Vol. 19 (2026): Proceedings of 2026 International Conference on Financial Technology, Mathematical Statistics, and Business Analytics (ICFMB 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/icfmb.202619568574

Application of Regression Analysis in Pricing of Fast-Moving Consumer Products

Wenxi Wang*

Sun Yat-sen University, Guangzhou, China

**Corresponding author: Wenxi Wang.*

Abstract

Against the backdrop of fiercer competition and diversified demands in the fast-moving consumer goods (FMCG) market, scientific new-product pricing is critical for enterprises to capture market share and maximize profits, yet most still rely on experience-based methods. Based on five years of relevant literature, this paper explores regression analysis's application in FMCG pricing: it clarifies the variable selection logic of pricing regression models, compares the fitting effects and applicable scenarios of mainstream and specialized regression methods, identifies practical implementation obstacles, and discusses the potential of integrating machine learning with regression analysis to capture product substitution effects and optimize profit-maximizing price combinations, providing targeted guidance for FMCG pricing research and practice.

Keywords

regression analysis, fast-moving consumer products, new-product pricing

1. Introduction

In recent years, regression analysis has assumed a growingly pivotal role in the business landscape. Amid the intensifying competition within the consumer goods market and the ever-growing diversification of consumer demands, the pricing strategy for new fast-moving consumer goods (FMCG) has emerged as a core link for enterprises striving to seize market share and maximize profits. Yet, a host of challenges persist in this domain: new FMCG products have extremely low tolerance for pricing deviations, yet most enterprises still rely on experience-driven pricing methods. Only a handful of leading brands have ventured to adopt statistical tools for pricing, and even these attempts often suffer from the issue of "tool-market mismatch".

As a time-tested predictive and attribution tool in statistics, regression analysis enables systematic disassembly and integration of pricing-related factors by constructing quantitative correlation models between independent and dependent variables. This provides enterprises with scientific, actionable decision support for new FMCG pricing. This paper aims to sort through the selection logic of core independent variables, dependent variables, and moderating variables in FMCG pricing regression models as documented in academic literature over the past five years. It will also review practical application cases of linear regression, multiple regression, and nonlinear regression in new FMCG pricing scenarios, examining each model from the perspectives of adaptability, variable selection rationale, and empirical application contexts.

Furthermore, the paper will conduct a comparative analysis of the fitting effects and applicable conditions of different regression models, while delving into the existing bottlenecks of this technology—in particular, constraints posed by data limitations and challenges in adapting to dynamic market conditions. Ultimately, it seeks to offer theoretical references and directional guidance for both follow-up academic research and industry practice in this field.

2. Theoretical Framework on FMCG Regression Model

2.1 Model Overview

Over the past five years, the academic community has conducted multi-dimensional explorations on the variable system of FMCG pricing regression model. The exploration has evolved from traditional basic indicators such as cost and competitor prices to dynamic variables like consumer channel preferences and promotional elasticity, and further to regulatory factors under algorithm optimization. The logic of variable selection has gradually shifted from a single cost-oriented approach to a multi-coordinated orientation that integrates cost, demand, and market environment—reflecting the industry's deepening understanding of how complex factors interact to shape FMCG pricing outcomes. For instance, Dachyar selected five key independent variables, as shown in Table 1, namely: product price, product cannibalization, price disparity, festival days, and weather (Rainfall). The rigorous logic behind the author's variable selection, particularly the way it aligns with FMCG's unique market dynamics, is worth in-depth study. As can be seen in Table 2, Dachyar initially selected 14 variables [1], and his stepwise screening process offers valuable insights into effective variable design for FMCG pricing models.

