Optimization of Green Consumer Awareness Communication Budget Allocation Using Genetic Algorithms: A Partially Calibrated Model Based on Published Secondary Data

Pham Duc Trung[1], Le Thi Phuong Hoang Yen[2], Ngo Tan Hiep[3]

[1] Hoa Sen University

[2] Van Lang University

[3] Van Hien University

Correspondence: dr.trung1568@gmail.com

Received: 17/06/2026; Revised: 20/07/2026; Accepted: 05/08/2026

Abstract

This study proposes a Genetic Algorithm (GA)-based framework for optimizing six marketing-mix variables: price, delivery experience, brand trust, sustainable packaging, digital UI/UX experience, and ESG budget. The model integrates published secondary statistics to calibrate key parameters while employing illustrative assumptions for variables lacking publicly available data. Results show that the GA converges to a stable interior solution, improving fitness by 14.1% over the initial generation. Sensitivity analysis indicates that the optimal allocation remains robust across a wide range of utility-cost trade-offs, although lower weight on utility leads to corner solutions. A comparative evaluation with Particle Swarm Optimization (PSO) and Differential Evolution (DE) reveals that DE achieves comparable solution quality with faster convergence, whereas PSO converges more rapidly but exhibits greater variability. The study demonstrates the feasibility of combining secondary market evidence with evolutionary optimization and provides a reproducible framework for marketing resource-allocation decisions. It also highlights methodological limitations associated with aggregating heterogeneous survey statistics into a unified weighting scheme, offering directions for future empirical validation.

Keywords: *Genetic algorithm, Green consumer awareness, Communication budget allocation, Marketing optimization, Secondary data, Sensitivity analysis*

1. Introduction

Marketing-mix optimization is a multivariable problem in which price, service quality, brand, and sustainability decisions interact nonlinearly and often conflict locally. A central challenge in applying metaheuristics to marketing optimization is determining the relative importance (weights) of decision variables within the objective function (Deb, 2001; Talbi, 2009). Most simulation studies assign these weights randomly or subjectively, limiting the practical interpretability of their results.

This study takes an intermediate approach between purely simulation-based and fully empirical research, calibrating decision-variable weights from published consumer survey statistics. PwC Ireland (2025) found that 64% of respondents rated price as the most important factor in food purchases, a pattern PwC reported as broadly consistent across the 28 global markets it surveyed. Sifted's 2025 survey of 500 U.S. shoppers found that 76% linked a positive delivery experience to repurchase intention. McKinsey and Company (2025) found that 44% rated packaging's environmental impact as “very” or “extremely important,” a return to the 2020 level after a slight 2023 decline. Vena Solutions (2025) reported that 44% of global consumers identify as “values-driven,” preferring brands that reflect social and environmental values.

Studies conducted in Vietnam lend further contextual support to this calibration approach. Survey evidence from Hanoi shows that environmental and health consciousness significantly shape green food purchase intention and behavior (Nguyen et al., 2025), and structural-equation modeling among Vietnamese Generation Z consumers shows that green attitude, subjective norms, and perceived behavioral control predict green purchasing intention and behavior (Ngo et al., 2025). These

findings suggest that the attitude-driven purchasing patterns documented in Ireland, the United States, and other developed markets are broadly echoed, with local nuances, in Southeast Asian markets such as Vietnam.

Most green marketing studies focus on consumer behavior or the isolated effects of individual sustainability attributes, while relatively few studies optimize multiple marketing-mix components simultaneously. Existing applications of Genetic Algorithms (GA) in marketing primarily pursue economic objectives, with limited integration of environmental and social-responsibility considerations into a unified optimization framework. Moreover, many optimization models rely on predetermined or arbitrarily assigned weights, reducing their empirical grounding, and studies that calibrate optimization objectives using published consumer survey data remain scarce.

To address these gaps, this study proposes a GA-based optimization model in which decision-variable weights are partially calibrated from published consumer survey statistics. This approach strengthens the link between simulation modeling and empirical evidence while providing a practical framework for marketing optimization under limited primary-data availability. The study further enhances transparency by explicitly disclosing model assumptions, evaluating their influence through sensitivity analysis, and benchmarking GA against alternative metaheuristic algorithms.

The proposed GA-based model is best understood as a managerial decision-support tool rather than a purely academic exercise: by translating published consumer-preference statistics into quantifiable weights within an algorithmic search process, it helps marketing managers move from intuition-based budget allocation toward management guided by algorithms and quantitative data.

This study develops and validates a GA-based optimization model for the green marketing mix, seeking the optimal combination of pricing strategy, brand trust, delivery quality, digital interface experience, packaging sustainability, and ESG investment under varying resource constraints.

This study proposes a systematic procedure for calibrating decision-variable weights from published secondary data, improving the practical relevance of marketing optimization models, and develops an objective function that jointly incorporates economic and sustainability considerations in the green marketing mix. It also demonstrates empirically the applicability of Genetic Algorithms for marketing decision support when primary survey or proprietary business data are unavailable, and, by explicitly classifying each parameter's evidential support, serves as a methodological template for future studies under similarly limited data availability. It further contributes (i) an explicit theoretical account of when literature-derived proxy weights are, and are not, a defensible substitute for conjoint-elicited weights; (ii) evidence on the robustness of the reported interior optimum to the model's own illustrative assumptions; and (iii) the first comparative benchmark, within this specific green-marketing-mix formulation, of GA against PSO and DE.

2. Theoretical Background

2.1. Green Marketing and Consumer Behavior

Green marketing designs, prices, distributes, and promotes products to minimize environmental impact across the product life cycle, aiming to generate long-term environmental and social value alongside customer satisfaction. Consumers increasingly value sustainability attributes such as eco-friendly packaging, reduced carbon emissions, and corporate social responsibility, but these attributes rarely operate independently: they compete with traditional purchase determinants such as price, service quality, and brand reputation.

