients does not provide a risk-aware budget plan. We therefore fit a conjugate Bayesian regression model with a normal–inverse-gamma prior, producing closed-form posteriors and Student-t predictive distributions [17]. These posterior samples propagate naturally to uncertainty bands on the budget sensitivity curve. Finally, because business decisions depend on stakeholder trust, we add an explainable scenario rollout layer: a deterministic, schema-based narrative generator that turns numeric plans into concise, audit-ready explanations grounded in computed marginal returns and posterior intervals. No external LLM is used. The generator is rule-based, but it is specialized for this decision setting because each sentence is linked to a predefined plan schema, including outcome differences, allocation shifts, marginal-return comparisons, and uncertainty intervals [18], [19].

In summary, this paper contributes: (1) full experimental evaluations on Advertising.csv with detailed figures and tables; (2) a reproducible Bayesian log-response model and uncertainty propagation; (3) a constrained budget optimizer that outputs budget–sales sensitivity curves and allocation stability analysis; (4) an explanation layer that generates consistent scenario narratives; and (5) a careful treatment of feasibility and extrapolation via per-channel caps based on observed data ranges. The remainder of the paper reviews related work, describes methods, and reports empirical results, positioning the proposed workflow within marketing response modeling, media allocation, and marketing analytics research [1], [2], [3], [14] [15][20][21].

2. Research Method

This section defines the dataset, the experimental protocol, the predictive baselines, the conjugate Bayesian log-response model, the constrained budget optimization problem, the sensitivity analysis outputs, and the deterministic scenario rollout layer. All numerical results and all figures are generated from the dataset and the stated procedures.

2.1. Dataset and Notation

We use Advertising.csv distributed with ISLR resources [4], [5]. Each observation corresponds to a market. Let $x_i = (TV_i, Radio_i, Newspaper_i)$ denote channel spending in thousands of dollars and y_i denote sales in thousands of units. The dataset contains $n=200$ observations and $p=3$ channels.

We remove the first column, which is an index, and treat the remaining columns as numeric. No additional feature engineering, scaling, imputation, or row deletion is applied at this stage. We verified that the modeling variables contain no missing values in the Advertising.csv file, so all n=200 observations are retained for descriptive analysis and subsequent modeling. Table 1 reports summary statistics and Table 2 reports the correlation structure. TV and radio spending exhibit strong positive associations with sales, while newspaper spending is weakly correlated with sales. Radio and newspaper show moderate correlation (0.354), consistent with the intuition that some channels covary in campaign intensity.

We evaluate two feature representations: raw spend features X_raw = [TV, Radio, Newspaper], and a concave transformation used for planning X_log = [log(1+TV), log(1+Radio), log(1+Newspaper)]. The log transform has two roles. First, it reduces the influence of extreme spends, which can stabilize estimation on small samples. Second, it encodes diminishing returns: the marginal increase in log(1+s) is 1/(1+s), which decreases as spend increases. This use of concave or saturating response transformations is consistent with MMM literature on advertising shape effects and diminishing returns [2], [3]. This property is central for producing non-corner solutions in budget optimization.

2.2. Experimental Protocol and Metrics

We conduct two complementary evaluation protocols.

Fixed hold-out benchmark. We split the data into 80% training (n=160) and 20% testing (n=40) using random_state=2026. This value has no statistical significance; it is an arbitrary but fixed seed chosen only to make the train/test split exactly reproducible. We use only the training set for hyperparameter tuning and report metrics on the held-out test set. Table 3 confirms that training and test distributions are comparable.

Repeated-split robustness. To quantify variability due to sample splitting, we run 100 random 80/20 splits using ShuffleSplit (random_state=2026) and report the mean and standard deviation of test RMSE and R2. This procedure measures expected predictive performance and its stability.

We report four metrics, defined as follows:

$$RMSE = \sqrt{1/n \sum_{i=1} n(y_i - \hat{y}_i)^2} \quad (1)$$

$$MAE = 1/n \sum_{i=1} n|y_i - \hat{y}_i| \quad (2)$$

$$R^2 = 1 - \sum_{i=1} n(y_i - \hat{y}_i)^2 / \sum_{i=1} n(y_i - \bar{y})^2 \quad (3)$$

For the Bayesian model, we also report average test log predictive density (LPD) under the Student-t predictive distribution. These metrics are standard in regression evaluation and are widely used to summarize prediction error and goodness of fit in supervised regression models [9], [22], [23].

2.3. Descriptive Statistics

Before turning to predictive and prescriptive modeling, this subsection characterizes the dataset itself. We summarize the distribution of each variable, examine how the channels relate to one another and to sales, and confirm that the fixed train/test split used throughout the paper is balanced. These descriptive results motivate two later design choices: the use of regularization to handle correlated predictors, and the use of a fixed, reproducible split for benchmarking. Table 1 summarizes the distribution of budgets and sales. Total budgets vary substantially across markets, with TV spending ranging from 0.7 to 296.4 and radio spending ranging from 0.0 to 49.6. Newspaper spending shows the widest relative spread among the three channels (mean 30.554, standard deviation 21.779), while sales itself ranges from 1.600 to 27.000 thousand units with a mean of 14.023. The gap between each variable's mean and its median (50%) is modest for all four columns, indicating that the distributions are reasonably symmetric rather than heavily skewed by a small number of extreme-spend markets.

Table 1. Descriptive statistics of Advertising.csv (n=200).

Variable	mean	std	min	25%	50%	75%	max
TV	147.042	85.854	0.700	74.375	149.750	218.825	296.400
Radio	23.264	14.847	0.000	9.975	22.900	36.525	49.600
Newspaper	30.554	21.779	0.300	12.750	25.750	45.100	114.000
Sales	14.023	5.217	1.600	10.375	12.900	17.400	27.000

Table 2 reports the Pearson correlation matrix, which motivates the inclusion of regularization methods to stabilize coefficients and support interpretable inference when predictors are correlated. TV shows the strongest association with sales (r=0.782), followed by radio (r=0.576), while newspaper is only weakly associated with sales (r=0.228). Among the predictors themselves, TV is nearly uncorrelated with radio (r=0.055) and newspaper (r=0.057), but radio and newspaper show a moderate correlation (r=0.354). This pattern, two largely independent strong channels alongside one weak channel that mildly covaries with radio, is the prototypical structure that motivates comparing OLS against Ridge, Lasso, and ElasticNet in the next subsection.

