A Decision Framework for Identifying High-Potential Emerging Consumer-Packaged Goods Brands

Kristin Henderson and Bivin Sadler
Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
{hendersonk, bsadler}@smu.edu

Abstract. Decision makers in consumer-packaged goods supply chains must decide which emerging brands to support or invest in. These decisions are often made under conditions of limited data and uncertain growth trajectories. Many prior studies focus on demand forecasting, despite evidence that brand performance is also influenced by factors such as distribution coverage, promotional efficiency, and category context. This study develops a multi-criteria decision analysis (MCDA) framework to evaluate and rank emerging brands. The framework uses demand-based features combined into a single opportunity score, with room for expansion to include additional retail performance indicators. To benchmark the MCDA rankings, an ordinal logistic regression model is used to assign brands to ordered opportunity tiers. Model performance is evaluated using a hand-labeled truth set and tested across forecast horizons of 6 to 18 months. Results indicate that median year-over-year case sales growth distinguishes low-performing from viable brands with up to 85% balanced accuracy, and that performance is stable up to 18 months in advance. The ordinal model provides finer-grained tier assignments with 93% adjacent accuracy. These findings suggest that a small set of demand-based features can support practical brand evaluation for investment and acquisition decisions.

1 Introduction

In supply chain management, investors, distributors, brokers, and retailers must make forward-looking decisions about which emerging brands to support or invest in. Assessing future demand informs these decisions and improves planning, resource allocation, and overall supply chain performance (Lee, 2002; Rojas, Rojas, & Wanke, 2024).

Poor assessments of demand and growth potential can lead to inefficient resource use and missed growth opportunities. Organizations rely on demand forecasts to inform planning and resource decisions across multiple departments, making the quality of these predictions important for overall business performance (Punia &

Shankar, 2022). Developing structured ways to assess demand alongside other indicators of growth potential can help organizations make clearer, more informed decisions about where to focus investment and growth efforts.

Prior research has focused on improving demand forecasting accuracy, using traditional time series models, and more recently, machine learning and deep learning approaches. Three prior SMU Master of Science in Data Science (MSDS) capstone projects applied these methods to forecast demand for an alcoholic beverage distributor, showing that demand patterns differ across products and that no single forecasting model performs well in all cases (Arora et al., 2020; Ford et al., 2020; Jiang et al., 2020). This work highlights both the value of demand forecasting and the challenges of model selection and scalability.

Separately, a body of operations and management science literature describes multi-criteria decision analysis (MCDA), also referred to as multi-criteria decision making (MCDM), as an approach for evaluating alternatives when decisions involve multiple, often competing criteria. MCDA methods are commonly used in supply chain management for problems such as supplier selection, resource allocation, and risk assessment, where decision makers need to evaluate the impact of both quantitative and qualitative factors (Bozorg-Haddad et al., 2021; Kizielewicz et al., 2021; Olson, 2025).

There is limited work that applies MCDA to brand-level investment decisions in consumer-packaged goods supply chains. Although MCDA is widely used to assess alternatives across multiple criteria in operations and management, its use in evaluating emerging brands using demand and growth-related signals is limited. As a result, decision makers lack a structured and robust method for assessing whether emerging brands meet minimum viable product (MVP) criteria.

This research develops an MCDA framework to identify emerging consumer-packaged goods brands with a high probability of near-term value creation. Rather than relying solely on demand forecasting, the framework integrates demand signals, distribution dynamics, promotional efficiency, and category performance into a weighted opportunity score used to rank brands for investment consideration. To benchmark these rankings, the study also evaluates an ordinal logistic regression model, a statistical approach appropriate for modeling ordered outcomes such as opportunity tiers (Parry, 2024).

Model performance is evaluated by comparing predicted opportunity tiers with realized brand outcomes over a 6–18-month period, including acquisitions, breakout growth, and business failure. The resulting rankings are intended to help a brokerage/distribution agency identify high-potential brands that meet MVP investment criteria for acquisition or incubation, while also highlighting the primary drivers of each opportunity.

2 Literature Review

This section reviews prior work on retail performance signals, multi-criteria decision analysis, and modeling approaches for ordered decision outcomes.

2.1 Brand-Level Performance Signals in Retail Markets

Prior research in marketing and retail analytics shows that early brand performance can be assessed using observable demand patterns and distribution structure, even when market coverage is uneven and some drivers of performance are not directly observed.

Bronnenberg and Sismeiro (2002) show that brand-level sales performance can be assessed even when direct sales data are incomplete by using information from related retail markets. Using data from multiple consumer-packaged goods markets, the study finds that differences in retailer presence explain substantial variation in brand sales across markets. This indicates that brands tend to perform similarly in markets that share the same retail chains rather than simply in geographically nearby locations. Although the analysis focuses on a well-established brand, the authors discuss applications in which initial sales performance in a subset of markets informs decisions about market expansion. These findings demonstrate that distribution patterns can provide useful information for assessing sales and growth potential when market coverage is uneven and some factors that drive performance are not available in the data.

Prior work by Michis (2023) shows that distribution coverage plays an important role in brand-level sales, but its value depends on the stores where products are carried rather than simply on how widely they are distributed. Using weekly SKU-level retail data in consumer-packaged goods, the study demonstrates that weighted distribution provides a more informative measure of coverage than numeric distribution, particularly when brands expand into lower-volume stores. The analysis uses multiple SKUs within each brand to capture variability between newer and more established products, so that observed sales patterns reflect differences across a brand's portfolio. Results indicate that expanding distribution into less important outlets or without sufficient in-store support is associated with diminishing sales. The study also finds that deep price discounts tend to be less effective once prices are already low. These findings support treating distribution gaps and promotional efficiency as features for assessing sales and growth potential.

Using data across multiple consumer-packaged goods categories and countries, Victory (2017) finds that new products typically enter the market with higher prices and more limited distribution than established products. Early growth in distribution is closely related to whether a product remains in the market beyond

its first year. Performance also differs by manufacturer size, with products from smaller manufacturers launching with more limited distribution than those from larger manufacturers. These findings indicate that evaluating early brand performance requires attention to both demand and distribution.

2.2 Multi-Criteria Decision Analysis in Supply Chain Decision-Making

Khan et al. (2018) review the use of MCDA in supply chain management and emphasize that many supply chain decisions involve balancing competing objectives rather than optimizing a single metric. The review discusses trade-offs such as maximizing profit, minimizing risk, and ensuring product availability across supply chain functions, including supplier selection, manufacturing, warehousing, and logistics.

The authors show that MCDA methods are primarily used for strategic and tactical decisions, where choices have longer-term consequences. MCDA techniques allow decision makers to combine performance across both quantitative and qualitative criteria, reflect how different criteria are prioritized, and compare alternatives when some inputs are uncertain or difficult to measure directly. MCDA provides a framework for evaluating and ranking alternatives rather than generating forecasts when decisions depend on multiple, potentially conflicting criteria.

2.3 Predictive and Ranking Models for Ordered Business Outcomes

In many applied decision situations, outcomes are often ordered rather than continuous or binary. Common examples include how often behavior occurs (e.g., never, sometimes, often), performance tiers (e.g., low, medium, high), or ordered ratings such as Likert-scale responses (e.g., strongly disagree to strongly agree). In these cases, categories have a clear order, but the differences between categories are not necessarily equal or measurable.

Ordinal logistic regression, often referred to as an ordered logit model, is designed for this type of response variable (Agresti, 2010; Ananth & Kleinbaum, 1997). Unlike linear regression, it does not assume equal spacing between categories, and unlike nominal models, it uses the ordering information in the outcome. Results are interpreted in terms of the odds of being in a higher versus lower category, which makes the model appropriate for evaluation and prioritization rather than predicting exact numerical values.

An applied example is provided by Giannikos and Korkou (2025), who model credit card repayment behavior using the ordered categories “hardly ever,” “sometimes,” and “always or almost always” paying off monthly balances. Rather

than predicting exact repayment amounts, the analysis estimates the odds of moving to a higher repayment category. This demonstrates how ordinal regression can be used to classify and rank outcomes based on relative ordering when the outcome represents ordered levels of performance.