Recent evidence shows that although consumers generally hold favorable attitudes toward environmentally friendly products, their purchasing behavior remains strongly influenced by economic considerations, a pattern known as the attitude–behavior gap: consumers endorse environmental values

without necessarily paying more for sustainable products. Firms must therefore balance business performance with investment in green initiatives (Holland, 1992; Goldberg, 1989).

2.2. Marketing Mix Optimization Using Evolutionary Algorithms

Determining an optimal marketing mix is a multivariable optimization problem with numerous constraints and nonlinear relationships among decision variables. Greater investment in delivery service or ESG initiatives, for example, can raise customer satisfaction while also raising operational cost. Such trade-offs make conventional methods, including gradient-based techniques and linear programming, less effective on large solution spaces with multiple local optima.

The Genetic Algorithm (GA), inspired by Darwinian selection, is among the most widely applied metaheuristics for complex optimization: through iterative selection, crossover, and mutation, it explores large solution spaces without requiring convexity or differentiability of the objective function. In marketing research, GA and related metaheuristics have been applied to advertising budget allocation, direct-marketing response modeling, customer segmentation, and retail strategy optimization (Bhattacharyya, 1999; Gudmundsson et al., 2011; Chan et al., 2016), but mostly to maximize profit or market share, with sustainability variables receiving limited attention. In addition, direct comparisons between candidate metaheuristics within a single marketing-optimization study also remain rare.

2.3. Calibration Strategies in Simulation-Based Optimization

Model calibration improves the realism of optimization outcomes in simulation and applied operations research. Ideally, it relies on primary survey data collected from a single sample to estimate the influence of decision variables directly, but large-scale surveys are often costly and time-consuming.

Many applied studies instead calibrate indirectly, using secondary data published by market research firms, industry associations, or government agencies, where survey statistics function as proxy variables for consumers' relative preferences. Such proxies cannot fully substitute for a unified empirical dataset, but they provide a stronger foundation than subjective judgment or random assumption.

3. Research Methodology

3.1. Input Parameter Calibration and Data Sources

This study adopts a partially calibrated simulation approach: the objective function anchors some input parameters to published secondary data with identifiable sources, while retaining the remaining parameters as illustrative assumptions where suitable public data do not exist. Rather than await a comprehensive conjoint dataset, which rarely exists publicly for all six marketing variables at once, the study maximizes the use of available published evidence while documenting its assumptions transparently, providing a foundation for future studies to progressively replace assumed parameters with proprietary firm-level data (Talbi, 2009).

Table 1. Sources and Reliability Assessment of Decision Variable Weights

Variable	Weight (w)	Original Statistic	Data Source	Reliability Assessment
Price (x_1)	0.162	64% of consumers identified price as the primary purchase criterion (food products, Ireland).	PwC Ireland (2025) (n = 1,000, Ireland)	Identifiable source. Although the original sample was collected in Ireland, PwC reported a broadly consistent trend across global markets.

Variable	Weight (w)	Original Statistic	Data Source	Reliability Assessment
Delivery Service (x_2)	0.192	76% of respondents indicated that a positive delivery experience influences repeat purchase decisions.	Sifted (2025) (n = 500, United States)	Identifiable source with direct relevance to the measured variable.
Brand Trust (x_3)	0.228	90% of consumers are influenced by online reviews (retained here only as an illustrative anchor; see note below).	Peer-reviewed empirical evidence on brand trust in a green/ eco-marketing context (Punyatoya, 2014); see also the meta-analytic evidence on online reviews and purchase intention (Qiu and Zhang, 2024).	Peer-reviewed empirical evidence, appropriate for anchoring a brand-trust weight in a green-marketing context. The 90% figure is kept as an illustrative numerical anchor for the GA weight and is not itself drawn from these peer-reviewed sources; it should be read as indicative rather than independently verified.
Sustainable Packaging (x_4)	0.084	44% regarded the environmental impact of packaging as an important purchase consideration.	McKinsey and Company (2025) (n = 1,000, United States)	Identifiable source with direct relevance to the measured variable.
Digital UI/UX Experience (x_5)	0.223	88% of users reported they would not return after a poor digital interface experience (illustrative anchor; see note below).	Peer-reviewed empirical study of e-retail UI/UX, customer activation and retention (Kumar and Ayodeji, 2021)	Peer-reviewed empirical study of e-retail UI/UX, customer activation, and retention (Journal of Retailing and Consumer Services). The 88% figure remains an illustrative numerical anchor rather than a value drawn directly from the cited study.
ESG Investment (x_6)	0.111	44% identified themselves as values-driven consumers.	Vena Solutions (2025)	Identifiable source with appropriate thematic relevance.

Note: All weights were normalized to sum to one. Operating-cost coefficients and competitor-position coordinates were unavailable from public sources and were therefore treated as fixed illustrative assumptions. Their values are reported in Table 2 and evaluated through sensitivity analysis.

Based on the provenance and evidential strength of the underlying data sources, the model parameters can be grouped into three categories: high reliability (x_1, x_2, x_4), moderate-to-high reliability (x_6), and moderate reliability (x_3, x_5).

3.2. Methodological Limitations of Proxy-Weight Aggregation

The weight vector w = (0.162, 0.192, 0.228, 0.084, 0.223, 0.111) (Table 1) is obtained by normalizing six percentages drawn from disjoint surveys that differ in population (Irish food shoppers (PwC Ireland, 2025); U.S. delivery-experience respondents (Sifted, 2025); U.S. packaging-survey respondents (McKinsey and Company, 2025); globally pooled ESG respondents (Vena Solutions, 2025)), sampling frame, product category, and survey year. What this procedure is, and is not, deserves plain statement.