Table 2. Pearson correlation matrix (Advertising.csv).

Variable	TV	Radio	Newspaper	Sales
TV	1.000	0.055	0.057	0.782
Radio	0.055	1.000	0.354	0.576
Newspaper	0.057	0.354	1.000	0.228
Sales	0.782	0.576	0.228	1.000

Table 3 reports the mean and standard deviation of each variable in the training and test sets for the fixed split. The split maintains similar ranges and variances, making it suitable for comparative evaluation. The test set

has a somewhat lower mean for every variable than the training set (for example, mean sales of 12.627 versus 14.371), but the standard deviations remain close across the two sets (e.g., TV standard deviation of 82.580 in test versus 85.633 in train). This indicates that random_state=2026 happened to place a slightly lower-spend, lower-sales subset of markets into the test fold, while preserving comparable spread, so the fixed split remains a fair basis for benchmarking model performance.

Table 3. Train/test summary for fixed 80/20 split (random_state=2026).

Variable	train_mean	test_mean	train_std	test_std
TV	153.649	120.617	85.633	82.580
Radio	23.724	21.425	14.929	14.551
Newspaper	30.741	29.807	21.733	22.221
Sales	14.371	12.627	5.330	4.537

The x-axis reports channel spend in thousands of dollars, and the y-axis reports sales in thousands of units; points show individual markets, and lines show fitted univariate OLS trends for visual comparison. Figure 1 visualizes the bivariate relationships summarized numerically in Table 2. The TV panel shows a clear upward trend with relatively tight scatter around the fitted line, consistent with its strong correlation with sales. The radio panel shows a similar but noisier upward trend, while the newspaper panel shows a nearly flat fitted line with wide scatter, visually confirming that newspaper spend carries little linear association with sales on its own. These univariate patterns motivate the multivariate models considered next, since none of the three channels alone explains the full variation in sales.

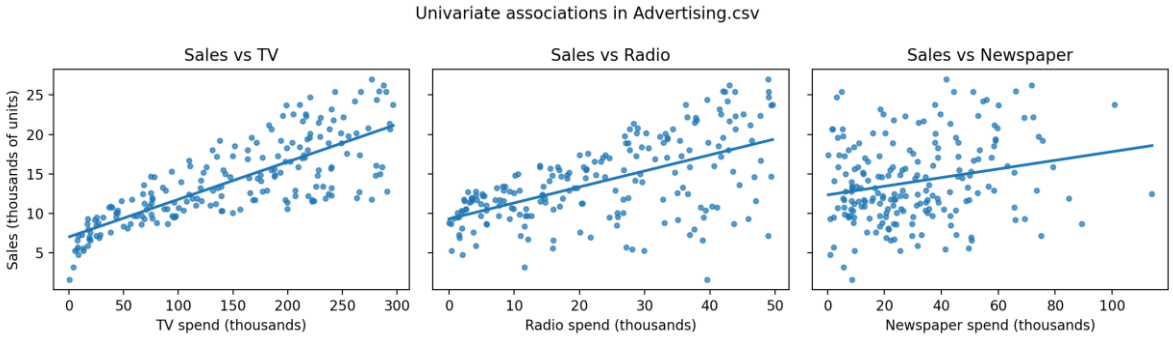

Figure 1. Sales vs. each channel in Advertising.csv.

2.4. Predictive Baselines: OLS and Regularization

We benchmark OLS, Ridge, Lasso, and ElasticNet on both feature representations. OLS fits β by minimizing $\Sigma_i (y_i - \beta_0 - x_i^T\beta)^2$. Ridge minimizes the same objective plus $\alpha\|\beta\|_2^2$, shrinking coefficients continuously [8]. Lasso uses $\alpha\|\beta\|_1$, inducing sparse solutions [7]. ElasticNet combines both penalties $\alpha(\rho\|\beta\|_1 + (1-\rho)\|\beta\|_2^2/2)$ [6].

We tune α over 25 values logarithmically spaced from 10^−3 to 10^3. For ElasticNet, we also tune $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. Within each cross-validation fold, we standardize predictors by subtracting the training-fold mean and dividing by the training-fold standard deviation, so that the penalty is applied on a comparable scale across predictors [9]. We implement all models using scikit-learn [23].

The evaluation for predictive baselines serves two purposes. First, it establishes how much predictive accuracy is available in the dataset without strong structural assumptions. Second, it reveals whether the log transformation improves or degrades prediction relative to raw spends, informing how we separate the predictive layer from the prescriptive planning layer.

2.5. Bayesian Log-Response Regression (Conjugate NIG)

For budget optimization, we require a response function that is concave in spend and admits uncertainty quantification. We therefore use a Bayesian linear regression on log-response features.

Model. Let $X \in \mathbb{R}^n \times (p+1)$ be the design matrix with an intercept and log-response predictors. The likelihood is $y \mid \beta, \sigma^2 \sim N(X\beta, \sigma^2 I)$. We use the conjugate normal–inverse-gamma prior $\beta \mid \sigma^2 \sim N(m_0, \sigma^2 V_0)$, $\sigma^2 \sim \text{Inv-Gamma}(a_0, b_0)$. We set $m_0 = 0$, $V_0 = 100I$, $a_0 = 2$, $b_0 = 1$. These values define a weakly informative but proper prior: m_0=0 centers coefficients before observing the data, V_0=100I gives relatively diffuse prior variance on the coefficient scale, and a_0=2, b_0=1 provide a proper inverse-gamma prior for σ^2. The prior is used for regularization and uncertainty quantification rather than for encoding strong domain beliefs. No separate prior-sensitivity experiment was used for model selection; therefore, posterior results should be interpreted conditional on this weakly informative prior specification [22], [17].

Posterior. Conjugacy implies the posterior is also normal–inverse-gamma with parameters:

$$V_n = \left(V_0^{-1} + X^T X\right)^{-1}, \tag{4}$$

$$m_n = V_n\left(V_0^{-1} m_0 + X^T y\right), \tag{5}$$

$$a_n = a_0 + n/2, \tag{6}$$

$$b_n = b_0 + 0.5\left(y^T y + m_0^T V_0^{-1} m_0 - m_n^T V_n^{-1} m_n\right). \tag{7}$$

These expressions follow standard derivations [9].

Predictive distribution. For a new x_* (including intercept), the posterior predictive distribution is Student-t with df=2an, mean μ_* = x_*^T mn, and scale^2 proportional to (bn/an)(1 + x_*^T Vn x_*). We use this distribution to compute log predictive density on held-out data and to generate uncertainty bands for budget sensitivity curves.