Table 1: Multi Linear Regression Analysis

Model		Unstandardized Coefficients		Standardized Coefficient	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper bound
1	(Constant)	-277.909	54,082.516		-0.005	0.966	-110,729.142	110,173.323
	Rain fall	-12.501	4.817	-0.292	-2.595	0.014	-22.339	-2.663
	Population	0.001	0.000	0.041	0.490	0.627	0.000	0.001
	Day of delivery	452.681	179.768	0.195	2.518	0.017	85.546	819.816
	Cannibalization	0.969	0.925	0.099	1.048	0.303	-0.919	2.857
	Price difference	-3.298	0.505	-0.652	-6.535	0.000	-4.329	-2.268

Table 2: Factors Affecting Demand

No.	Factor	No.	Factor
1	Number of Consumers	8	Holiday
2	Prices	9	Weather
3	Type of Consumers	10	Future Forecast
4	Cannibalization	11	Advertisement
5	Product Quality	12	Product Shelf Life
6	Income	13	Number of Stores
7	Prices of Competitor	14	Type of Sales

2.2 Variable Design

In response to the low accuracy problem of traditional historical data prediction methods (time series, qualitative methods) with a Multi Linear Regression Analysis (MAPE) over 20%, the author broke away from the reliance on a single type of data and introduced non-sales multi-dimensional variables (such as climate and festivals). At this point, he selected 8 variables: number of consumers, price, cannibalization, income, price of competitors / other items, holiday, weather, and number of stores.

In detail, firstly the author combined the characteristics of the specific FMCG product in focus (pesticides) — whose demand is directly affected by seasonality (weather), consumption scenarios (festivals), and market competition (price differences, cannibalization effect) — to screen out variables strongly associated with product demand. Here, the inclusion of product cannibalization (the substitution effect) holds significant practical relevance for FMCG brand operations. Secondly, another key insight from Dachyar's variable selection lies in the emphasis on festival days and weather—two factors that exert a far more pronounced influence on FMCG than on durable goods, and understanding this sensitivity is essential to justifying their inclusion in the regression model. The fundamental difference between FMCG and durable goods lies in their consumption patterns: FMCG products are consumed quickly, require regular repurchase, and are often tied to immediate or short-term needs, while durable goods (e.g., electronics, furniture, appliances) are purchased infrequently and serve long-term functional purposes. This distinction explains why FMCG demand is far more susceptible to weather and festival fluctuations.

2.3 Price Elasticity

Among these, the price variable carries clear economic implications centered on price elasticity—a core concept in FMCG pricing that measures the degree to which consumer demand for a product responds to changes in its price. Unlike durable goods, which often have inelastic demand due to their long service life and high purchase threshold, FMCG products (especially daily necessities, snacks, and personal care items) typically exhibit higher price elasticity: small fluctuations in price can trigger significant shifts in consumer choices, as buyers tend to have lower switching costs and more substitute options. For example, a 5% price increase for a bottle of laundry detergent may lead some consumers to switch to a competing brand with a more favorable price point, while a temporary price reduction often drives a noticeable surge in sales. By incorporating the price variable into the regression model, Dachyar was able to quantify this elasticity, enabling more precise predictions of how pricing adjustments would impact overall sales volume and profit margins—addressing a critical practical need for FMCG brands that operate in highly competitive, low-margin markets.

2.4 Cannibalization

The inclusion of product cannibalization (the substitution effect) holds significant practical relevance for FMCG brand operations. In the FMCG sector, brands often maintain diverse product lines within the same category (e.g., different sizes of beverages, varying formulations of skincare products) or launch new SKUs to cater to segmented market needs. However, such product proliferation can lead to cannibalization: when a new product is priced too close to an existing one, or targets an overlapping consumer group, it may divert sales from the original product rather than expanding the brand's total market share. For instance, if a brand launches a mid-priced fruit juice alongside its existing premium juice line without sufficient differentiation, price-sensitive consumers who previously purchased the premium variant may switch to the cheaper option, while new customers remain limited—resulting in no net growth in overall revenue. By treating cannibalization as an independent variable, Dachyar's model accounts for this internal competition, helping brands balance product portfolio expansion with the need to protect existing sales. This is particularly valuable for FMCG brands, as their short product life cycles and frequent new product launches make cannibalization a common and costly challenge; accurately measuring this effect allows for more informed pricing and product positioning decisions.

2.5 Weather

For weather, changes in temperature, precipitation, or seasonal conditions directly drive immediate consumption needs: hot weather boosts sales of bottled water, ice cream, and sunscreen; rainy days increase demand for umbrellas, wet wipes, and instant meals; and cold seasons drive up purchases of warm beverages and moisturizers. In contrast, durable goods purchases are rarely impelled by short-term weather changes—consumers are unlikely to buy a new refrigerator or sofa simply because of a heatwave or rainstorm.