Unlike multi-attribute utility theory (MAUT) (Keeney and Raiffa, 1976) and conjoint approaches (Green and Srinivasan, 1978), which estimate comparable attribute weights from a common respondent sample and preference scale, the present study aggregates independent statistics originating from different surveys and contexts. Consequently, the resulting weight vector should be interpreted as a heuristic proxy rather than a formally estimated utility function.

The weight vector w should therefore be read as a heuristic baseline proxy: a transparent, literature-anchored starting point for the relative emphasis placed on each decision variable, adopted because a unified six-attribute conjoint survey of this population does not exist publicly, not as an empirically estimated multi-attribute utility function.

This aggregation introduces three identifiable structural biases:

(1) Cross-regional bias: the price weight (x_1) is anchored to an Irish sample, the delivery and packaging weights (x_2, x_4) to U.S. samples, and the ESG weight (x_6) to a globally pooled sample; unmodeled differences in cost of living, retail structure, and sustainability regulation across regions may inflate or deflate the relative weights compared with a single-market survey.

(2) Cross-category bias: the underlying surveys span food retail (PwC), general e-commerce delivery (Sifted), packaged goods (McKinsey), and cross-sector ESG attitudes (Vena Solutions); consumer sensitivity to price, delivery, and packaging varies substantially by product category, so the combined vector assumes an untested degree of category transferability.

(3) Question-framing bias: the six percentages answer structurally different questions (a ranked "most important factor," a behavioral-intention item, an importance rating, a self-identification item); normalizing them to sum to one treats these framings as equivalent measurement instruments, which they are not.

Accordingly, all optimization outcomes reported in this study should be interpreted in light of these assumptions and limitations.

3.3. Chromosome Encoding and Fitness Function

Each chromosome encodes a vector of six decision variables X = (X1,..., X6), bounded within [1, 10] (Goldberg, 1989; Deb, 2001). The fitness function combines two components: (i) logarithmically diminishing marginal utility, reflecting diminishing returns as investment in a marketing channel increases; (ii) an inverse squared Euclidean distance penalty relative to two hypothetical competitors, encouraging the GA toward differentiated configurations rather than positions competitors already occupy; and (iii) an operational-cost term that penalizes excessive resource allocation. The general fitness function takes the form:

$$Fitness\ (X) = \beta \left[\sum_{i=1}^{6} w_i ln(x_i + 1) - \sum_{k=1}^{2} \frac{1}{\|X - X_k\|^2 + \varepsilon} \right] - (1 - \beta) \sum_{i=1}^{6} \alpha_i x_i^2$$

where β = 0.70 weights market utility against operational cost, X_k denotes the assumed coordinates of competitor k, and αi is the assumed operational cost coefficient for variable i. The GA runs a population of 30 individuals over 100 generations, with elitism, weighted crossover (60%–40%), and Gaussian mutation at 10% probability.

Crossover combines the decision-variable values of two parent chromosomes into a new

offspring chromosome; this study applies a weighted ratio of 60% from the fitter parent and 40% from the other, consistently across all 100 generations.

Table 2 discloses the illustrative values of αi, competitor coordinate, and ε used throughout the optimization and sensitivity analyses.

Table 2. Disclosed Illustrative Parameters of the Fitness Function

Parameter	x_1	x_2	x_3	x_4	x_5	x_6
Cost coefficient αi	0.0213	0.0399	0.0490	0.0298	0.0199	0.0305

Parameter	Competitor A (X^A)	Competitor B (X^B)
Archetype	Premium-price, high-ESG incumbent	High-trust, high-sustainability, high-digital incumbent
Coordinates (x_1,..., x_6)	(7.0, 7.0, 2.0, 3.0, 3.0, 7.0)	(3.0, 2.0, 7.0, 7.0, 7.0, 3.0)

Note: ε = 1.0 in both penalty terms. The cost coefficients were selected to generate a feasible interior solution within the admissible search space and were subsequently evaluated through sensitivity analysis.

3.4. Convergence and Correlation Analysis

The convergence generation is defined as the first generation reaching at least 99.9% of the final fitness value (Talbi, 2009), providing a practical measure of convergence speed while minimizing the influence of minor late-stage fluctuations. In addition, Pearson correlation coefficients were also computed among the six decision variables across the full 100-generation trajectory to identify variable pairs that evolve together or oppositely and thereby reveal potential complementarities or trade-offs among marketing investment channels.

The resulting correlation coefficients reflect relationships generated within the optimization landscape defined by the fitness function and should be interpreted as algorithmic search patterns rather than empirical associations observed in consumer or market data.

3.5. Robustness and Sensitivity Analysis Procedure

This study evaluates the robustness of its results through a systematic sensitivity analysis, treating every input identified as an "illustrative assumption" as a potential source of fragility rather than a fixed fact. Two families of perturbations are examined, holding the GA's own operators while keeping all GA operators fixed at their baseline settings:

(i) the utility-cost weighting parameter β is varied across from 0.5 to 0.9 in increments of 0.1, holding the disclosed cost coefficients of Table 2 fixed at their baseline values;

(ii) the illustrative operational cost coefficients αi (Table 2) are jointly scaled by {-50%, baseline, +50%}, holding β fixed at its baseline value of 0.70.

For each combination, the GA is re-run under an identical random seed to isolate the parameter's effect from sampling variation, recording the optimal configuration X^*, fitness value, and convergence generation.