Sampling. To propagate uncertainty through the optimizer and sensitivity analysis, we draw 5,000 posterior samples. Because these are direct independent samples from the closed-form normal-inverse-gamma posterior rather than MCMC draws, Markov-chain convergence diagnostics are not required. The number 5,000 was chosen as a practical Monte Carlo size for estimating posterior means and 5th/95th predictive percentiles with stable interval estimates. We sample σ^2 from Inv-Gamma(an, bn) and then sample β from N(mn, σ^2 Vn). Each sampled β defines a response curve that can be evaluated at a candidate allocation.

2.6. Budget Allocation Optimization with Feasibility Constraints

Decision variables. Let $s = (s_{TV}, s_{Radio}, s_{News})$ denote the decision variables, representing channel spends in thousands of dollars.

Objective. We maximize the posterior mean prediction under the Bayesian log-response model:

$$f(s) = \beta_0 + \beta_1 log(1 + s_{TV}) + \beta_2 log(1 + s_{Radio}) + \beta_3 log(1 + s_{News}). \tag{8}$$

Constraints. We impose a total budget constraint and non-negativity:

$$s_{TV} + s_{Radio} + s_{News} = B, \quad s_i \geq 0. \tag{9}$$

Per-channel caps (support constraint). Because Advertising.csv has limited range for each channel (e.g., radio spend never exceeds 49.6), small-data optimization must avoid out-of-support extrapolation. We therefore impose upper bounds equal to observed maxima:

$$0 \le s_{TV} \le 296.4, \quad 0 \le s_{Radio} \le 49.6, \quad 0 \le s_{News} \le 114.0. \tag{10}$$

These caps ensure that allocations remain within the empirical support of the dataset and that recommended spends are consistent with the observed data.

Convexity and solution. Since log(1+s) is concave for s ≥ 0, f(s) is concave. Since log(1+s) is concave for s ≥ 0, f(s) is concave. Maximizing a concave function over linear equality, inequality, and bound constraints can be equivalently written as minimizing −f(s), a convex objective, over a convex feasible set; therefore, the problem has a globally optimal solution under the stated feasible constraints [16]. The caps introduce piecewise regimes where some constraints are active (e.g., radio hits its cap). We solve the constrained problem using SLSQP (Sequential Least Squares Programming) as implemented in SciPy [24]. In all budgets reported, SLSQP terminates successfully with ftol=1e−12. In SciPy's SLSQP implementation, ftol controls the stopping precision for objective changes, step size, Lagrangian-gradient optimality, and absolute constraint violation. We set maxiter=5000 and verify that the equality constraint |Σs_i−B| is satisfied to numerical tolerance. Because the transformed problem is convex, the solution is not expected to depend on initialization; nevertheless, projected feasible initializations were checked for representative budgets and produced the same optimized allocations within numerical tolerance [16], [24].

Implementation details for optimization and heuristic projection. We solve the bounded optimization for each budget B with SciPy's SLSQP solver [24]. For numerical stability we initialize from an equal-split allocation clipped to caps and then redistribute any remainder proportional to remaining capacity. The solver tolerance is ftol=1e−12 and maxiter=5000. Across all B in the grid, the solver terminates successfully and satisfies the equality constraint to numerical tolerance.

Heuristic baselines must also respect feasibility. Simple share-based rules (e.g., 1/3–1/3–1/3) can violate the radio cap at moderate budgets, creating invalid comparisons. We therefore project each heuristic plan into the feasible set using a deterministic redistribution procedure: (i) compute target spends as shares×B, (ii) cap any channel that exceeds its limit, (iii) compute the remaining budget, and (iv) redistribute the remainder across channels with slack capacity proportional to their target shares (or equally if target shares are zero on the remaining channels). The output plan exactly satisfies the sum constraint and all caps, and is reproducible because the projection rule is deterministic and does not depend on random initialization.

Computational cost. With p=3, the 39 constrained solves for B=20…400 complete quickly. The optimizer uses posterior mean coefficients to obtain s*(B), while posterior bands are obtained afterward by evaluating 5,000 coefficient draws at these optimized allocations. Thus, the mean curve represents the posterior-mean decision rule, and the shaded band represents coefficient uncertainty around that decision rule.

2.7. Budget Sensitivity Outputs

For a grid of budgets B ∈ {20, 30, …, 400}, we compute the bounded optimal allocation s*(B) by maximizing the posterior mean prediction subject to the total-budget constraint and channel caps. We then record the corresponding posterior-mean predicted sales f(s*(B)), which forms the central "budget → optimized sales" curve. Marginal returns are computed using finite differences, Δf(B) = f(s*(B)) − f(s*(B−10)), and are reported per $1k of additional budget by dividing by 10.

Uncertainty bands. For each budget B, we evaluate f_β(s*(B)) across the 5,000 posterior samples β and extract the 5th and 95th percentiles, yielding a 90% credible band. The posterior samples are therefore used to quantify uncertainty around the optimized allocation obtained from the posterior mean, not to solve 5,000 separate allocation problems. This band captures coefficient uncertainty given the model and data.

Interpretation. The sensitivity curve answers the direct managerial question: increasing the total budget from B to B+ΔB changes optimized expected sales by approximately f(s*(B+ΔB))−f(s*(B)). The marginal return curve additionally indicates where returns begin to flatten (saturation) under the assumed response form and feasibility constraints.

2.8. Explainable Scenario Rollouts (Deterministic and Rule-Based)

Scenario definition. We define a plan schema with fields: total budget B, allocation s, predicted sales mean, credible interval, and implied marginal returns by channel. We compare an optimized plan against heuristic plans defined by target shares (e.g., balanced, TV-heavy). Each heuristic plan is projected into the feasible set using the same caps and total budget constraint.

Narrative generator. Given two plans A and B, we generate a narrative with four steps:

(1) Outcome summary: report predicted sales and credible intervals.

(2) Allocation delta: report how budget shifts across channels.

(3) Reasoning: compute and report marginal returns for channels at the operating point; identify the dominant ROI gaps.

(4) Risk: report interval widths and whether outcome ranking is preserved under uncertainty.

No actual LLM is called. The generator is rule-based, but it differs from a generic explanation template because each sentence is bound to a scenario schema and populated only with computed quantities from the posterior summaries, feasible allocations, marginal-return calculations, and credible intervals. The design therefore emphasizes traceability, numerical consistency, and auditability rather than natural-language generation capability [18], [19].