Drugova and Curtis (2022) use ordered logit models to study how buyers at different stages of the organic wheat supply chain, including millers, distributors, and bakers, evaluate product quality using ordered categories ranging from very low to very high. The analysis focuses on how specific product attributes affect the odds of receiving a higher quality rating rather than estimating a precise quality score.

Although ordered logit models are well established for evaluating ordered outcomes, there is limited evidence of their use in ranking brands by investment or growth opportunity. This study applies ordinal logistic regression to brand-level evaluation using ordered opportunity tiers to support decision-making.

This literature supports evaluating emerging brands using multiple performance signals in a decision-based framework. We hypothesize that models incorporating demand signals, distribution dynamics, and promotional efficiency will better distinguish higher- and lower-opportunity brands than uninformed benchmarks.

3 Data

3.1 Data Source and Structure

The dataset was provided by a consumer-packaged goods brokerage/distribution agency. The data consist of weekly records with sales, distribution, and promotion measures aggregated across U.S. retail outlets. Each record corresponds to a single Universal Product Code (UPC) in a given week. The full dataset contains approximately 77 million observations collected between 2021 and 2026, spanning up to 260 weekly periods per product.

For this analysis, the data were restricted to one primary product megacategory and a single market with the widest coverage. After filtering, the resulting dataset contains approximately 7.3 million observations representing 46,008 unique UPCs and 22,495 unique brand extensions across 8,739 brand families.

Products in the dataset are organized into a multi-level hierarchy. At the lowest level, each UPC identifies a single product. UPCs are grouped into brand extensions, which represent a product variant within a brand. Brand extensions are grouped into brand families, which represent the broader product line. To illustrate with a hypothetical example: a single UPC might identify a 12-oz bag of salt and vinegar

potato chips, the brand extension would be the salt and vinegar flavor line across all sizes, and the brand family would be the potato chip brand. The dataset also includes product category and subcategory classifications and a price tier (e.g., value, premium, luxury) for each product.

The variables used in this analysis include dollar sales, case sales, and percent all-commodity volume (%ACV). %ACV is a measure of distribution coverage that reflects how much of the market a product has access to, weighted by store importance: a product with a %ACV of 5% is carried in stores that represent 5% of total market sales volume. The dataset also includes base and promotional pricing variables, which are not used in the current analysis but may be used for future versions of the decision framework.

Each record is associated with a week-ending date, which was converted to a consecutive week index from 1 to 260. Most analysis in this study is conducted at the brand extension level rather than the individual UPC level.

Weekly UPC-level data are aggregated to the extension level by summing dollar sales and cases sold across all UPCs within each extension, since total volume is a better measure of an extension's performance than a per-UPC average. %ACV is aggregated by taking the maximum across UPCs, since a brand extension has distribution in any store where at least one of its UPCs is carried.

All data are proprietary and were analyzed in accordance with the terms of a non-disclosure agreement. Results are presented in anonymized form.

3.2 Exploratory Analysis

Total market sales show a stable trend with seasonal patterns across all five years (see Fig. 1). Sales rise in summer months, peak again around Thanksgiving, and reach their highest point at the end of the year. Distribution coverage follows a similar pattern, with an additional dip early in each year. Year-over-year growth, used later, helps adjust for these seasonal effects.

Product categories show different trends over the five-year period. Some are stable, while others are expanding or declining. These differences suggest that a brand's performance may be better interpreted in the context of its category.

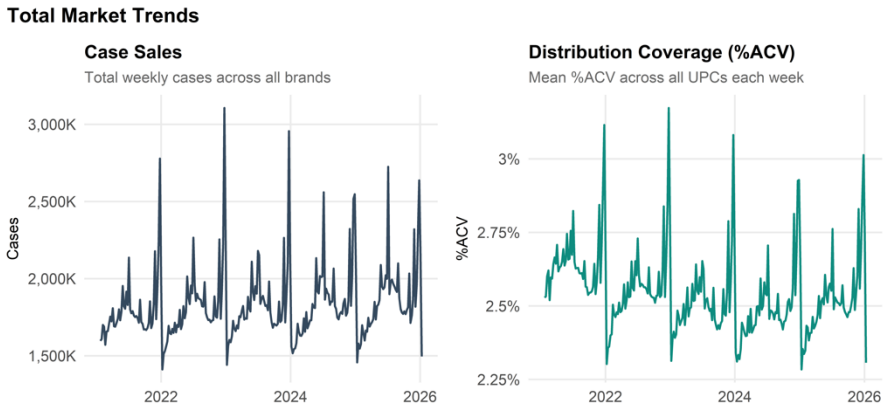

Fig. 1. Total weekly case sales (*left*) and mean distribution coverage (*right*) over 260 weeks. Both series show stable trends with large seasonal peaks at year-end and a dip early in each year.

Sixty percent of brand families contain a single extension, and the median number of UPCs per extension is one. However, within families that contain multiple extensions, individual extensions can show different performance trajectories. Some extensions may be growing while others in the same family are declining. This supports analyzing performance at the extension level, where individual patterns are preserved, rather than only at the family level, where aggregation may obscure and dilute them.

Visual inspection of brand trajectories reveals different performance patterns, from sustained growth to collapse with no recovery (see Fig. 2). These patterns motivated the labeling approach described in Section 4.1 and the demand features described in Section 4.3.

4 Methods

4.1 Truth Set Construction

Two truth sets were developed for this study. The first is a hand-labeled set of 112 brands within a single product subcategory, used for model training and evaluation. The second is a set of 18 brand families provided by a stakeholder, used for external validation.

4.1.1 Hand-Labeled Subcategory Truth Set. A single subcategory with the largest number of single-extension brand families was selected for labeling. It also had a wide range of case sales and distribution coverage across brands and a mix of performance trajectories, from little to no activity to steady growth. Having a large pool made it possible to label only brands with clear trajectory patterns. The subcategory's overall sales trend is stable, which avoids confounding decline at the brand level with decline in the subcategory.

The sample was restricted to brand families with only one extension. When a family has only one extension, the two levels are identical. This avoids complications where some extensions in a family are growing while others are declining, which would make the family harder to classify. Brands were also required to have at least 52 weeks of data, ensuring enough history for a confident assessment of the trajectory.

For each eligible brand, trajectory plots of weekly case sales and %ACV were generated with 13-week moving averages to smooth weekly noise. Brands were classified into one of four ordered tiers based on their trajectory over approximately the last two years. Brands labeled as discontinued had sales and distribution that collapsed to near zero with no recovery or disappeared from the dataset entirely. Brands labeled as decline showed a sustained downward trend in sales or distribution but had not declined to zero. Stable brands had relatively flat sales and distribution with no strong trend in either direction, and growing brands showed a sustained upward trend in sales or distribution. Fig. 2 shows one representative brand from each tier.

The final truth set contains 112 brands: 29 discontinued, 30 decline, 28 stable, and 25 growing. The distribution is reasonably balanced across tiers.

Three outcome variables were derived from these tiers. The first groups discontinued and decline brands together as low-performance, and stable and growing brands together as viable. The second separates discontinued brands from all others. The third is the full four-tier ordinal variable (discontinued < decline < stable < growing), used for ordinal logistic regression.

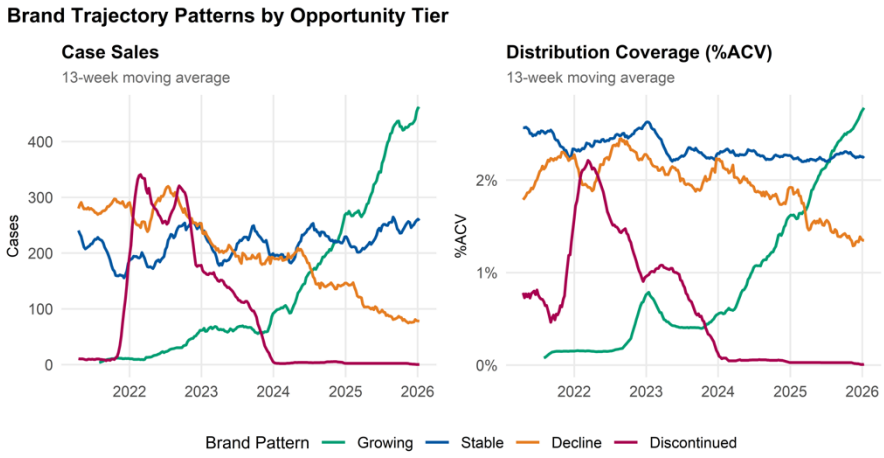

Fig. 2. Representative brand trajectories from the hand-labeled truth set, showing 13-week moving averages of case sales (*left*) and distribution coverage (*right*). Each line represents one brand from a different opportunity tier: growing, stable, decline, and discontinued.