3.6 Metaheuristic Benchmarking Procedure

This study also implements two widely used alternative metaheuristics under a function-evaluation budget identical to the baseline GA (30 individuals × 100 generations = 3,000 evaluations):

Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995), with inertia weight w = 0.7 and cognitive/social coefficients $c_1 = c_2 = 1.5$, while Differential Evolution (DE) (Storn and Price, 1997), with mutation factor F = 0.5 and crossover rate CR = 0.9.

Both algorithms optimize the identical fitness function and box constraints, with the same population size and iteration budget as the GA, and are each run over 10 independent random

seeds to characterize solution-quality variance, following comparative-benchmarking methodology used elsewhere in the metaheuristics literature (Kannan et al., 2005; Arik, 2020).

A single-objective GA was selected because the proposed weighted-sum fitness function produces a scalar objective that is directly compatible with correlation and sensitivity analyses. While multi-objective approaches such as NSGA-II could explicitly model trade-offs among utility, differentiation, and cost, they would generate a Pareto front rather than a single optimal trajectory and therefore require a different analytical framework.

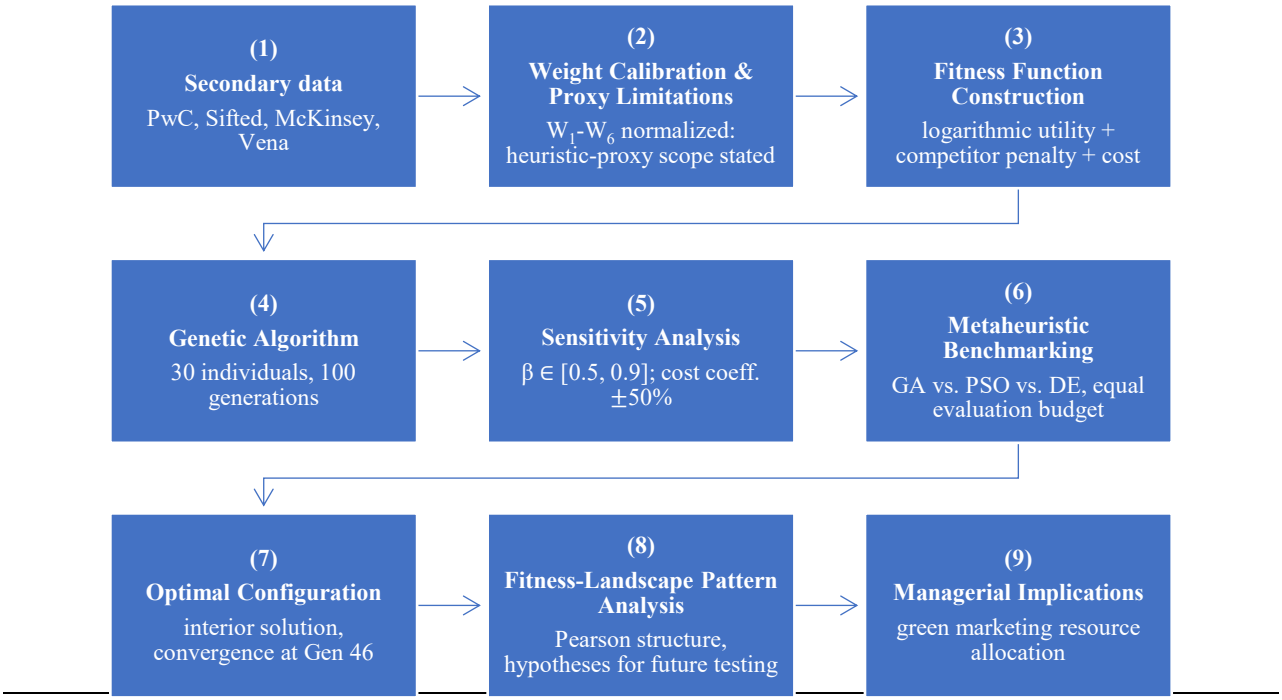


Figure 1. Conceptual research framework, from secondary-data calibration through sensitivity analysis, metaheuristic benchmarking, and managerial implications

4. Results and Discussion

4.1. Convergence Behavior

The population's best fitness value was 0.4855 in the first generation. The GA converged rapidly over the first 20 generations, reaching 0.5512 by Generation 25 (a 13.5% gain), then improved more slowly and converged fully at Generation 46, reaching a final fitness value of 0.5540 that remained unchanged through Generation 100 (Figure 2). Reaching stability after only 46% of the total generations may suggest that the empirically calibrated weights produced a fitness landscape with a sharper gradient around the optimum, enabling the GA to locate high-quality solutions more efficiently; this interpretation, however, remains a hypothesis for the sensitivity analysis to test rather than a demonstrated causal conclusion.