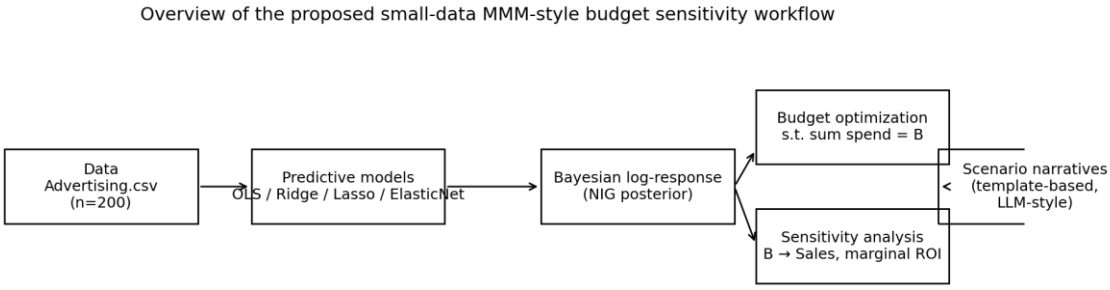

Figure 2. Overview of the proposed small-data MMM workflow

Figure 2 summarizes the end-to-end pipeline. It starts from the public Advertising.csv dataset, applies explicitly stated preprocessing, evaluates predictive baselines with cross-validation and hold-out testing, fits a Bayesian log-response model for diminishing returns, solves constrained budget allocation problems under observed channel caps, converts the optimized allocations into budget-sensitivity curves with posterior credible bands, and finally generates deterministic scenario explanations from the computed quantities.

3. Result and Discussions

We present results in four parts: (i) model selection and predictive accuracy; (ii) coefficient interpretation and Bayesian posterior summaries; (iii) budget sensitivity curves and allocation stability under per-channel

caps; and (iv) scenario comparisons with deterministic explanations. Predictive benchmarks use the fixed hold-out split unless otherwise noted. Planning analyses fit the Bayesian log-response model on all n=200 observations to maximize available information for prescriptive recommendations, consistent with common practice of refitting on full data after selecting a model class.

3.1. Cross-Validation and Hold-Out Benchmarking

Table 4 shows that cross-validation selects very small penalty values for the raw-feature models (alpha between 0.001 and 0.0056), which pulls Ridge, Lasso, and ElasticNet close to unregularized OLS and yields nearly identical mean CV RMSE (around 1.636 for all three). The log-feature models select larger penalties (alpha near 0.056 to 1.78) and show higher CV RMSE (around 2.14), reflecting that the log transformation removes some of the raw predictive signal in exchange for the concavity needed later in budget optimization. The ElasticNet l1_ratio of 0.9 in both feature spaces indicates that the selected solution leans toward the Lasso penalty rather than the Ridge penalty.

Table 4. Best 5-fold CV settings on the training split (80% train; random_state=2026).

Model	Features	alpha	l1_ratio	CV_RMSE_mean	CV_RMSE_std
Ridge_raw	Raw	0.001000		1.636424	0.375525
Lasso_raw	Raw	0.010000		1.636361	0.376164
ENet_raw	Raw	0.005623	0.900000	1.636403	0.376058
Ridge_log	log1p	1.778279		2.144373	0.227132
Lasso_log	log1p	0.100000		2.140777	0.233212
ENet_log	log1p	0.056234	0.900000	2.141887	0.227744

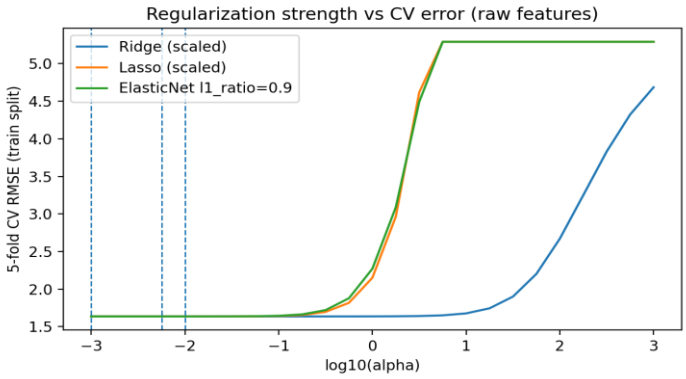

Figure 3. 5-fold CV RMSE versus log10(alpha) for raw features.

Lines show mean 5-fold CV RMSE across folds for Ridge, Lasso, and ElasticNet; fold-level standard deviations are reported numerically in Table 4 rather than plotted. Figure 3 shows that, for raw features, CV RMSE stays essentially flat across a wide range of small-to-moderate alpha values before rising sharply once the penalty becomes large enough to shrink coefficients away from their OLS values. This flat region explains why the optimal alpha selected in Table 4 is so small: the data offer little benefit from regularization beyond mild numerical stabilization, since TV and radio are only weakly correlated with each other.

Table 5. Predictive performance on the fixed hold-out split (train n=160, test n=40).

model	train_rmse	test_rmse	train_r2	test_r2	test_mae	test_lpd
Lasso_log	2.083	1.829	0.846	0.833	1.447	
ENet_log	2.080	1.842	0.847	0.831	1.449	
Ridge_log	2.078	1.856	0.847	0.828	1.451	
Lasso_raw	1.627	1.865	0.906	0.827	1.425	
ENet_raw	1.627	1.867	0.906	0.826	1.427	
OLS_log	2.078	1.868	0.847	0.826	1.454	
Ridge_raw	1.627	1.871	0.906	0.826	1.430	
OLS_raw	1.627	1.871	0.906	0.826	1.430	
BayesNIG_log	2.078	1.868	0.847	0.826	1.453	-2.052

Table 5 shows that the raw-feature models (Lasso_raw, ENet_raw, Ridge_raw, OLS_raw) reach the lowest training RMSE (around 1.627) but slightly higher test RMSE (1.865–1.871) than the best log-feature model, Lasso_log, which achieves the lowest test RMSE overall (1.829) despite a higher training RMSE (2.083). This crossover, where a model with worse training fit generalizes marginally better on this particular split, illustrates why a single fixed split is not sufficient evidence that log features are superior; Section F revisits this question with 100 repeated splits. The Bayesian model (BayesNIG_log) matches the conventional log-feature models closely on RMSE, R^2, and MAE, confirming that adding posterior uncertainty does not come at the cost of point-prediction accuracy.