4.1.2 Stakeholder Truth Set. A domain expert at the brokerage/distribution agency identified 18 brand families as either bankrupt (10 families, 37 extensions) or acquired (8 families, 123 extensions) (see Table 1). This set is used for external validation only.

Table 1. First 10 rows of the stakeholder truth set (anonymized). Each row represents one brand family identified by a domain expert as either bankrupt or acquired. The full set contains 8 acquired families (123 extensions) and 10 bankrupt families (37 extensions).

Brand Owner	Brand Family	Outcome	Year	Month
Owner A	Family 1	Acquired	2024	May
Owner B	Family 2	Acquired	2025	Sep
Owner C	Family 3	Acquired	2025	Jan
Owner C	Family 4	Acquired	2025	Jan
Owner C	Family 5	Acquired	2025	Jan
Owner D	Family 6	Acquired	2024	Apr
Owner E	Family 7	Acquired	2023	Sep
Owner F	Family 8	Acquired	2023	Nov
Owner G	Family 9	Bankrupt	2025	Apr
Owner H	Family 10	Bankrupt	2025	Aug
…	…	…	…	…

Bankruptcy and acquisition records serve as proxies for brand outcomes but are imperfect. A brand may file for bankruptcy for restructuring or other financial reasons while its sales and distribution remain stable. Similarly, a brand may be acquired despite declining performance. The reasons behind each bankruptcy or acquisition are not known. These labels were provided at the family level, but individual extensions within the same family may show different trajectories. These limitations are discussed further in Section 4.4.

4.2 Models

4.2.1 MCDA Framework and Opportunity Scoring. The goal of this analysis is to evaluate and rank brand extensions by opportunity potential using multiple performance signals. MCDA provides a structured way to combine different types of information into a composite score that can be used for ranking and comparison. As reviewed in Section 2.2, MCDA is commonly used for supply chain decisions involving multiple criteria (Khan et al., 2018), but its use for brand-level investment decisions in consumer-packaged goods is limited.

The framework includes four criteria. Demand momentum measures recent sales growth trajectory. Distribution coverage captures how widely a brand is carried and whether it has room to grow. Promotional efficiency reflects how effectively a brand generates sales from promotional activity. Category context considers whether a brand's product category is growing, shrinking, or changing in composition. Each criterion measures a different aspect of a brand's opportunity potential. This study implements the demand momentum criterion. The framework can be extended as the other criteria are added iteratively in future work.

Each criterion is scaled from 0 to 1 using min-max normalization (Vafaei et al., 2022). Scaling is applied so that higher values are always better. Volatility was negated before normalization so that lower standard deviation receives a higher score, since more consistent growth provides a clearer signal of a brand's trajectory. The normalized features were combined into an opportunity score:

$$\text{Score} = w_1 \times \text{Level} + w_2 \times \text{Acceleration} + w_3 \times \text{Volatility}, \quad (1)$$

where weights sum to 1. Weights can be adjusted through sensitivity analysis or stakeholder input.

Multiple weight configurations were tested across four feature combinations: level only, level plus acceleration, level plus volatility, and all three. Level's weight ranged from 60% to 100%, with the remaining weight allocated to one or both secondary features. Both year-over-year (YoY) and quarter-over-quarter (QoQ) versions of acceleration and volatility were tested. For each weight configuration,

the mean opportunity score was computed for each of the four truth-set tiers. Separation was evaluated using the gap between the means of adjacent tiers. Larger gaps indicate better separation between performance levels.

For comparison with the logistic regression models, the opportunity score was converted to a binary classification. Brands scoring above the median were classified as viable; brands scoring below the median were classified as low-performance. The median was used as a simple threshold, though alternative thresholds could be explored depending on decision priorities.

4.2.2 Binary Logistic Regression. Two binary logistic regression models were fit to benchmark the MCDA opportunity scores, each using a different outcome definition. Each was built incrementally, starting with level alone and adding acceleration and volatility one at a time. The same four feature combinations used in the MCDA scoring were tested: level only, level plus acceleration, level plus volatility, and all three. Likelihood ratio tests and the Akaike information criterion (AIC) were used to test whether each added feature improved model fit. The likelihood ratio test checks whether adding a feature produces a statistically significant improvement in model fit. AIC penalizes model complexity, so lower values indicate a better balance of fit and parsimony.

The first binary model classifies brands as low-performance or viable. The low-performance group includes brands labeled as discontinued or decline in the truth set. The viable group includes brands labeled as stable or growing.

The second binary model classifies brands as discontinued or not discontinued, combining decline, stable, and growing brands into the not-discontinued group. This produces an imbalanced outcome (26% discontinued, 74% not discontinued). Inverse frequency class weights were applied during model fitting so that the total weight of each class is equal. An alternative would be to adjust the classification threshold to change the decision boundary.

4.2.3 Ordinal Logistic Regression. A third benchmarking model uses ordinal logistic regression, also called ordered logit (Agresti, 2010), to classify brands into all four tiers: discontinued, decline, stable, and growing. The same feature combinations and model selection approach described in Section 4.2.2 were used. Rather than collapsing tiers into two groups like the binary models, this model uses the ordering of the tiers. The model estimates three intercepts, one for each boundary between adjacent tiers (discontinued to decline, decline to stable, and stable to growing), and one coefficient per feature. The intercepts set the baseline odds at each boundary. The feature coefficients shift those odds by the same amount at every boundary. This is called the proportional odds assumption. For example, the odds that a brand is classified above the discontinued-decline boundary increase by the same factor as

the odds that a brand is classified above the stable-growing boundary, for a given increase in level.

4.3 Features

Demand momentum summarizes each brand's recent sales growth using three features computed from a year-over-year growth history.

4.3.1 Year-over-Year Growth Time Series. Weekly YoY growth is computed for each brand extension as

$$\left(X_t - X_{(t-52)}\right)/X_{(t-52)}\,, \tag{2}$$

where X_t is weekly case sales and $X_{(t-52)}$ is the same week one year prior. Weeks with zero sales the previous year are treated as missing to avoid division by zero. Brands with fewer than 52 weeks of valid YoY data use QoQ growth, computed as proportional change with a 13-week lag. QoQ features are used only when YoY growth cannot be computed due to insufficient history.

Both case sales and dollar sales growth were computed. Cases were selected as the primary demand measure because dollar sales reflect both volume and pricing effects. Divergence between case and dollar growth could be explored in future analysis to identify promotional or pricing patterns.

4.3.2 Summary Features. Three summary features are computed from each brand's 52 most recent weeks of growth. Level is the median YoY growth rate over the 52-week window. It measures how much a brand is growing or shrinking. The median is used rather than the mean to reduce the influence of extreme values. Acceleration is the slope of the YoY growth series, computed using ordinary least squares regression of YoY growth on week number. A positive slope indicates growth is speeding up; a negative slope indicates growth is slowing down. Volatility is the standard deviation of the YoY growth series. It measures how consistent a brand's growth is. Lower volatility indicates more consistent growth.

A 52-week window was selected over a longer 104-week window. Recent trends are better captured by the shorter window and are diluted by the longer one. Acceleration and volatility were also computed on a QoQ basis to test whether the shorter period captured growth changes that YoY missed.

Features were computed at three earlier time points: 6, 12, and 18 months before the end of the data. At each time point, only data available up to that point is used to compute the growth series and summary features. The truth set labels remain the same because they are based on observed outcomes.