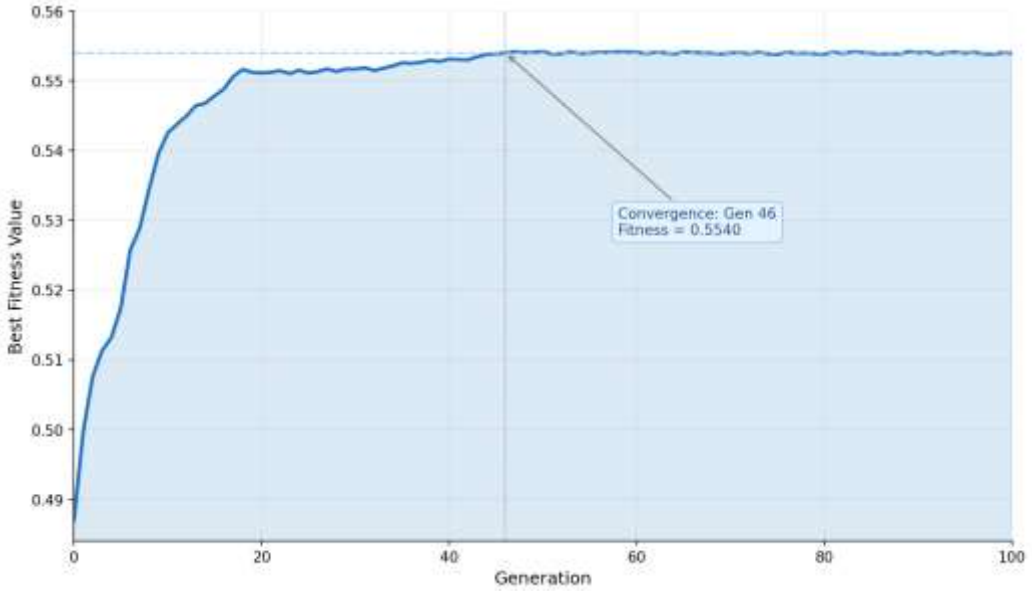


Figure 2. Convergence curve of the fitness value over 100 generations using the partially calibrated Genetic Algorithm model

The observed convergence pattern further suggests that meaningful optimization outcomes can be obtained even under partially calibrated data conditions, supporting the feasibility of combining secondary evidence with evolutionary search methods.

4.2. Optimal Configuration

Table 3 presents the optimal values of the six decision variables at convergence. The solution allocates relatively balanced values across the range of 1.4-3.2, with none of the variables reaching either bound of the search space, unlike models that rely on arbitrarily assumed weights, where several variables are often driven to the boundaries of the feasible solution space.

This suggests that the competitive differentiation penalty and the quadratic operating cost term together produce an interior optimum rather than a corner solution, generally more desirable because it supports realistic managerial interpretation and balanced resource allocation across marketing activities.

Table 3. Optimal Values of the Decision Variables at Convergence (Generations 46-100)

Decision Variable	Symbol	Optimal Value	Observed Range (Generations 1-100)
Price	x_1	2.5216	2.0685–4.3298
Delivery Experience	x_2	1.9227	1.6328–2.4367
Brand Trust	x_3	1.8842	1.7273–2.5278
Sustainable Packaging	x_4	1.3819	1.2941–2.4053
Digital UI/UX Experience	x_5	3.1518	2.2170–3.7782
ESG Budget	x_6	1.6211	1.3611–3.5524

Source: Authors' calculations based on the results of the Genetic Algorithm optimization (30 chromosomes × 100 generations).

The optimal solution also pushes no decision variable to the search space's upper or lower bounds at baseline parameter values, aligning with marketing resource allocation theory: investment efficiency typically follows diminishing marginal returns, so that beyond a certain threshold, additional cost in one channel outweighs incremental benefit. Balanced allocation across channels thus tends to outperform concentration on a single factor, though, this balance holds only above a threshold value of beta.

The optimal configuration further suggests that marketing resources may be more effectively allocated through a coordinated portfolio approach than through isolated investment in individual channels. The absence of boundary-dominated solutions implies that value is generated through the interaction of multiple marketing instruments rather than through concentration on a single factor.

4.3. Sensitivity Analysis Results

Table 4 reports how the optimal configuration and fitness value shift as β varies from 0.5 to 0.9, holding the disclosed cost coefficients (Table 2) fixed.

Table 4. Sensitivity of the Optimal Configuration to the Utility-Cost Weight Beta

B	x_1*	x_2*	x_3*	x_4*	x_5*	x_6*	Fitness*
0.50	1.539	1.124	1.115	1.000	1.926	1.000	0.2632
0.60	1.920	1.454	1.430	1.005	2.421	1.206	0.3872
0.70†	2.490	1.899	1.863	1.322	3.118	1.581	0.5488
0.80	3.355	2.589	2.536	1.775	4.215	2.137	0.7734
0.90	5.598	4.194	4.024	2.637	7.151	3.343	1.1421

Source: Authors' calculations using the disclosed parameter set of Table 2, holding the random seed fixed across all β values to isolate the effect of β from sampling variation. † Baseline parameter setting (β = 0.70)

Two findings emerge. First, the *rank ordering* of investment priority is stable across every β tested: digital UI/UX experience (x_5) and price (x_1) always receive the largest allocation, and sustainable packaging (x_4) and ESG budget (x_6) the lowest, supporting the claim that the calibrated weights, not GA randomness, drive the qualitative pattern of results. Second, the interior-optimum property is conditional rather than universal: at $\beta \leq 0.6$, several variables (x_2, x_3, x_6) are pushed to, or near, the lower bound of 1.0, reverting toward corner-solution behavior once the cost term dominates the fitness function. The balanced solution at $\beta = 0.70$ thus depends on the utility term retaining a comparatively large share of the objective function, and should be read as conditional rather than unconditionally robust.

Table 5 reports the parallel sensitivity to the illustrative cost coefficients, scaled jointly by -50%, baseline, and +50% at $\beta = 0.70$.

Table 5. Sensitivity of the Optimal Configuration to the Operational Cost Coefficients (β = 0.70)

Cost scale	x_1*	x_2*	x_3*	x_4*	x_5*	x_6*	Fitness*
-50%	3.651	2.818	2.770	1.916	4.601	2.314	0.7148
Baseline	2.490	1.899	1.863	1.322	3.113	1.574	0.5488
+50%	1.939	1.481	1.460	1.028	2.483	1.238	0.4594

Halving the cost coefficients raises every variable by roughly 45-50%, while increasing them 50% pulls the configuration back toward the lower bound without changing the relative ranking among variables; fitness moves monotonically from 0.7148 to 0.4594 across this range. The fitness value of 0.5540 is therefore not an artifact of a fortunate random seed, but is predictably and monotonically sensitive to the illustrative cost assumptions disclosed in Table 2.