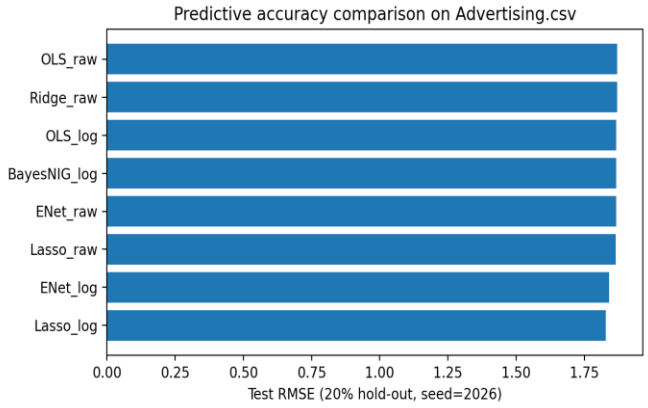

Figure 4. Test RMSE comparison on the fixed 80/20 split (random_state=2026). Lower is better.

On the fixed split, log-feature models slightly outperform raw-feature models in test RMSE (best 1.829 vs. 1.865–1.871). This comparison is descriptive rather than a statistical significance claim, and the gap is small relative to the variability observed in the repeated-split robustness analysis. Cross-validation selects very small penalties for raw features, yielding solutions close to OLS. These results suggest that, in this low-dimensional dataset, regularization primarily provides stability rather than dramatic accuracy improvements.

A key observation is that prediction and decision support need not use the same feature map. Raw features provide strong predictive performance within the observed range. The log feature map is most valuable for the prescriptive layer because it encodes diminishing returns and ensures smooth marginal ROI calculations.

3.2. Coefficient Interpretation Under OLS and ElasticNet

To connect results to classic MMM intuition, we report coefficients estimated on the full dataset using raw features. Table 6 compares OLS and ElasticNet coefficients. The magnitudes match the standard ISLR

regression: TV has a coefficient near 0.046 and radio near 0.188, while newspaper is close to zero and slightly negative.

Because TV and radio are only weakly correlated (Table 2), ElasticNet does not need to drop or group variables aggressively. Instead, it performs mild shrinkage, which slightly stabilizes coefficients while preserving their interpretation. In larger MMM settings with many correlated channels and features, ElasticNet's grouping behavior becomes more important [6], but Advertising.csv already exhibits the prototypical pattern: one weak channel (newspaper) and two strong channels (TV and radio).

Table 6. Coefficients on raw features (fit on all n=200 for interpretability).

Parameter	OLS_raw	ENet_raw
Intercept	2.938889	2.953632
TV	0.045765	0.045680
Radio	0.188530	0.187901
Newspaper	-0.001037	-0.000633

3.3. Bayesian Posterior for Log-Response Coefficients

For prescriptive modeling, we fit the conjugate Bayesian regression on log-response features and summarize its posterior. Table 7 lists posterior means, standard deviations, and 90% credible intervals. The posterior indicates that TV and radio have reliably positive effects, while newspaper's coefficient is not credibly separated from zero.

This uncertainty pattern has direct decision implications. In the optimizer, a channel with a near-zero coefficient contributes little to marginal ROI and is deprioritized unless other channels are constrained by caps or saturate. The Bayesian posterior also enables uncertainty bands on predicted sales and therefore on the sensitivity curve, which is essential for risk-aware decision making in small samples.

Table 7. Posterior summaries for Bayesian log-response regression (Normal–Inverse-Gamma prior; 5,000 draws).

Param	mean	sd	p05	p50	p95
Intercept	-13.961677	0.941149	-15.511318	-13.961563	-12.425028
log1p(TV)	4.050001	0.148045	3.808959	4.050628	4.291258
log1p(Radio)	2.983260	0.161925	2.718465	2.981079	3.250362
log1p(Newspaper)	0.113179	0.170351	-0.167362	0.115438	0.393408

3.4. Constrained Budget Sensitivity Curves with Channel Caps

Table 8 reports bounded optimal allocations and predicted sales for representative budgets. Figures 5 and 6 report the full sensitivity curve and allocation shares. Figure 5 should be read as a policy curve: for each possible total budget, it shows the expected sales under the optimized feasible allocation and compares this policy with capped fixed-share rules. Figure 6 explains how the optimizer achieves the curve in Figure 5 by showing which channel receives incremental budget and when channel caps force reallocations.

Table 8. Bounded optimal allocations and predicted sales at representative total budgets.

B_total	TV	Radio	Newspaper	TV_share	Radio_share	News_share	pred_sales_mean	pred_sales_p05	pred_sales_p95	marginal_sales_per_$1k
60.000000	34.710157	25.289519	0.000324	0.578503	0.421492	0.000005	10.270482	9.292023	11.256937	0.112520
100.000000	57.383353	41.981195	0.635452	0.573834	0.419812	0.006355	13.782874	12.922586	14.672749	0.069033
150.000000	98.609699	49.600000	1.790301	0.657398	0.330667	0.011935	16.493623	15.746982	17.250446	0.040461

B_total	TV	Radio	Newspaper	TV_share	Radio_share	News_share	pred_sales_mean	pred_sales_p05	pred_sales_p95	marginal_sales_per_$1k
200.000000	147.247233	49.600000	3.152767	0.736236	0.248000	0.015764	18.149074	17.491782	18.814528	0.027230
250.000000	195.884792	49.600000	4.515208	0.783539	0.198400	0.018061	19.330380	18.720544	19.942239	0.020519
300.000000	244.522352	49.600000	5.877648	0.815075	0.165333	0.019592	20.249527	19.679245	20.830073	0.016463
350.000000	293.159873	49.600000	7.240127	0.837600	0.141714	0.020686	21.001997	20.444816	21.563606	0.013745
400.000000	296.400000	49.600000	54.000000	0.741000	0.124000	0.135000	21.261723	20.803699	21.726681	

The solid optimized curve reports posterior-mean predicted sales at the allocation s*(B) that maximizes the posterior mean for each total budget B. The shaded band reports the 90% posterior credible interval obtained by evaluating posterior coefficient samples at s*(B), and capped fixed-share heuristics are shown as feasible comparison policies. Table 8 reports bounded optimal allocations and predicted sales for representative budgets. Figures 5 and 6 report the full sensitivity curve and allocation shares. Figure 5 should be read as a policy curve: for each possible total budget, it shows the expected sales under the optimized feasible allocation and compares this policy with capped fixed-share rules.