4.3.3 Forecast-Based Alternative. An alternative approach was explored using linear regression to forecast sales values, intended as a starting point that could be replaced by more sophisticated time series models in future iterations. A regression was fit on each brand's recent 104 weeks of case sales, and implied YoY growth was computed from the forecasted values. For most brands, the rankings were similar to the summary features approach. However, for brands with noisy, near-zero data, the regressions produced inflated growth estimates, and clearly inactive brands received the highest scores. Unlike the acceleration feature, which fits a line through the growth rates themselves, the forecast approach fits a regression on raw sales values and converts the predictions back to implied growth. This conversion amplified noise for brands with near-zero sales.

The summary features approach was more reliable in these cases because the median captures the typical growth level rather than fitting a trend line through noise. It is also more scalable, computing three statistics per brand rather than fitting a separate regression model for each of approximately 22,500 extensions.

4.4 Model Evaluation

Each modeling approach is evaluated on the hand-labeled truth set using the same metrics. The approaches compared are a rule-based baseline, MCDA scoring with a median threshold, the two binary logistic regression models, and the ordinal logistic regression model. Confusion matrices are reported in the results for each approach and classification task.

4.4.1 Evaluation Metrics. The primary evaluation metrics are balanced accuracy and macro F1. Balanced accuracy averages the recall for each class, so that a small class counts as much as a large one (Brodersen et al., 2010). Macro F1 averages the F1 score for each class and penalizes models that perform well on one class but poorly on another. Raw accuracy is reported for reference but can be misleading when classes are imbalanced. For the ordinal model, adjacent accuracy is also reported. It is the proportion of predictions that are within one tier of the true label.

4.4.2 Cross-Validation. With only 112 labeled brands, leave-one-out cross-validation (LOOCV) is used to produce out-of-sample predictions for the logistic regression models. MCDA weights were selected through sensitivity analysis rather than fitted to the data, so cross-validation is less critical for the MCDA scores.

4.4.3 Rule-Based Baseline. A rule-based labeling system serves as a baseline for comparison. The system flags weeks as inactive when distribution coverage (%ACV) falls below 1.5% or when sales and distribution are both very low. It counts consecutive inactive weeks and labels a brand as discontinued when three

conditions are met: the stretch of inactive weeks at the end of the data exceeds the brand's longest past inactive gap, mean distribution over the last 13 weeks is below 2%, and the last active week was at least 13 weeks before the end of the data. Parameters were tuned via grid search against the stakeholder truth set.

Brands not labeled as discontinued are then checked for steady decline using the ratio of average distribution in the second half of the brand's history to the first half. Brands where this ratio falls below 0.75 or distribution peaked above 2% in the brand's early history but averaged below 2% in the most recent 13 weeks are labeled as declining. The remaining brands are labeled as active. These three labels are also combined into a binary classification: low-performance (discontinued or declining) versus viable (active). The rule-based approach was developed as an initial method for identifying discontinued and declining brands and is used here as a benchmark. One limitation is that its fixed distribution thresholds do not account for brand size. Small niche brands with naturally low distribution can be labeled as discontinued even when they are actively selling, because their distribution falls below the same threshold used for larger brands.

4.4.4 Forecast Horizon Evaluation. The evaluation described above used features computed from the most recent 52 weeks of data. To test whether the framework can identify high-opportunity brands in advance, the same evaluation was repeated at three forecast horizons: 6, 12, and 18 months before the end of the data. At each horizon, features were computed from the 52 most recent weeks of data available at that time. Features were renormalized, MCDA scores recomputed, and the median threshold recalculated at each time point. Logistic regression models were retrained on the truth set using features computed from data available at that horizon and evaluated with LOOCV. Comparing accuracy across horizons shows how far in advance the framework can reliably predict brand performance.

4.4.5 External Validation. The stakeholder truth set is used to test whether the framework generalizes beyond the hand-labeled truth set. As noted in Section 4.1, these labels reflect business events rather than demand trajectory patterns and are therefore an imperfect proxy for validation. Models trained on the hand-labeled truth set are applied to the stakeholder brands, which were never seen during model fitting. The two truth sets differ in two additional ways: the stakeholder brands span multiple product categories rather than one, and they were labeled at the brand family level rather than the extension level.

Because the truth labels are at the family level, extension-level predictions are combined into a single family-level prediction before evaluation. Each extension is scored individually. For the logistic regression model, predicted probabilities are averaged across extensions, weighted by total case volume, and the 0.5 threshold is applied to the family-level average. For the MCDA score, extension-level scores are

averaged the same way, and the median threshold is applied. For the ordinal model, each extension is assigned a predicted tier, and the family is assigned the tier with the highest total case volume. MCDA scores are normalized using the min and max of the hand-labeled truth set, so a score of 0.5 on the stakeholder set means the same as a score of 0.5 on the validation set. Scores that fall outside the truth set range are set to 0 or 1.

Features are computed at a brand-specific cutoff: 6 months prior to each family's known event date. Results are reported at the family level only.

4.5 Application to the Full Dataset

To generate predictions on the full dataset, the logistic regression models trained on the hand-labeled truth set were used to classify all approximately 22,500 brand extensions. MCDA scores were computed using the same features and weights, with normalization applied across all extensions. Prior to normalization, level and QoQ volatility were winsorized at the 5th and 95th percentiles. Values outside this range were set to the nearest percentile boundary. This prevented brands with extreme growth rates at either end of the distribution from compressing all other scores toward the center.

4.6 Planned Framework Extensions

As described in Section 4.2.1, the MCDA framework defines four criteria. This study implements demand momentum. The remaining three are not yet implemented. Category context would measure whether a brand's product category is growing or shrinking overall, so that a brand's performance can be interpreted relative to its category's direction. Distribution coverage would measure the gap between a brand's current distribution and that of a leading brand in the same category, with larger gaps suggesting room for expansion. Promotional efficiency would measure how effectively a brand converts promotions into sales, using promotional versus non-promotional pricing and sales features.

A dashboard was developed to allow stakeholders to filter brand rankings and explore individual brand trajectories. Future work will allow uploading new data. This will make the framework a practical decision support tool rather than a static analysis.

5 Results

5.1 Feature Distributions by Opportunity Tier

Fig. 3 shows the distribution of the three demand momentum features across the four opportunity tiers. Level, the median YoY case sales growth rate, shows the clearest separation. Discontinued and decline brands are concentrated below zero, stable brands near zero, and growing brands above zero. Acceleration, the slope of the YoY growth series, shows little separation across tiers. Volatility, the standard deviation of the YoY growth series, has similar medians across tiers but substantially wider distributions for stable and growing brands. These patterns suggest that level is the primary differentiating feature, and that the secondary features contribute most at the lower end of the performance spectrum.

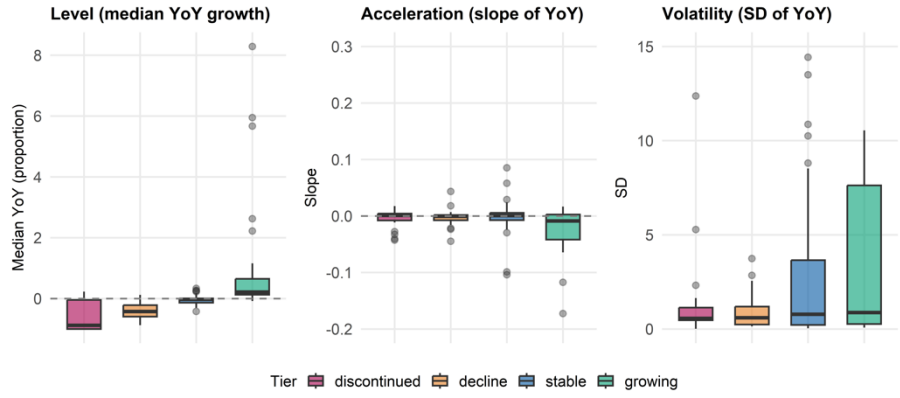

Fig. 3. Distribution of demand momentum features across the four opportunity tiers (n = 112). Level (median YoY case sales growth) shows clear separation across tiers, with discontinued and decline brands below zero, stable near zero, and growing above zero. Acceleration shows little separation, while volatility has similar medians but wider spread for higher tiers. Boxes show the interquartile range; the horizontal line inside each box is the median. The dashed line at zero indicates neutral growth. Features are expressed as proportions (YoY case sales growth rates). SD = standard deviation.