2.6 Festivals

Festivals, meanwhile, create concentrated consumption scenarios for FMCG: holidays such as Lunar New Year, Christmas, or local cultural festivals often involve family gatherings, gift-giving, or travel, leading to surge in demand for food and beverages, personal care products, and small daily necessities. For example, snack and beverage sales typically spike during holiday seasons as households stock up for gatherings, while travel-sized toiletries see increased demand from travelers. Durable goods, by contrast, are rarely tied to festival-related consumption—purchases are more often driven by life events (e.g., moving to a new home, replacing a broken appliance) or long-term planning, making festival periods have little impact on their demand. By incorporating weather and festival days as variables, Dachyar's model captures these short-term, demand-driven fluctuations that are inherent to FMCG, addressing a gap in traditional pricing models that often overlooked such context-dependent factors.

2.7 Multiple Linear Regression

Finally, through multiple linear regression (MLR), Dachyar selected effective variables and eliminated redundant factors, ultimately retaining five core variables and reducing the MAPE to 8.66%—achieving a balance between prediction accuracy and variable explanatory power. This outcome underscores the importance of aligning variable selection with the unique characteristics of FMCG products: by incorporating the economic logic of price elasticity, the practical implications of cannibalization, and the context-specific sensitivity to weather and festivals, the model not only improves predictive performance but also provides actionable insights for brand managers. As the FMCG industry continues to evolve with changing consumer behaviors and market dynamics, such context-aware variable design will remain a key direction for future research on pricing regression models—helping brands navigate the sector's inherent instability and time sensitivity to achieve sustainable profit growth.

3. Literature Review

3.1 The Linear Regression

In the study on vegetable retail pricing, Zhu took "cost-plus pricing" as the sole independent variable and "quarterly sales volume" as the dependent variable to construct a linear regression model [2]. After optimization with the PSO particle swarm algorithm, a linear equation was obtained, with a P-value of 0.032564 and an R^2 coefficient of 0.9997, proving a strong linear correlation between cost-plus pricing and sales volume. Based on this, a basic pricing and replenishment strategy for supermarket vegetables was formulated for the period from July 1st to 7th.

In the field of FMCG beverages, when a company analyzed the reference value of a single competitor's price for the pricing of a new product, it took "the unit price of the direct competitor" as the independent variable and "the estimated sales volume of the new product" as the dependent variable to establish a linear regression. When the direct competitor's product was priced at 35 yuan for 500ml with the claim of "0 sugar, 0 calories", the model calculated that pricing the new product at 29.9 yuan could balance sales and profit. Subsequently, by combining a "buy two, get one free" promotion, the decision-making threshold for consumers was further lowered, demonstrating the practicality of linear regression in single-variable pricing reference.

In multiple regression: Multiple independent variables can be included to precisely capture the synergistic influence of multiple factors in the pricing of FMCG new products, making it a core tool in complex scenarios. For instance, in the pricing prediction of pesticides, a fast-moving consumer good, Dachyar broke through the limitation of using only historical sales data and included "product price, cannibalization effect, price difference, number of festival days, and weather" as independent variables and "product demand" as the dependent variable to construct a multiple linear regression model [1]. Compared to the traditional time series method with a MAPE over 20%, this model reduced the MAPE to 8.66%, formulating a more accurate pricing strategy for new pesticide products and effectively improving the accuracy of demand prediction.

Finally, in nonlinear regression: It is applicable to scenarios where there are dynamic and nonlinear relationships between variables in the pricing of FMCG new products, especially in dynamic pricing and demand learning. A study proposed the "Elastic - ARIMA demand model", combining price elasticity with the ARIMA time series process to construct a nonlinear regression model. This model can more accurately reflect the nonlinear relationship between demand and price changes in dynamic pricing of FMCG. For example, during the promotion period of FMCG on e-commerce platforms, it can capture the nonlinear relationship of "initial sales volume surge followed by a slowdown in growth rate" when the price drops, providing support for optimizing the pricing rhythm of promotions, such as determining a "buy one, get one free" offer in the first week and gradually adjusting the discount rate subsequently.