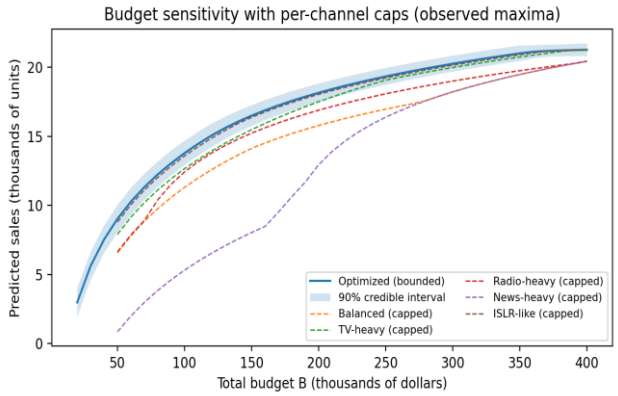

Figure 5. Budget sensitivity curve under per-channel caps

Each line in Figure 6 reports the share of the total budget allocated to TV, radio, or newspaper under the bounded optimum; flat or abrupt changes indicate regimes where a channel cap becomes binding and additional budget must be reallocated to channels with remaining feasible capacity. Figure 6 explains how the optimizer achieves the curve in Figure 5 by showing which channel receives incremental budget and when channel caps force reallocations.

The cap constraints create economically meaningful regimes. For small budgets, radio receives a substantial share because its log-response coefficient is high and its marginal returns are large at low spend. As the total budget increases, radio quickly reaches its cap of 49.6, after which additional budget flows to TV (which has a much larger feasible range). Only after TV approaches its own cap does newspaper receive substantial allocations. This behavior is consistent with MMM and advertising-allocation literature in which saturating media response curves and feasibility constraints imply that marginal dollars should move across channels as returns decline or caps become binding [2], [25], [3], [14], [15].

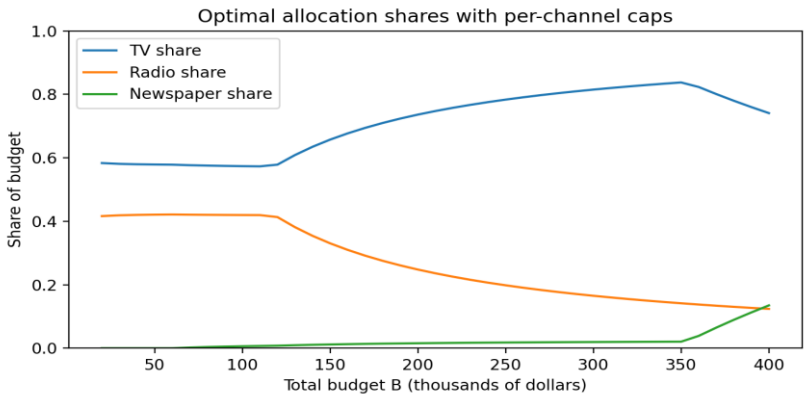

Figure 6. Optimal allocation shares as total budget increases.

The optimized curve is concave and displays strong diminishing returns, which has a direct business implication: beyond moderate budget levels, additional spending still increases predicted sales, but the incremental gain per additional $1k becomes smaller. Therefore, managers should evaluate not only whether a larger budget improves sales, but whether the marginal gain justifies the additional spend. Marginal sales per additional $1k falls from approximately 0.113 at B=60 to 0.014 at B=350 (Table 8). At B=200, the bounded optimizer allocates 73.6% to TV, 24.8% to radio, and 1.6% to newspaper, predicting 18.149 sales with a 90% credible interval of 17.492–18.815. The interval width at B=200 is 1.323, which remains small relative to the gaps between high-performing and low-performing heuristic plans.

Comparing optimized and heuristic curves in Figure 5 illustrates a key advantage of sensitivity analysis: it provides a consistent mapping from budget changes to expected outcome changes under a fixed decision policy (optimal or heuristic). In addition, the curvature and the flattening of the curve after caps become active provide a quantitative signal of saturation and diminishing returns, which is central to MMM decision making.

Regime thresholds and Lagrange multiplier interpretation. Under caps, radio becomes binding at B=120 (Table 8) and remains fixed at 49.6 for all larger budgets. At B=200, TV and newspaper are interior, and the KKT conditions imply a common marginal return $\lambda \approx 0.0273$ sales per additional $1k. This value matches both $\beta_{TV}/(1+s_{TV})$ and $\beta_{News}/(1+s_{News})$ at the optimum. Radio is capped with marginal return $\beta_{Radio}/(1+49.6) \approx 0.0589$, which exceeds λ. This inequality suggests a managerial implication: if the radio cap were relaxed, the optimizer would allocate additional budget to radio before TV or newspaper under the specified model.

Bounded vs. unbounded optimization. Without caps, the same model allocates radio well beyond its observed range (for example, an unconstrained solve at B=200 assigns approximately $83.7k to radio), because radio has a high estimated effect and the decision model has no mechanism to keep spending within the empirical domain. The bounded formulation prevents this extrapolation and yields allocations that remain supported by the data. This design choice is essential for small-data MMM-style decision making: it ensures that sensitivity curves and scenario comparisons are grounded in the dataset rather than in unvalidated extrapolation.

3.5. Scenario Comparisons and Deterministic Explanations

Bars report posterior-mean predicted sales, and error bars report the 5th and 95th percentiles of posterior predictive sales, corresponding to 90% credible intervals. The bounded optimized plan yields the highest predicted sales (18.149). The ISLR-like capped heuristic is close (18.038), and TV-heavy is slightly lower

(17.503). Plans that allocate excessive budget to newspaper perform substantially worse, consistent with the weak posterior evidence for newspaper impact. The bounded optimized plan yields the highest predicted sales (18.149). The ISLR-like capped heuristic is close (18.038), and TV-heavy is slightly lower (17.503). Plans that allocate excessive budget to newspaper perform substantially worse, consistent with the weak posterior evidence for newspaper impact.