5.2 MCDA Score Configuration and Tier Separation

Fourteen weight configurations were evaluated across four feature combinations: level only (L), level plus volatility (LV), level plus acceleration (LA), and all three

features (LVA). Separation between adjacent tiers was measured as the difference in mean opportunity score between each pair of adjacent tiers. Table 2 reports mean scores by tier for selected configurations, and Table 3 reports the corresponding adjacent-tier gaps.

Table 2. Mean opportunity score by tier and feature configuration. Scores are normalized to a 0–1 scale within the 112-brand truth set. Parentheses show weight percentages (Level / Volatility). The selected configuration is marked with an asterisk.

Tier	L	LV YoY (85/15)	LV QoQ (85/15)	LV QoQ (90/10)*
Discontinued	0.038	0.173	0.138	0.105
Decline	0.064	0.199	0.191	0.149
Stable	0.105	0.220	0.224	0.184
Growing	0.238	0.325	0.329	0.299

Table 3. Adjacent-tier gap (difference in mean opportunity score between adjacent tiers) by feature configuration. Parentheses show weight percentages (Level / Volatility). The selected configuration is marked with an asterisk.

Boundary	L	LV YoY (85/15)	LV QoQ (85/15)	LV QoQ (90/10)*
Discontinued to Decline	0.026	0.026	0.053	0.044
Decline to Stable	0.041	0.021	0.033	0.035
Stable to Growing	0.134	0.105	0.105	0.115
Low-Performance to Viable	0.117	0.084	0.108	0.111

Mean opportunity scores increased from discontinued to growing across all configurations, indicating that higher-performing brands received higher scores. The L configuration produced the widest gap between stable and growing brands (0.134) and the widest binary gap between low-performance and viable brands (0.117). However, the discontinued-to-decline gap was narrow under L (0.026).

Adding volatility computed on a QoQ basis improved separation at the discontinued-to-decline boundary. The LV configuration with a 90/10 weight split (90% level, 10% QoQ volatility) produced a discontinued-to-decline gap of 0.044, compared to 0.026 under L, while retaining a stable-to-growing gap of 0.115 and a binary gap of 0.111. Adding QoQ acceleration as a third feature did not further improve separation at any boundary. The LV QoQ (90/10) configuration was selected as the primary scoring configuration because it improved separation at the most compressed boundary without substantially reducing separation elsewhere.

A robustness check using median-based gaps confirmed that the stable-to-growing gap under L is partially driven by outliers. The median gap between stable and growing brands is 0.029, compared to a mean gap of 0.134. The difference suggests that a few high-growth brands pull the mean upward. By contrast, the

discontinued-to-decline gap under LV QoQ (90/10) is more consistent across mean and median (0.044 and 0.058, respectively), suggesting that separation at that boundary is less influenced by extreme values than under the L configuration.

5.3 Out-of-Sample Model Performance

All metrics in this section are out-of-sample, based on LOOCV on the 112-brand truth set.

5.3.1 Binary Logistic Regression. For the low-performance versus viable (LP vs. Viable) outcome, level and volatility were the only significant predictors (level $p < 0.0001$, volatility $p = 0.042$). Acceleration was not significant in any model and did not improve fit (likelihood ratio (LR) test $p = 0.959$). Level + Volatility had the lowest AIC (75.3) and was selected. The model achieved 84% balanced accuracy and 84% macro F1. The model correctly classified 48 of 59 low-performance brands and 46 of 53 viable brands (Table 4).

Table 4. Confusion matrix for binary logistic regression, low-performance versus viable (LOOCV, n = 112). Rows are true labels; columns are predicted labels.

Truth \ Predicted	Low-Performance	Viable
Low-Performance	48	11
Viable	7	46

For the discontinued versus not-discontinued (Disc. vs. Not) outcome, level was the only significant predictor. Neither volatility nor acceleration improved fit over level alone (LR test $p = 0.454$ and 0.755, respectively). Level only had the lowest AIC (141.0) and was selected. The model achieved 75% balanced accuracy and 72% macro F1. The model correctly classified 21 of 29 discontinued brands and 64 of 83 not-discontinued brands.

A sensitivity check replacing YoY-based secondary features with QoQ-based equivalents did not improve either binary model. For the low-performance versus viable outcome, YoY volatility (AIC 75.3) outperformed QoQ volatility (AIC 80.0), and QoQ volatility was not significant ($p = 0.212$). For the discontinued model, neither QoQ volatility nor QoQ acceleration improved on level alone.

5.3.2 Ordinal Logistic Regression. Level was the only significant predictor across all four-tier models ($p < 0.0001$). Neither volatility nor acceleration was significant (volatility $p = 0.158$, LR test $p = 0.149$; acceleration $p = 0.765$, LR test $p = 0.761$). Level only was selected (AIC 207.8). Level + Volatility had nearly identical fit (AIC 207.7), but volatility did not reach significance, so the simpler model was preferred.

QoQ-based features did not improve fit. The model achieved 60% balanced accuracy and 61% macro F1. Adjacent accuracy was 93%, meaning that in 104 of 112 cases the predicted tier was within one step of the true tier (Table 5). Per-class recall was 66% for discontinued, 47% for decline, 64% for stable, and 64% for growing. Decline was the hardest tier to classify.

Table 5. Confusion matrix for ordinal logistic regression, four-tier classification (LOOCV, n = 112). Rows are true labels; columns are predicted labels.

Truth \ Predicted	Discontinued	Decline	Stable	Growing
Discontinued	19	2	7	1
Decline	9	14	7	0
Stable	0	5	18	5
Growing	0	0	9	16

5.3.3 MCDA Opportunity Score. For comparison with the regression models, the LV configuration was converted to a binary classification using the median score as a threshold. The MCDA LV score achieved 85% balanced accuracy on the low-performance versus viable task (Table 6). Unlike the regression models, the MCDA score produces a continuous ranking that does not require a fixed classification threshold.

5.3.4 Summary. Table 6 compares all approaches. The MCDA LV score and binary logistic regression perform comparably on the low-performance versus viable task, both outperforming the rule-based baseline (Section 4.4.3). For MCDA configurations, volatility is computed on a QoQ basis when included, whereas the logistic regression models use YoY-based features. The ordinal model distinguishes between all four tiers at the cost of lower overall balanced accuracy.

Table 6. Out-of-sample performance comparison across all approaches (LOOCV, n = 112). Balanced accuracy and macro F1 are the primary metrics. Adjacent accuracy for the ordinal model (93%) is reported in the text. Bal. Acc. = balanced accuracy; Raw Acc. = raw accuracy.

Approach	Task	Bal. Acc.	Macro F1	Raw Acc.
Rule-based baseline	LP vs. Viable	75%	75%	76%
MCDA LV	LP vs. Viable	85%	85%	85%
Binary logistic regression	LP vs. Viable	84%	84%	84%
Binary logistic regression	Disc. vs. Not	75%	72%	76%
Ordinal logistic regression	4-Tier	60%	61%	60%

5.4 Forecast Horizon Evaluation

To test whether the framework can identify brand performance in advance, the evaluation was repeated at three earlier horizons: 6, 12, and 18 months before the end of the observation window. At each horizon, features were recomputed from the 52 most recent weeks of data available up to that point. Truth labels remained the same. A small number of brands are excluded at longer horizons because they do not have enough data history to compute YoY or QoQ features, reducing the sample from 112 to 106 at 18 months.

Table 7 reports balanced accuracy across approaches and horizons. Performance does not degrade monotonically; several approaches dip at 12 months before partially recovering at 18 months. With only 106 to 112 brands, a small number of brands shifting across the decision boundary at a given horizon could move accuracy by several percentage points.

Table 7. Balanced accuracy (%) by approach and forecast horizon. Features are recomputed from the 52 most recent weeks available at each horizon. Truth labels remain the same. N decreases at longer horizons as some brands lack sufficient history to compute YoY or QoQ features. Logistic regression results are LOOCV; MCDA results use the median threshold.