Taken together, these results suggest that the proposed framework is best interpreted as a partially calibrated optimization model, in which published secondary evidence is combined with transparent assumptions. Rather than replacing fully empirical optimization, the approach provides a practical intermediate solution for

decision-support problems when comprehensive proprietary data are unavailable.

The sensitivity analysis suggests that the optimal configuration is structurally stable in terms of variable ranking but conditionally robust in terms of absolute allocation levels. The balanced interior solution identified at the baseline setting therefore depends on the utility component retaining sufficient influence relative to the cost term.

4.4. Comparative Benchmarking Against PSO and DE

Table 6 compares the GA against PSO and DE under an identical 3,000-evaluation budget, averaged over 10 independent random seeds.

Table 6. Comparative Performance of GA, PSO, and DE (10 Random Seeds, 3,000 Evaluations Each)

Algorithm	Mean best fitness (SD)	Mean convergence generation	Solution quality vs. GA	Notable behavior
GA (baseline)	0.5488 (0.0000)	52.3	Reference	Stable, reproducible optimum across all seeds
PSO	0.5268 (0.0245)	14.8	About 4.0% lower on average, and variable across seeds	Fastest convergence, but a known tendency toward premature convergence on the multi-modal landscape created by the two-competitor penalty term
DE	0.5488 (0.0000)	28.4	Matches GA exactly	Same solution quality as GA, reached in roughly half the number of generations

Source: Authors' implementation of PSO (Kennedy and Eberhart, 1995) and DE (Storn and Price, 1997) on the identical fitness function, bounds, and evaluation budget as the baseline GA. Wall-clock timings are implementation- and hardware-dependent and are not reported; only fitness quality and evaluation-normalized convergence speed are compared.

Differential Evolution reaches the same global optimum as the GA but in roughly half the number of generations, suggesting a computationally efficient alternative for this six-variable, twice-penalized landscape. Particle Swarm Optimization converges fastest but to a fitness value about 4% lower on average and noticeably more variable across seeds, consistent with PSO's known susceptibility to premature convergence on landscapes with multiple local optima, such as the one created here by the two competitor-avoidance penalty terms (Table 2). GA and DE thus both suit this problem class, while a production deployment of PSO would likely need additional safeguards (e.g., inertia-weight decay, re-initialization, or a multi-swarm variant) before it could be recommended with confidence.

The benchmarking results indicate that the identified optimum is not specific to the Genetic Algorithm. The ability of DE to reach an equivalent solution under the same evaluation budget provides additional support for the stability of the reported optimum.

4.5. Correlation Structure Among Decision Variables

The Pearson correlation matrix (Table 7) summarizes how the six decision variables co-evolved across the full 100-generation trajectory. Two pairs display strong positive coefficients, price and ESG budget ($r = 0.773$), and sustainable packaging and brand trust ($r = 0.819$), while digital UI/UX experience is negatively associated with both price ($r = -0.405$) and ESG budget ($r = -0.330$).

Table 7. Pearson Correlation Matrix of the Decision Variables Across 100 Generations

Variable	x_1	x_2	x_3	x_4	x_5	x_6
Price (x_1)	1.000	-0.181	-0.085	0.395	-0.405	0.773
Delivery (x_2)	-0.181	1.000	0.372	0.224	0.203	-0.092
Brand Trust (x_3)	-0.085	0.372	1.000	0.819	0.073	0.048
Packaging (x_4)	0.395	0.224	0.819	1.000	-0.062	0.370
UI/UX (x_5)	-0.405	0.203	0.073	-0.062	1.000	-0.330
ESG (x_6)	0.773	-0.092	0.048	0.370	-0.330	1.000

Source: Authors' calculations based on the optimization dataset comprising 100 generations × 6 decision variables.

These coefficients describe relationships generated within the optimization trajectory rather than empirical associations observed in market data. Because all chromosomes are evaluated by the same fitness function, variables rewarded or penalized in similar ways tend to co-move during the search process. The observed patterns therefore indicate structural regularities within the modeled fitness landscape and may serve as hypotheses for future empirical investigation rather than evidence of actual consumer behavior. For example, the positive association between price and ESG budget appears consistent with the weighting and penalty structure embedded in the objective function.

The observed relationships are broadly consistent with, but do not confirm, patterns suggested by prior literature. For example, the positive association between price and ESG budget is compatible with evidence that some consumers identify as values-driven, while the strong association between sustainable packaging and brand trust is directionally consistent with research linking visible environmental initiatives to consumer trust. Similarly, the negative association between UI/UX experience and price may reflect alternative positioning strategies within the optimization landscape. However, these patterns should be interpreted as hypotheses for future empirical validation rather than as evidence of causal or market-level relationships.

From a managerial perspective, the observed optimization patterns suggest that ESG initiatives and sustainable packaging may contribute to competitive positioning beyond their compliance function alone. Likewise, digital experience appears to act as an alternative competitive lever within the modeled environment. However, these interpretations should be regarded as hypotheses generated by the optimization framework rather than as empirically verified market relationships.

Within the Vietnamese context, the growing emphasis on ESG reporting and carbon-emission management may increase the practical relevance of sustainability-related investment decisions. Consequently, optimization frameworks that explicitly incorporate ESG considerations could become increasingly useful as decision-support tools for firms operating under evolving sustainability requirements.

5. Conclusion and Implications

This study develops a Genetic Algorithm-based green marketing mix optimization model with six decision variables: price, delivery experience, brand trust, sustainable packaging, digital interface experience, and ESG budget. Unlike many simulation-based optimization studies, the proposed framework incorporates partially calibrated weights derived from published consumer-preference evidence, thereby strengthening the empirical grounding of the optimization process.