Table 9. Scenario comparison at B=200 with per-channel caps (Bayesian 90% interval).

scenario	TV	Radio	Newspaper	pred_sales_mean	pred_sales_p05	pred_sales_p95
Optimized (bounded)	147.247233	49.600000	3.152767	18.149074	17.491782	18.814528
ISLR-like (capped)	138.830769	49.600000	11.569231	18.038003	17.599080	18.474398
TV-heavy (capped)	140.000000	40.000000	20.000000	17.502705	17.163449	17.839678
Radio-heavy (capped)	100.266667	49.600000	50.133333	16.890377	16.501198	17.293875
Balanced (capped)	75.200000	49.600000	75.200000	15.783851	15.319426	16.257103
Newspaper-heavy (capped)	36.400000	49.600000	114.000000	12.948243	12.346731	13.544792

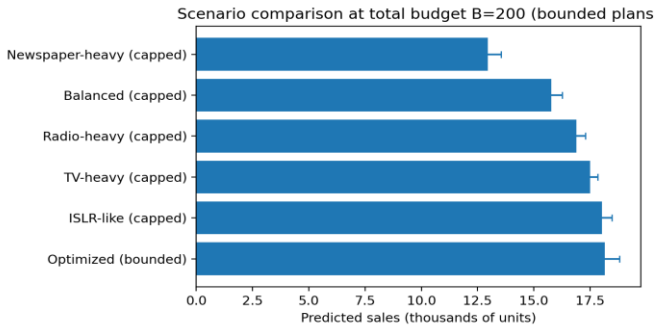

Figure 7. Scenario comparison at B=200 with 90% credible intervals

Deterministic scenario rollout (example at B=200):

Plan A: Optimized (bounded). Allocation = (TV $147.25k, Radio $49.60k, Newspaper $3.15k). Predicted sales = 18.149 with 90% CI [17.492, 18.815].

Plan B: Balanced (capped). Allocation = (TV $75.20k, Radio $49.60k, Newspaper $75.20k). Predicted sales = 15.784 with 90% CI [15.319, 16.257].

Step 1—Outcome: Plan A exceeds Plan B by 2.365 sales units at the same total budget.

Step 2—Budget shift: Compared to Plan B, Plan A shifts $72.05k from newspaper to TV while holding radio at its cap.

Step 3—Reason: Under the Bayesian posterior, the log-response coefficient for newspaper is near zero and its 90% credible interval includes zero (Table 7). Therefore, increasing newspaper spend yields low

expected marginal returns. In contrast, TV has a reliably positive coefficient, and at these spend levels the implied marginal effect $\beta_{TV}/(1+s_{TV})$ remains larger than the implied newspaper marginal effect $\beta_{News}/(1+s_{News})$. Because radio is capped, the optimizer reallocates marginal dollars to TV until TV approaches its own cap.

Step 4—Risk: The 90% intervals of the two plans do not overlap (Plan A lower bound 17.492 > Plan B upper bound 16.257), so the superiority of Plan A is robust to posterior uncertainty under the specified model.

This explanation contains only quantities computed from the posterior and the optimization constraints; it is therefore auditable and consistent with the numeric results. The explanation is validated through internal consistency checks: each reported outcome, allocation shift, marginal-return comparison, and interval statement can be traced back to Tables 7–9 and the optimization outputs. No user-study-based interpretability evaluation was conducted, so the claim is limited to numerical auditability and consistency rather than verified user preference or decision quality.

3.6. Robustness Across Random Splits

Because n=200 is small, a single train/test split is not sufficient to characterize expected predictive performance. Table 10 reports results across 100 random splits. Raw-feature OLS achieves the lowest mean RMSE (1.671) and the highest mean R2 (0.886), while log-feature OLS has higher mean RMSE (2.023). These numbers suggest that the log transform is not selected because it improves prediction on this dataset; rather, it is selected because it imposes diminishing returns and enables stable optimization.

This separation of concerns is central to the paper's approach. The predictive layer focuses on accuracy, where raw-feature OLS is strong. The prescriptive layer focuses on stability and realism of allocations under constraints, where the log-response model with caps provides a defensible decision rule within the empirical support of the dataset.

Table 10. Robustness across 100 random 80/20 splits (ShuffleSplit, random_state=2026).

Model	Features	Splits	RMSE_mean	RMSE_std	R2_mean	R2_std
OLS_raw	Raw	100	1.671	0.253	0.886	0.040
ENet_raw (fixed best)	Raw	100	1.669	0.253	0.886	0.040
OLS_log	log1p	100	2.023	0.243	0.837	0.030

These findings can be situated against prior MMM work. A constrained, data-driven budgeting framework that integrates macro demand forecasting with marketing response modeling on the same Advertising dataset used here combines Hill-function saturation curves with a constrained portfolio optimizer calibrated against real financial filings, finding that nonlinear response curves reveal strong diminishing returns with negligible incremental value for newspaper spend [26]. This closely parallels our own finding that the posterior coefficient for newspaper is near zero with a credible interval spanning zero (Table 7), and that a constrained, model-based allocation rule meaningfully outperforms a naive equal-split heuristic even on a small cross-sectional dataset (2.365 sales units, or roughly 15% higher than the balanced plan, at B=200).

Our uncertainty quantification is consistent with more recent stochastic budget-optimization studies. A stochastic mixed-integer nonlinear programming model for digital marketing budget allocation evaluates campaign outcomes across many randomly generated scenarios and reports the resulting confidence interval around the optimized objective value, finding that the interval narrows and stabilizes once enough scenarios are sampled [27]. We adopt a similar evaluate-under-uncertainty philosophy with a closed-form posterior rather than scenario sampling: at B=200, the 90% credible interval around our optimized plan is comparatively narrow (width 1.323) given the small sample

size, yet it is wide enough that the optimized and balanced plans' intervals do not overlap, indicating the ranking between a well-chosen allocation and a naive one survives posterior uncertainty even with only n=200 cross-sectional observations.

Our workflow differs from production-oriented open-source tools such as Robyn in both scope and validation strategy. Runge et al. combine ridge regression with an evolutionary search over adstock and Hill-curve saturation parameters, optimize three weighted objectives simultaneously (predictive error, a business-plausibility penalty, and an experimental-calibration error), and explicitly recommend validating MMM-derived effects against field experiments because observational allocation models can otherwise mislead decision-makers [3]. Our workflow is narrower by design: it uses a single conjugate Bayesian model rather than an evolutionary multi-objective search, and it has no adstock or carryover component because Advertising.csv is cross-sectional. We do not claim our allocations are causally validated; like Runge et al., we treat the optimizer's output as a model-conditional recommendation that would ideally be checked against experimental or quasi-experimental evidence before being implemented at scale. What our workflow adds relative to these heavier frameworks is a fully closed-form, reproducible alternative for settings where a full evolutionary or hierarchical Bayesian pipeline is not practical, together with explicit per-channel caps and credible bands that make the limits of the recommendation visible to a non-technical stakeholder.