Approach	Full Data	6 Months	12 Months	18 Months
N	112	112	110	106
MCDA L	85%	83%	80%	82%
MCDA LV	85%	83%	82%	72%
Binary logistic regression (LP vs. Viable)	84%	80%	77%	82%
Binary logistic regression (Disc. vs. Not)	75%	78%	73%	75%
Ordinal logistic regression (balanced)	60%	69%	57%	57%
Ordinal logistic regression (adjacent)	93%	93%	93%	93%

MCDA L (level only) is the most consistent approach, with balanced accuracy ranging from 80% to 85% across all horizons. The binary logistic regression (LP vs. Viable) model shows a similar pattern (84% full data, 80% at 6 months, 77% at 12 months, 82% at 18 months). The binary logistic regression (Disc. vs. Not) model is also stable, though consistently lower (75% full data, 78% at 6 months, 73% at 12 months, 75% at 18 months). MCDA LV degrades more sharply at 18 months (85% to 72%).

The ordinal model balanced accuracy ranges from 57% to 69% across horizons, with a spike at 6 months (69%). Adjacent accuracy is stable at 93% across all horizons, meaning the ordinal model rarely misclassifies a brand by more than one tier at any of the three tested time horizons.

Although MCDA L is more stable across horizons, MCDA LV widens the gap between the discontinued and decline tiers and performs comparably at the 6-

month horizon. Fig. 4 shows MCDA LV scores for all 112 brands at the 6-month horizon, sorted lowest to highest, with bar height indicating score and bar color indicating truth tier. Discontinued and decline brands are concentrated at the left of the distribution and stable and growing brands at the right, with mixing in the middle range. The Logistic and Ordinal rows show model predictions for each brand in the same order. Both models align with the truth tier colors for most brands.

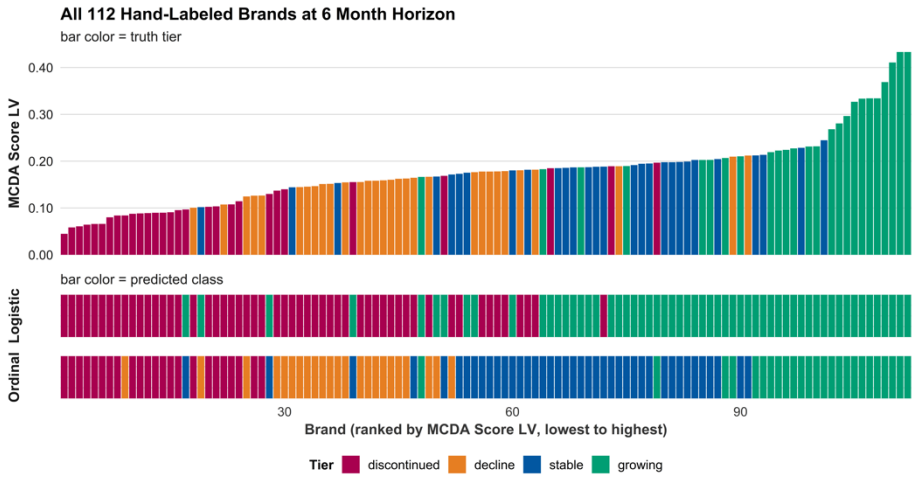

Fig. 4. MCDA LV score and model predictions for all 112 hand-labeled brands at the 6-month horizon. Brands are ranked left to right by MCDA LV score (lowest to highest). Bar height indicates the MCDA score (*top*); bar color indicates the truth tier (*top*) or predicted class (*Logistic and Ordinal rows*). Logistic predicted class: red = low-performance, green = viable. Ordinal and truth tier colors follow the legend.

5.5 External Validation

The 18 stakeholder-labeled brand families (10 bankrupt, 8 acquired) were scored using features computed 6 months prior to each family's known event date, as described in Section 4.4.5. All three approaches achieved similar balanced accuracy: 61% for MCDA LV, 63% for the binary logistic regression model, and 63% for the ordinal model collapsed to binary.

Acquired brands received higher MCDA LV scores on average than bankrupt brands, but the two groups overlap substantially in the middle of the score range. Brands at the top and bottom of the score distribution were correctly classified by all three approaches. Misclassifications were concentrated among bankrupt brands with scores in the 0.17 to 0.19 range, where demand signals alone were not

sufficient to distinguish them from acquired brands. Fig. 5 shows this distribution, along with logistic model predictions.

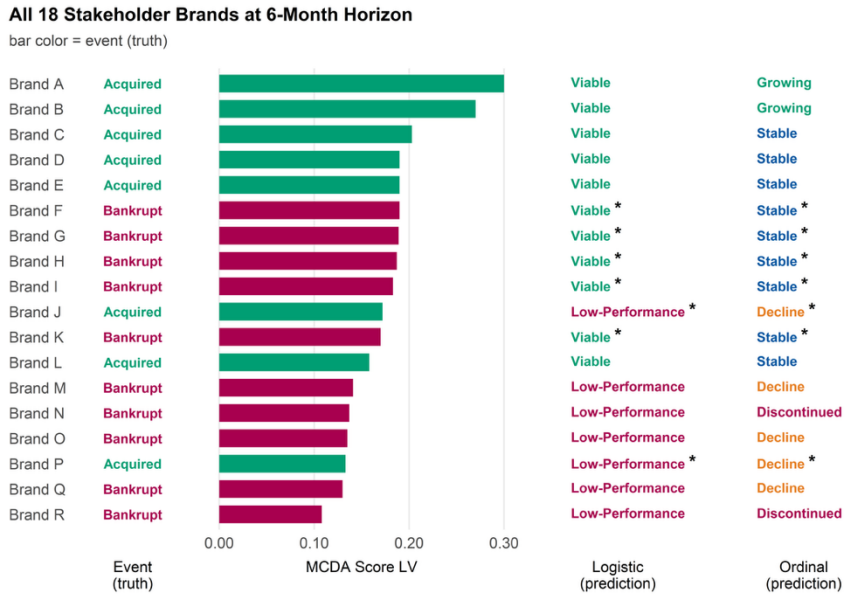

Fig. 5. MCDA LV score and model predictions for all 18 stakeholder-labeled brand families at the 6-month horizon. Brands are ranked top to bottom by MCDA LV score (highest to lowest). Bar color indicates event (truth), and logistic colors indicate predicted class: green = acquired/viable, red = bankrupt/low-performance. Ordinal prediction colors follow the truth tier palette. Asterisks indicate misclassifications.

5.5.1 Binary Logistic Regression. The model correctly classified 5 of 10 bankrupt brands and 6 of 8 acquired brands (Table 8). The five misclassified bankrupt brands had predicted probabilities between 0.55 and 0.66.

Table 8. Confusion matrix for binary logistic regression (n = 18). Features were computed 6 months prior to each family's event date. Rows are true events (bankrupt as a proxy for low-performance, acquired as a proxy for viable); columns are predicted classes.

Truth \ Predicted	Low-Performance	Viable
Bankrupt	5	5
Acquired	2	6

5.5.2 Ordinal Logistic Regression. Table 9 shows the full four-tier predictions. No bankrupt brand was predicted as growing, and no acquired brand was predicted as discontinued. Of the 10 bankrupt brands, 5 were predicted in a higher tier than expected, while 2 of 8 acquired brands were predicted in a lower tier.

Table 9. Confusion matrix for ordinal logistic regression, four-tier classification (n = 18). Features were computed 6 months prior to each family's event date. Rows are true events (bankrupt or acquired); columns are predicted tiers.

Truth \ Predicted	Discontinued	Decline	Stable	Growing
Bankrupt	2	3	5	0
Acquired	0	2	4	2

5.5.3 MCDA Opportunity Score. The continuous MCDA LV score captures both the ordering and the distance between brands, providing more information than a binary classification. Acquired brands cluster at higher scores and bankrupt brands at lower scores, with overlap in the middle. Fig. 5 shows this distribution.

5.6 Application to the Full Dataset

The scoring framework and regression models were applied to the full dataset of approximately 22,500 brand extensions to examine how scores and classifications distribute across all brands. MCDA scores were computed using the selected LV QoQ (90/10) configuration. Winsorization at the 5th and 95th percentiles was applied to both features before min-max normalization, as described in Section 4.5. A total of 1,978 extensions lacked sufficient observations to compute growth features and were excluded from scored results.