3.2 The Cross-Model Comparison

In another piece of literature, Maria adopted a distinct research methodology: building a semi-parametric regression model grounded in support vector machine (SVM) technology and pairing this model with an analysis of daily scanner sales data from a Spanish supermarket dating back to 1999 [3]. Concurrently, Maria conducted a comparative assessment of the fitting performance of traditional parametric models—including the linear additive and multiplicative models built on ordinary least squares (OLS)—against that of standard semi-parametric models [3]. From these analyses, a series of key conclusions emerged: the specific shape of the discount effect curve (with a discount threshold falling between 4%–8% and a saturation level ranging from 14%–20%); the sales-accelerating impact of weekend promotions, which proves more pronounced for high-priced brands; and the asymmetry inherent in cross-brand price effects, where promotions for premium-priced brands exert a stronger pull on the sales of low-cost alternatives, while cross-effects between brands at the same price tier tend to be more conspicuous.

A cross-model comparison of these regression frameworks yields two core takeaways:

1) In terms of fitting accuracy, the nonlinear SVM-semi-parametric regression (SVM-SR) model outperforms all others overall. By blending the flexibility of its non-parametric component with the interpretability of its parametric segment, it delivers more stable test set error rates and can capture nonlinear patterns that elude conventional linear or multiple regression approaches.

2) In terms of scenario applicability, linear regression is only suited for the preliminary exploratory phase of new product pricing. Multiple regression works well for routine multi-factor pricing exercises, whereas nonlinear regression emerges as the core tool for pricing new FMCG during promotional campaigns and for multi-brand bundle pricing—particularly in scenarios that demand precise control over discount ranges and promotion timelines.

3.3 The Semi-Logarithmic Regression Model

In addition to the above-mentioned studies, Robert established a semi-logarithmic regression model including variables such as relative prices, promotional discounts, advertising, and display, using weekly scan data of 12 "chain-store-brand" combinations of four mainstream bath tissue brands from three chain supermarkets as samples, and reached the following key conclusions [4]:

(1) In terms of fitting accuracy: Linear regression can meet the basic correlation verification, but the reliability of the coefficients is insufficient; multiple regression improves the explanatory power of multiple factors, but it is limited by the linear assumption; the nonlinear shrinkage estimation model takes into account both fitting accuracy and the rationality of the coefficients, and has lower errors in long-term prediction, making it the optimal choice for complex pricing scenarios.

(2) At the level of applicable scenarios: Linear regression is suitable for the initial exploration of new product pricing, multiple regression is adapted to regular multi-factor pricing, and the non-linear shrinkage estimation model is the core tool for cross-channel, cross-brand, and multi-promotion strategies of FMCG new products, especially suitable for the refined pricing needs that require balancing "common patterns" and "individual differences".

4. Key Finding Synthesis

Regression analysis effectively enhances the precision of FMCG pricing. Multiple linear regression yields higher accuracy than conventional methods. Nonlinear and semi-parametric models better capture promotion effects and product substitution. Valid variables include price elasticity, cannibalization, weather and festivals. Machine learning strengthens forecasting and profit optimization. Implementation barriers involve channel conflicts and insufficient dynamic adjustment.

5. Discussion

5.1 The Obstacles to The Implementation of Regression Models in Fast-Moving Consumer Pricing Decisions

Even though regression analysis models have been widely applied in the field of new fast-moving consumer products, the obstacles to their implementation in fast-moving consumer products pricing decisions still cannot be ignored. In a study conducted in China, Si Y. (2023) [5] took the BOPS channel as an example and compared the benefits that the brand can obtain when different denominations of coupons are issued. The study found that when the coupon denomination is greater than the threshold $(1-\theta)v$, the brand is difficult to obtain benefits, which obviously has an impact on the brand's pricing decisions. The author also proposed a solution, that is, the brand and the retailer build a BOPS cooperation. The brand is the leader of the omni-channel system, responsible for the online channel, and relies on the brand's advantages to drive the retailer to jointly build the BOPS channel and provide products for BOPS channel consumers. The retailer is the distributor of the brand, mainly responsible for the offline channel, and at the same time provides store experience, pick-up and other services for BOPS channel consumers.