4. Conclusions and Future Works

This paper demonstrates that "how to split the budget" can be written and evaluated in an MMM style even when the available dataset has a limited sample size. Using Advertising.csv, we conduct complete experimental evaluations of OLS and regularized regressions (Ridge, Lasso, ElasticNet) with cross-validation and hold-out testing, and we provide repeated-split robustness results. The predictive benchmarks show that raw-feature OLS performs best on average, while log-feature models are competitive on a single split but not superior in expectation.

For decision support, we fit a conjugate Bayesian log-response regression that encodes diminishing returns, supplies closed-form uncertainty quantification, and supports convex budget optimization. Crucially, we impose per-channel caps derived from observed maxima to keep recommendations within the empirical support of the dataset. Under these constraints, the optimizer yields stable allocations and a budget–sales sensitivity curve whose marginal returns decline as budget increases and as caps become active.

The resulting sensitivity curve and scenario comparisons provide direct, actionable outputs: expected sales at each budget, credible intervals, and clear statements of which channel becomes saturated first and where incremental dollars should go next. We also attach a deterministic, rule-based narrative generator that produces audit-ready explanations grounded in computed marginal returns and posterior uncertainty, enabling traceable stakeholder communication without external model calls.

Although Advertising.csv is cross-sectional and lacks many complexities of production MMM (carryover, seasonality, competitive effects), the workflow generalizes: predictive benchmarking, transformation for diminishing returns, feasibility constraints to control extrapolation, Bayesian uncertainty propagation, and optimization-based sensitivity analysis. Future work can incorporate adstock, richer saturation families, and broader marketing analytics validation protocols, as established in Bayesian media mix modeling, advertising-allocation studies, and modern open-source toolkits, while retaining the reproducible experimental discipline demonstrated here.

5. References

[1] D. M. Hanssens, L. J. Parsons, and R. L. Schultz, *Market Response Models. Econometric and Time Series Analysis*, 2nd Edition. Boston, MA, USA: Kluwer Academic Publishers, 2001.

[2] Y. Jin, Y. Wang, Y. Sun, D. Chan, and J. Koehler, "Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects," 2017.

[3] J. Runge, I. Skokan, G. Zhou, and K. Pauwels, “Packaging up media mix modeling: An introduction to Robyn’s open-source approach,” 2024.

[4] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 103. New York, NY: Springer New York, 2013. doi: 10.1007/978-1-4614-7138-7.

[5] G. James, D. Witten, T. Hastie, and R. Tibshirani, “ISLR Data Sets: Advertising.csv.”

[6] H. Zou and T. Hastie, “Regularization and Variable Selection Via the Elastic Net,” *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, Apr. 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[7] R. Tibshirani, “Regression Shrinkage and Selection Via the Lasso,” *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[8] A. E. Hoerl and R. W. Kennard, “Ridge Regression: Biased Estimation for Nonorthogonal Problems,” *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970, doi: 10.1080/00401706.1970.10488634.

[9] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009. doi: 10.1007/978-0-387-84858-7.

[10] M. Stone, “Cross-Validatory Choice and Assessment of Statistical Predictions,” *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 36, no. 2, pp. 111–133, Jan. 1974, doi: 10.1111/j.2517-6161.1974.tb00994.x.

[11] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” 1995, pp. 1137–1143.

[12] S. Arlot and A. Celisse, “A survey of cross-validation procedures for model selection,” *Stat. Surv.*, vol. 4, no. none, Jan. 2010, doi: 10.1214/09-SS054.

[13] M. W. Browne, “Cross-Validation Methods,” *J. Math. Psychol.*, vol. 44, no. 1, pp. 108–132, Mar. 2000, doi: 10.1006/jmps.1999.1279.

[14] P. Doyle and J. Saunders, “Multiproduct Advertising Budgeting,” *Marketing Science*, vol. 9, no. 2, pp. 97–113, May 1990, doi: 10.1287/mksc.9.2.97.

[15] P. J. Danaher and T. S. Dagger, “Comparing the Relative Effectiveness of Advertising Channels: A Case Study of a Multimedia Blitz Campaign,” *Journal of Marketing Research*, vol. 50, no. 4, pp. 517–534, Aug. 2013, doi: 10.1509/jmr.12.0241.

[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004. doi: 10.1017/CBO9780511804441.

[17] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. Chapman and Hall/CRC, 2013. doi: 10.1201/b16018.

[18] T. Miller, “Explanation in artificial intelligence: Insights from the social sciences,” *Artif. Intell.*, vol. 267, pp. 1–38, Feb. 2019, doi: 10.1016/j.artint.2018.07.007.

[19] R. Venkatesan, P. W. Farris, and R. T. Wilcox, *Marketing Analytics*. University of Virginia Press, 2021. doi: 10.2307/j.ctv1bd4mz9.

[20] N. Arora, R. Berman, E. Feit, D. Hanssens, A. Li, M. Lovett, C. Mela, K. Wilbur, and J. Lynch, “Charting the Future of Marketing Mix Modeling Best Practices,” 2023.

[21] P. S. H. Leeflang, P. C. Verhoef, P. Dahlström, and T. Freundt, “Challenges and solutions for marketing in a digital era,” *European Management Journal*, vol. 32, no. 1, pp. 1–12, Feb. 2014, doi: 10.1016/j.emj.2013.12.001.

[22] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[24] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nat. Methods*, vol. 17, pp. 261–272, 2020.

[25] P. A. Naik and K. Raman, "Understanding the Impact of Synergy in Multimedia Communications," *Journal of Marketing Research*, vol. 40, no. 4, pp. 375–388, Nov. 2003, doi: 10.1509/jmkr.40.4.375.19385.

[26] Y. Lu, H. Zhou, and Y. Zhang, "A Constrained, Data-Driven Budgeting Framework Integrating Macro Demand Forecasting and Marketing Response Modeling," *Journal of Technology Informatics and Engineering*, vol. 4, no. 3, pp. 493–520, Dec. 2025, doi: 10.51903/jtie.v4i3.466.

[27] Z. Yahia and M. ElBolok, "A stochastic nonlinear programming model for budget mix optimization of digital marketing campaigns under uncertainty," *Future Business Journal*, vol. 11, art. 238, Dec. 2025, doi: 10.1186/s43093-025-00664-x.