The distribution of MCDA LV scores across all brand extensions is shown in Fig. 6. Scores span the full 0–1 range, with most brands concentrated in the middle of the distribution, fewer in the lower range, and relatively few at the highest score levels. The concentration near 1.0 reflects brands whose growth rates exceeded the 95th percentile winsorization threshold.

Using the binary logistic regression model, 54% of extensions were classified as viable and 46% as low-performance. The ordinal model assigned 21% of extensions to discontinued, 22% to decline, 49% to stable, and 8% to growing. Stable was the most common predicted tier, while growing was the least common. The combined discontinued and decline proportion from the ordinal model (43%) is consistent with the low-performance proportion from the binary model (46%).

Across the 20,517 extensions with sufficient data, the model results identify a subset for review as high-potential brands. To support this use, scored outputs were compiled into an interactive HTML dashboard allowing stakeholders to filter and review individual brand scores and model predictions.

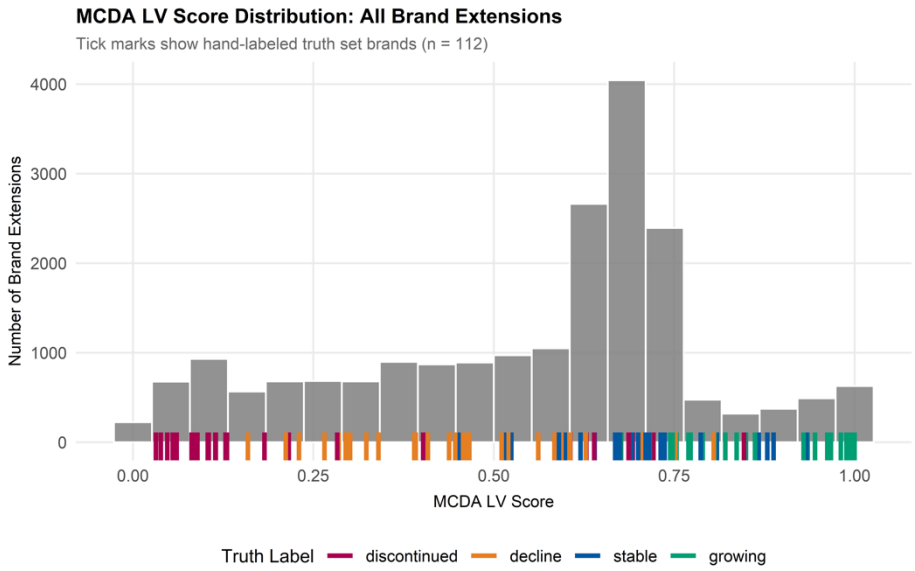

Fig. 6. MCDA LV score distribution across all 20,517 scored brand extensions. Tick marks along the x-axis show the 112 hand-labeled truth set brands, colored by opportunity tier.

6 Discussion

6.1 Interpretation of Results

Median YoY case sales growth (level) was the dominant predictor across all models and configurations. In the binary logistic regression model for the low-performance versus viable outcome, level was the primary predictor ($p < 0.0001$) and volatility was marginally significant ($p = 0.042$); acceleration was not a significant predictor in any model. In the discontinued versus not-discontinued binary model and the ordinal model, level was the only significant predictor. In the MCDA configurations, level-only scoring produced the widest separation between stable and growing brands and the widest binary gap between low-performance and viable brands.

The practical implication depends on the use case. When the primary goal is identifying high-potential brands for investment or acquisition, level alone maximizes separation at the higher end of the spectrum. When the goal also includes identifying brands that are not worth retaining, adding QoQ volatility

improves the ability to distinguish discontinued from declining brands. MCDA LV is better suited for decisions that span both ends of the distribution. MCDA L is sufficient when the focus is identifying high-potential brands.

Performance was fairly stable across forecast horizons, though not uniform. MCDA L was the most consistent approach across all three horizons. MCDA LV degraded more sharply at 18 months, likely because QoQ volatility is less stable over time than level. Median YoY growth tends to be consistent across horizons, while QoQ variation in growth may change as a brand matures, for example shifting from an erratic early growth pattern to a more stable one. The dips and spikes at 6 and 12 months across several approaches, including the ordinal model's jump to 69% at 6 months, are likely due to the combination of week-to-week variability in YoY growth and the small truth set. Weekly YoY growth is sensitive to which specific weeks fall in the 52-week window, and with a small truth set that variability is not smoothed out across brands. The choice between L and LV depends on both the decision goal and the horizon. For longer-horizon decisions, level alone is the more stable choice.

The binary and ordinal regression models address the same question with different levels of detail. The binary model answers whether a brand is worth attention at all, with higher accuracy than the ordinal model. The ordinal model answers where in the performance spectrum a brand falls, with more granularity but lower overall accuracy. Adjacent accuracy was 93% across all forecast horizons, meaning when it did misclassify a brand, it almost always placed it in a neighboring tier. This has practical value because a brand incorrectly placed in the growing tier is unlikely to be low-performing, and a brand incorrectly placed in the discontinued tier is unlikely to be viable. The MCDA score, binary prediction, and ordinal tier can also be used together. Agreement across the three strengthens confidence in a decision. Disagreement flags a brand for closer review.

The external validation results should be interpreted carefully. Balanced accuracy of 61–63% on the 18-brand stakeholder sample reflects meaningful signal. Bankrupt brands tended to score lower, and acquired brands tended to score higher, consistent with the pattern of low-performing and viable brands. However, these are business events with causes that may not appear in retail sales data at all, including debt and cash flow problems, operational challenges, and ownership changes. A brand may file for Chapter 11 reorganization while still selling product on shelves, which would not appear as a demand collapse. These outcomes serve as proxies for low-performing and viable brands, not as ground truth labels.

What the result demonstrates is that demand signals computed six months in advance can partially separate these two groups, which is informative though imperfect. Misclassifications were concentrated among bankrupt brands in the middle of the score distribution, where retail sales data alone were not sufficient to predict the event. Prior work shows that distribution coverage explains substantial

variation in brand sales performance (Bronnenberg & Sismeiro, 2002), suggesting that adding distribution and promotional features in addition to demand momentum could improve classification.

6.2 Implications

The framework provides a structured and transparent approach to evaluating emerging brands under conditions of limited ground truth. Rather than requiring a large pre-labeled dataset, the framework infers opportunity tiers from observed sales and distribution behavior and a small hand-labeled truth set, making it applicable in settings where explicit outcome labels are not available.

The three approaches work together at different levels of resolution. The binary model identifies whether a brand warrants closer attention. The ordinal model places it in one of four performance tiers. The MCDA score provides a continuous ranking with finer detail than a tier assignment alone. Used in combination, the three approaches support consistent decision making. Because the scoring weights and criteria are explicit, practitioners can see why a brand received a particular score and adjust the framework to reflect different priorities. In practice, the results can be used to identify brands that meet MVP criteria for investment or acquisition, brands to monitor, and brands to deprioritize, based on observed performance patterns in the data.

6.3 Limitations

The size and scope of the hand-labeled truth set is an important limitation of this study. The 112 brands used for model training and evaluation come from a single product subcategory. Results may not generalize well to other subcategories with different growth and discontinuation patterns, variability, and seasonality. With 112 brands, a small number of brands near a decision boundary can shift accuracy estimates, which contributes to the variability seen across forecast horizons. Expanding the truth set to include additional subcategories would strengthen confidence in the model estimates.

The rule-based labeling approach used early in the analysis applies a single distribution threshold regardless of brand size. Small, newer, or niche brands can be flagged as inactive even when they are genuinely active. This can lead to disproportionate low-performance labeling for smaller brands. The model-based approaches are less sensitive to this because they use growth rates rather than absolute sales and distribution levels. A small brand and a large brand growing at the same rate are treated the same way, which is one reason the model-based approaches outperformed the rule-based baseline.

The scoring framework applies to all brands with sufficient data history, but not all brands are realistic targets for investment or acquisition. Well-known high-volume brands that occupy shelf space across nearly all retailers are unlikely acquisition or incubation candidates. Growth expectations differ between small emerging brands and large established ones. Scores are most meaningful when comparing brands within a similar size and maturity range. Including large established brands in the scoring pool can compress or alter the scores of smaller emerging brands because all brands are scored relative to one another. To best serve its intended purpose, the truth set and model training should focus on smaller emerging brands rather than large established ones.