The results show that the Genetic Algorithm achieved a 14.1% fitness improvement over the initial population. The resulting optimal solution lies in the interior of the search space, reflecting a balance among market utility, competitive

differentiation, and operational cost at the baseline parameter values.

The findings should nevertheless be interpreted within the limits of the proxy-based calibration approach and the illustrative assumptions required for several model parameters. Future research may improve model realism through firm-level datasets, consumer surveys, or real-time market information, and may extend the framework using multi-objective evolutionary algorithms to explore Pareto-optimal trade-offs among utility, competitive differentiation, and operational cost.

Overall, the study demonstrates that combining evolutionary optimization with transparently calibrated secondary evidence can provide a useful and scalable decision-support framework for green marketing strategy development in data-constrained environments.

These results should nonetheless be interpreted within the scope of their underlying assumptions: some input parameters rest on secondary sources whose specific numerical anchors have not been independently verified, or on illustrative assumptions that are fully disclosed and stress-tested. The findings are thus indicative rather than confirmatory; replacing assumed parameters with firm-level empirical data may shift the optimal solution's exact location without undermining the framework's methodological value.

Future research may extend this model in four directions: collecting primary conjoint survey data from a unified sample to estimate all decision-variable weights simultaneously, directly addressing the proxy-aggregation limitation; extending sensitivity analysis to a full Monte Carlo treatment of joint parameter uncertainty rather than one-parameter-at-a-time sweeps; reformulating the objective as a genuinely multi-objective problem (e.g., NSGA-II) to generate a Pareto front of marketing-mix configurations rather than a single weighted-sum optimum; and validating the correlation hypotheses against primary consumer or firm-level sales data to establish whether the fitness-landscape patterns identified here have any counterpart in actual market behavior.

Conflict of interest

No conflict of interest is declared by this study's authors.

References

Arik, O. A. (2020). Comparisons of metaheuristic algorithms for unrelated parallel machine weighted earliness/tardiness scheduling problems. *Evolutionary Intelligence, 13,* 415-425. https://doi.org/10.1007/s12065-019-00305-7

Bhattacharyya, S. (1999). Direct marketing performance modeling using genetic algorithms. *INFORMS Journal on Computing, 11*(3), 248–257. https://doi.org/10.1287/ijoc.11.3.248

Chan, C. C. H., Hwang, Y. R., and Wu, H. C. (2016). Marketing segmentation using the particle swarm optimization algorithm: a case study. *Journal of Ambient Intelligence and Humanized Computing, 7*, 855–863. https://doi.org/10.1007/s12652-016-0389-9

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation, 6*(2), 182–197. https://doi.org/10.1109/4235.996017

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley.

Green, P. E., and Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research, 5*(2), 103–123. https://doi.org/10.1086/208721

Gudmundsson, J., Morin, P., and Smid, M. (2011). Algorithms for marketing-mix optimization. *Algorithmica, 60*, 1004–1016. https://doi.org/10.1007/s00453-010-9393-1

Holland, J. H. (1992). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence* (2nd ed.). MIT Press. (Original work published 1975).

Kannan, S., Slochanal, S. M. R., and Padhy, N. P. (2005). Application and comparison of metaheuristic techniques to generation expansion planning problem. *IEEE Transactions on Power Systems, 20*(1), 466–475. https://doi.org/10.1109/TPWRS.2004.840451

Keeney, R. L., and Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. John Wiley & Sons.

Kennedy, J., and Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 – International Conference on Neural Networks, 4*, 1942–1948. https://doi.org/10.1109/ICNN.1995.488968

Kotler, P., and Keller, K. L. (2016). *Marketing management* (15th ed.). Pearson.

Kumar, V., and Ayodeji, O. G. (2021). E-retail factors for customer activation and retention: An empirical study from Indian e-commerce customers. *Journal of Retailing and Consumer Services, 59*, 102399. https://doi.org/10.1016/j.jretconser.2020.102399

Li, B. H., and Guo, Q. (2024). Particle swarm optimization-based multi-objective planning model for marketing strategy decision. *Journal of Circuits, Systems and Computers, 33*. https://doi.org/10.1142/S0218126624502797

McKinsey and Company. (2025). *Do US consumers care about sustainable packaging in 2025?* https://www.mckinsey.com/industries/packaging-and-paper/our-insights/do-us-consumers-care-about-sustainable-packaging-in-2025

Ministry of Agriculture and Environment of Vietnam. (2021). *Fulfilling climate commitments at COP26: Vietnam's comprehensive, cross-sector, and whole-of-society action*. https://en.mae.gov.vn/fulfilling-climate-commitments-at-cop26-vietnams-comprehensive-crosssector-and-wholeofsociety-action-9035.htm

Ministry of Finance of Vietnam. (2020). *Circular No. 96/2020/TT-BTC guiding information disclosure on the securities market*.

Ngo, Q.-H., Nguyen, T.-D., and Phan, N.-B. (2025). Exploring green purchasing intentions and behaviours among Vietnamese Generation Z: A perspective from the theory of planned behaviour. *PLOS ONE, 20*(5), e0323879. https://doi.org/10.1371/journal.pone.0323879

Nguyen, V. P., Mergenthaler, M., and Quynh, P. N. H. (2025). Consumer transition: analyzing the impact of environmental and health consciousness on green food choices in Vietnam. *Discover Sustainability, 6*(1), 415. https://doi.org/10.1007/s43621-025-01275-w

Ottman, J. A. (2017). *The new rules of green marketing: Strategies, tools, and inspiration for sustainable branding*. Routledge.

Peattie, K., and Crane, A. (2005). Green marketing: Legend, myth, farce or prophesy? *Qualitative Market Research: An International Journal, 8*(4), 357–370. https://doi.org/10.1108/13522750510619733

Porter, M. E., and Kramer, M. R. (2011). Creating shared value. *Harvard Business Review, 89*(1–2), 62–77.