In another piece of literature, Liu reviewed the successful strategies in the pricing decision of new products. Liu thus summarized three major strategies: skimming pricing strategy, penetration pricing strategy and satisfactory pricing strategy. In addition, Liu also summarized the common problems in the pricing of the FMCG industry [6]:

(1) The pricing results of the regression model are disconnected from the profit distribution of the channel, failing to balance the interests of multiple levels, which leads to the dealers' resistance, weak terminal promotion, and the obstruction of the strategy implementation.

(2) The implementation of the price promotion strategy is arbitrary, without formulating the plan based on the price elasticity analysis of the regression model, resulting in frequent short-term price cuts and chaotic price levels, weakening the long-term effectiveness of the model pricing.

(3) There is a lack of a review and correction mechanism after implementation. The pricing suggestions of the regression model are not dynamically adjusted in combination with market feedback, causing the model results to be disconnected from the actual sales performance.

5.2 The Potential of Machine Learning in Pricing New Fast-Moving Consumer Products

In 2023 research on the application of machine learning in product pricing, Lee first precisely calculated the core variables that affect product sales, such as the price sensitivity of FMCG and cross-product price elasticity, through exploratory data analysis (EDA) combined with statistical and time series analysis [7]. Subsequently, to construct a demand forecasting model that reflects the product substitution effect within the category, the authors compared simple averaging, naive forecasting, and other benchmark models with traditional models such as time series and multiple linear regression. Ultimately, they selected a machine learning ensemble model, combined with hyperparameter tuning and rolling forecast methods, to complete the screening of the optimal demand forecasting models for each individual product. Based on the demand forecast, the authors used the results as input and embedded them into the price optimization model with business rules constraints in the FMCG industry. They also introduced a constraint programming (CP) improved algorithm to solve the efficiency problem caused by the massive price combinations resulting from the substitution effect. From the research findings, the machine learning ensemble model significantly outperformed traditional statistical models and various benchmark methods in capturing the product substitution effect and ensuring the accuracy of demand forecasting. The price optimization model that incorporates the substitution effect can output the optimal price combination that maximizes weekly profits, and the improved algorithm can reduce the number of price combinations and execution time by more than ten thousand times compared to the brute force enumeration method. The overall decision-making framework can assist FMCG retailers in precise pricing and inventory control, as well as uncover consumer behavior patterns, bringing direct financial benefits to the enterprise. Applying machine learning to regression models can directly use the regression demand forecast results that incorporate the substitution effect as the input for the price optimization model and combine the business rules of the FMCG industry (such as inventory upper limit and gross margin lower limit) to construct constraints, outputting the globally optimal price combination that maximizes weekly profits.

6. Conclusion

FMCG new products are an important part of modern commodities. Their characteristics lie in instability and strong time sensitivity. Therefore, being good at applying various statistical models to the pricing of FMCG new products can effectively help brand owners create greater profits. Although the practical application of regression analysis in FMCG new products still faces some obstacles at present, there is every reason to believe that with the advancement of technology, regression analysis can greatly promote the development of modern business.

References

[1] Dachyar, M., Taurina, Z. & F.F., (2021). Disclosing fast moving consumer goods demand forecasting predictor using multi linear regression. EASR, 7.

[2] Zhu, Z., Wu, B. & Wang, T., (2024). Research on Vegetable Cost Pricing and Sales Forecast Based on PSO Optimization and Forecast Model. Academic Journal of Business & Management, 6(5), pp.141–146.

[3] Maria, S. & Luis, G., (2020). Evaluating temporary retail price discounts using semiparametric regression. Journal of Retailing and Consumer Services, 55, pp.101890–101905.

[4] Robert, L. & Jennifer, K., (2012). Shrinkage estimation of price and promotional elasticities: seemingly unrelated equations. Journal of Marketing Research, 3, pp.289–302.

[5] Si, Y., Meng, Q. & Yang, W., (2023). Omnichannel Pricing Strategy Based on Consumers' Channel Preference and Electronic Coupon Delivery. Journal of Systems & Management, 6, pp.1021–1032.

[6] Liu, L., (2023). New product pricing strategies and post-launch success review in the fast-moving consumer goods industry. Times of Economy & Trade, 11, pp.52–54.

[7] Lee, K.H. & Abdollahian, M., (2023). USING MACHINE LEARNING APPROACHES TO DEVELOP PRICE OPTIMISATION AND DEMAND PREDICTION MODELS. Mathematics, 11(11), p.2502.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access