6.4 Future Work

There are several ways to extend and improve this framework. The most straightforward is expanding the hand-labeled truth set to include additional subcategories. The labeling workflow developed for this study could be extended to 200-300 brands across two or three subcategories with minimal additional effort. This would allow model estimates, scores, and tier assignments to be evaluated across a broader range of brand patterns.

The current framework uses demand momentum as the only input. A natural next step is incorporating additional performance signals, including distribution coverage and headroom for growth, promotional efficiency, category context, and social trend data. Category context is worth prioritizing. Scoring brands relative to their own category would make scores more interpretable and fair. A brand declining in a declining category may still be performing well relative to its peers, while the same pattern in a growing category is a stronger warning sign. Highly seasonal brands may also benefit from being modeled separately, since their growth patterns differ from year-round brands.

The demand features used in this study are summary statistics of past growth. Statistical time series models such as autoregressive integrated moving average (ARIMA) are designed to handle seasonality and autocorrelation. These could provide forward-looking demand estimates that serve directly as MCDA inputs. However, forecast-based demand estimates are absolute rather than relative. The advantage of the current summary features is that growth rates preserve comparability across brands of different sizes. This tradeoff warrants further consideration.

Divergence between dollar and case sales may indicate that growth is driven by price changes rather than underlying demand. This relationship could be incorporated as an additional feature, being careful to avoid multicollinearity with existing demand and promotional features.

Finally, establishing a feedback loop would strengthen the framework over time. As stakeholders use the framework to make decisions and those outcomes are later observed, the results can be used to retrain the model and improve accuracy. This would generate real labeled data organically and reduce dependence on the hand-labeled truth set over time.

6.5 Ethics

The data used in this analysis are proprietary and were accessed under a non-disclosure agreement. All results are presented in anonymized form. The framework is intended to support investment decision making and should be used alongside domain knowledge and other available information. Brand performance scores are based on observed retail sales patterns and do not account for all factors that influence business outcomes. Investment decisions carry inherent risk and should not be made based on this framework alone.

7 Conclusion

This study demonstrates that a small set of demand-based features can distinguish brand performance tiers. Median year-over-year case sales growth was the dominant predictor across all models and configurations. The framework provides a forward-looking signal of brand performance. Most approaches maintained stable accuracy for 12 months, and the binary model and MCDA L remained stable up to 18 months.

The framework is a proof of concept designed to grow. It is validated by statistical models and is designed to expand as new criteria are tested and additional ground truth outcome data become available. The current study uses demand momentum as the only input, and future work will include distribution coverage, promotional efficiency, category context, and social trend data alongside a larger and more diverse truth set. The goal is a structured, transparent, and scalable tool that supports consistent investment decision making in the consumer-packaged goods industry.

Acknowledgments. The authors thank the contributing brokerage/distribution agency for providing access to the data used in this study, for their domain knowledge and guidance, and for their feedback on this work. The authors also thank Jacquelyn Cheun and Diana Sherman for feedback on the manuscript and presentation, Nibhrat Lohia for support throughout the course, students of DS6210

Capstone at SMU for feedback on the presentation and slide deck, and colleagues at Howard Hughes Medical Institute for their comments.

References

1. Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Wiley. https://doi.org/10.1002/9780470594001
2. Ananth, C. V., & Kleinbaum, D. G. (1997). Regression models for ordinal responses: A review of methods and applications. *International Journal of Epidemiology, 26*(6), 1323. https://doi.org/10.1093/ije/26.6.1323
3. Arora, T., Chandna, R., Conant, S., Sadler, B., & Slater, R. (2020). Demand forecasting in wholesale alcohol distribution: An ensemble approach. *SMU Data Science Review, 3*(1), Article 7. https://scholar.smu.edu/datasciencereview/vol3/iss1/7
4. Bozorg-Haddad, O., Zolghadr-Asli, B., & Loáiciga, H. A. (2021). *A handbook on multi-attribute decision-making methods* (1st ed.). Wiley. https://doi.org/10.1002/9781119563501
5. Brodersen, K. H., Cheng, S. O., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In *Proceedings of the 20th International Conference on Pattern Recognition* (pp. 3121–3124). IEEE. https://doi.org/10.1109/ICPR.2010.764
6. Bronnenberg, B. J., & Sismeiro, C. (2002). Using multimarket data to predict brand performance in markets for which no or poor data exist. *Journal of Marketing Research, 39*(1), 1–17. https://doi.org/10.1509/jmkr.39.1.1.18939
7. Drugova, T., & Curtis, K. R. (2022). Why can't the supply chain keep up with organic bakery product demand? Understanding miller, distributor, and baker organic wheat quality perceptions and needs. *The International Food and Agribusiness Management Review, 25*(4), 601–618. https://doi.org/10.22434/IFAMR2021.0138
8. Ford, J., Nava, C., Tan, J., & Sadler, B. (2020). Automated machine learning framework for demand forecasting in wholesale beverage alcohol distribution. *SMU Data Science Review, 3*(3), Article 7. https://scholar.smu.edu/datasciencereview/vol3/iss3/7
9. Giannikos, C. I., & Korkou, E. D. (2025). Financial literacy and credit card payoff behaviors: Using generalized ordered logit and partial proportional odds models to measure American credit card holders' likelihood of repaying their credit cards. *International Journal of Financial Studies, 13*(1), 22. https://doi.org/10.3390/ijfs13010022
10. Jiang, L., Rollins, K. M., Ludlow, M., & Sadler, B. (2020). Demand forecasting for alcoholic beverage distribution. *SMU Data Science Review, 3*(1), Article 5. https://scholar.smu.edu/datasciencereview/vol3/iss1/5
11. Khan, S. A., Chaabane, A., & Dweiri, F. T. (2018). Multi-criteria decision-making methods application in supply chain management: A systematic literature review. In V. A. P. Salomon (Ed.), *Multi-criteria methods and techniques applied to supply chain management* (pp. 3–31). IntechOpen. https://doi.org/10.5772/intechopen.74067
12. Kizielewicz, B., Więckowski, J., Shekhovtsov, A., Wątróbski, J., Depczyński, R., & Sałabun, W. (2021). Study towards the time-based MCDA ranking analysis – a supplier

selection case study. *Facta Universitatis. Series: Mechanical Engineering, 19*(3), 381–399. https://doi.org/10.22190/FUME210130048K

13. Lee, H. L. (2002). Aligning supply chain strategies with product uncertainties. *California Management Review, 44*(3), 105–119. https://doi.org/10.2307/41166135
14. Michis, A. A. (2023). Retail distribution evaluation in brand-level sales response models. *Journal of Marketing Analytics, 11*(3), 366–378. https://doi.org/10.1057/s41270-022-00165-8
15. Olson, D. L. (2025). How to use multiple criteria selection methods to select supply chain vendors in business operations [How-to guide]. *SAGE Research Methods: Business*. SAGE Publications. https://doi.org/10.4135/9781036213572
16. Parry, S. (2024). *Ordinal logistic regression models and statistical software: What you need to know*. Cornell Statistical Consulting Unit. https://cscu.cornell.edu/wp-content/uploads/ordlogistic.pdf
17. Punia, S., & Shankar, S. (2022). Predictive analytics for demand forecasting: A deep learning-based decision support system. *Knowledge-Based Systems, 258*, 109956. https://doi.org/10.1016/j.knosys.2022.109956
18. Rojas, F., Rojas, J., & Wanke, P. (2024). Enhancing supply chain decision-making through forecast fit: A qualitative and quantitative analysis. *International Journal of Professional Business Review, 9*(8), 4880. https://doi.org/10.26668/businessreview/2024.v9i8.4880
19. Vafaei, N., Ribeiro, R. A., & Camarinha-Matos, L. (2022). Assessing normalization techniques for simple additive weighting method. *Procedia Computer Science, 199*, 1229–1236. https://doi.org/10.1016/j.procs.2022.01.156
20. Victory, K. (2017). *Understanding new product performance: A descriptive investigation across multiple categories and two countries* (Master's thesis, University of South Australia). https://searchlibrary.adelaide.edu.au/permalink/61USOUTHAUS_INST/17gltgr/alma9916147810401831