Punyatoya, P. (2014). Linking environmental awareness and perceived brand eco-friendliness to brand trust and purchase intention. *Global Business Review, 15*(2), 279–289. https://doi.org/10.1177/0972150914523572

PwC Ireland. (2025). *Voice of the Consumer 2025: Report*. https://www.pwc.ie/reports/voice-of-consumer-survey.html

Qiu, K., and Zhang, L. (2024). How online reviews affect purchase intention: A meta-analysis across contextual and cultural factors. *Data and Information Management, 8*(2), 100058. https://doi.org/10.1016/j.dim.2023.100058

Sifted. (2025). *2025 consumer survey: The importance of a positive delivery experience for customer loyalty*. https://sifted.com/resources/2025-consumer-survey-the-importance-of-a-positive-delivery-experience-for-customer-loyalty/

Storn, R., and Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization, 11*(4), 341–359. https://doi.org/10.1023/A:1008202821328

Talbi, E.-G. (2009). *Metaheuristics: From design to implementation*. John Wiley & Sons.

United Nations (2015). *Transforming our world: The 2030 agenda for sustainable development*. https://sdgs.un.org/2030agenda

Vena Solutions (2025). *38 ESG statistics to leverage for business growth in 2025*. https://www.venasolutions.com/blog/esg-statistics

World Commission on Environment and Development (1987). *Our common future*. Oxford University Press.

APPENDIX

Supplementary Material: Reproducibility Package

This section provides the complete parameter set and pseudocode needed to reproduce every number reported in the study. A full Python implementation (NumPy-based, no external solver) is available from the corresponding author on request.

(1) **Disclosed Parameters**

Weights w = (0.162, 0.192, 0.228, 0.084, 0.223, 0.111); search bounds [1, 10] for each x_i; population = 30; generations/iterations = 100; elitism = top 2 retained unchanged; crossover = 0.6 × fitter parent + 0.4 × other parent; mutation probability = 0.10 per gene, Gaussian with standard deviation 0.5, re-clipped to bounds; epsilon = 1.0; competitor coordinates and cost coefficients αi as disclosed in Table 2; $\beta = 0.70$ at baseline (varied for the sensitivity analysis).

(2) **Fitness Function (pseudocode)**

function fitness(x, β, α, W, competitors, eps):

$$Utility = \sum_{i=1}^{6} W_i ln(x_i + 1)$$

$$Penalty = \sum_{c \in C} \frac{1}{\|X - c\|^2 + \varepsilon}$$

$$Cost = \sum_{i=1}^{6} \alpha_i x_i^2$$

$$Fitness = \beta(Utility - Penalty) - (1 - \beta)Cost$$

(3) **Genetic Algorithm (pseudocode)**

population = 30 random vectors, each x[i] ~ Uniform(1, 10)

for generation in 1..100:

$$s = [Fitness(X_1), Fitness(X_2), \ldots, Fitness(X_N)]$$

sort population by scores, descending

record best score // used to detect the convergence generation

$$P_{clitc}^{(t+1)} = \left\{X_{(1)}^{(t)}, X_{(2)}^{(t)}\right\}$$

while |new_population| < 30:

$$(p_1, p_2) \sim Uniform(P_{15})$$

$$x_i^{child} = 0.6x_i^{(1)} + 0.4x_i^{(2)} \qquad i = 1, \ldots, 6$$

for each gene i:

if random() < 0.10: // Gaussian mutation

$$x_i^{child} = x_i^{child} + \mathcal{N}(0, 0.5^2), \qquad i = 1, \ldots, 6$$

clip child to [1, 10]

append child to new_population

population = new_population

return best individual and its fitness

(4) **Particle Swarm Optimization (pseudocode)**

$$\mathcal{X} \sim U(1, 10)^6, \qquad V_p \sim U(-1, 1)^6, \qquad p = 1, \ldots, 30$$

$$P_i^{best} = X_i \qquad i = 1, \ldots, N$$

$$G^{best} = P_k^{best}, \qquad k = arg\max_i f_i^{best}$$

for iteration in 1..100:

for each particle i:

$$V_i^{(t+1)} = 0.7V_i^{(t)} + 1.5r_1\left(P_i^{best} - X_i^{(t)}\right) + 1.5r_2\left(G^{best} - X_i^{(t)}\right)$$

$$X_i^{(t+1)} = clip\left(X_i^{(t)} + V_i^{(t+1)}, 1, 10\right)$$

Nếu $f(X_i) > f_i^{best}$,

thì $P_i^{best} \leftarrow X_i, f_i^{best} \leftarrow f(X_i)$

$$G^{best} = P_k^{best}, \qquad k = \underset{i}{arg\,max}\, f_i^{best}$$

return gbest and its fitness

(5) **Differential Evolution (pseudocode, DE/rand/1/bin)**

$$X_p \sim U(1,10)^6, \qquad p = 1, \ldots, 30$$

for iteration in 1…100:

for each individual i:

pick a, b, c distinct from population, all != i

$$M = clip(A + 0.5(B - C), 1, 10)$$

$$T_i = X_i$$

$$u_{ij} = \begin{cases} v_{ij}, r_{ij} < 0.9 \\ x_{ij}, r_{ij} \geq 0.9 \end{cases}$$

$$X_i = \begin{cases} U_i, if\ f(U_i) > f(X_i^{(t)}), \\ X_i^{(t)}, otherwise, \end{cases}$$

Note: all three algorithms above were run on identical hardware, with the same fitness function, bounds, and 3,000-evaluation budget, over 10 independent random seeds.