

A Study on the Nonlinear Relationship Between E-commerce Discount and Sales Profit—An Empirical Analysis Based on the Superstore Dataset

Bohan Ren

*School of Software Engineering, Shenyang University of Chemical Technology, Shenyang, China
1635862398@qq.com*

Abstract. In the context of e-commerce promotion normalization, discount has become a common marketing tool. However, excessive discounts may not bring profit growth and may even lead to losses. This paper investigates the nonlinear relationship between discount depth and sales profit based on the Superstore dataset (9,994 orders, 9,794 after cleaning). Using descriptive statistics, OLS regression with quadratic term, and subsample analysis by product category, the study examines whether discount has a U-shaped or inverted U-shaped effect on profit, and whether the relationship differs between absolute profit and profit margin, as well as across product categories (Furniture, Office Supplies, Technology). The results show that discount has a significant U-shaped relationship with absolute profit (quadratic coefficient positive, $p < 0.001$) with a turning point at about 79.5%, while discount has an inverted U-shaped relationship with profit margin (quadratic coefficient negative, $p < 0.001$). Heterogeneity across categories is also found: Technology products show a monotonically decreasing trend, while other categories show no significant nonlinearity. This study provides empirical evidence for pricing strategy optimization and enriches the price-quality heuristic and reference price theories.

Keywords: E-commerce discount, sales profit, nonlinear relationship, product category heterogeneity

1. Introduction

With the rapid development of e-commerce, platform promotional activities have become increasingly frequent. Major sales events such as "Double Eleven" and "618" have become the norm, and discount intensity has been increasing year by year. However, in practice, the phenomenon of "deeper discounts, lower profits" or even losses often occurs. For example, in the Superstore dataset, an order with a 45% discount generated a profit of -383 US dollars. Traditional marketing theories typically assume a linear relationship between discount and profit—discounts increase sales volume, which in turn increases profits [1]. However, recent studies have found that discounts have diminishing returns or even negative effects, and excessive discounts may damage consumers' quality perceptions and brand equity [2]. Nevertheless, systematic empirical tests of the nonlinear relationship between discount and profit based on large-scale real transaction data remain

insufficient, particularly lacking comparative studies on the differences in relationship patterns across different profit metrics (absolute profit vs. profit margin) and different product categories.

Based on this, this study takes the publicly available Sample Superstore dataset from the Kaggle platform (containing 9,994 order records, spanning 2014-2017) as the analytical object, employing descriptive statistics, OLS linear regression, and quadratic term testing to investigate the following three questions: Does a nonlinear relationship (U-shaped or inverted U-shaped) exist between discount intensity and order profit? Is the effect of discount on absolute profit consistent with its effect on profit margin? Does this relationship differ across product categories (Furniture, Office Supplies, Technology)? This study aims to provide empirical references for e-commerce enterprises in formulating refined discount strategies, and to supplement the Price-Quality Heuristic and reference price theories with evidence based on real transaction data.

2. Literature review and research hypotheses

2.1. Theoretical foundations

2.1.1. Price-Quality Heuristic theory

The Price-Quality Heuristic theory suggests that consumers tend to positively associate price with product quality. Excessively low selling prices or overly high discounts may raise consumers' doubts about product quality, thereby reducing purchase intentions and brand evaluations [1]. Accordingly, higher discounts are not always better; beyond a certain threshold, negative effects may emerge.

2.1.2. Reference Price Theory

Reference Price Theory holds that consumers use prices stored in memory as references when making purchases, and the magnitude of a discount relative to the reference price affects transaction utility [2]. Excessive discounts lower consumers' expectations of future prices, causing reference prices to decline, which makes subsequent sales at normal prices more difficult and damages long-term profitability.

2.2. Related studies on the relationship between discount and profit

Existing studies have mostly focused on the effects of discount on consumer perception, purchase intention, or sales volume [3], while some have begun to explore the nonlinear relationship between discount and profit. Grewal et al. found through experiments that moderate discounts can enhance perceived value, but excessive discounts reduce quality perception [1]. Huang and Yu, based on e-commerce data, identified a threshold effect between discount depth and profit, with profits declining after exceeding the optimal discount rate [4]. However, most of these studies use small-sample experimental data or only test linear models, and thus do not adequately characterize nonlinear patterns.

2.3. Limitations of existing research

Existing research has the following limitations. First, there is no consensus on the shape of the nonlinear relationship (U-shaped or inverted U-shaped) between discount and profit. Second, few studies directly compare the differential effects of discount on "order absolute profit" versus "profit

margin." Third, empirical exploration of differences in nonlinear relationships across product categories remains insufficient.

2.4. Research hypotheses

Based on the theoretical review and identified research gaps, this study proposes the following hypotheses:

H1: There is a significant nonlinear relationship between discount intensity and order profit.

H2: The effect of discount on absolute profit differs from its effect on profit margin. Specifically, absolute profit may exhibit a U-shape due to sales volume recovery, while profit margin may exhibit an inverted U-shape due to compression of per-unit profitability.

H3: The nonlinear relationship between discount and profit exhibits heterogeneity across product categories (Furniture, Office Supplies, Technology).

3. Research design

3.1. Data source and sample description

The data are obtained from the Sample Superstore dataset [5] on Kaggle (Tableau public data), containing 9,994 order records from 2014 to 2017. After removing records with missing profit or discount values, excluding discount rates outside [0,1], and trimming profit values beyond the 1st and 99th percentiles, the final sample comprises 9,794 observations.

3.2. Variable definition and measurement

The dependent variables in this study are Profit and Profit_margin. Profit is taken directly from the Profit field in the dataset, measured in US dollars. Profit_margin is calculated as Profit/Sales, reflecting the profitability efficiency per unit of sales revenue. The independent variable is Discount, taken from the Discount field in the dataset, ranging from 0 to 1. To capture the potential nonlinear relationship, a squared term of Discount, denoted as Discount_sq and calculated as Discount × Discount, is also included in the regression models. In an ideal specification, the following control variables should be incorporated: Sales (sales revenue), Quantity (purchase quantity), Sub-category (product sub-category), Region (geographic region), Segment (customer type), Ship Mode (shipping method), and Order Year (order year). These variables are not included in the current model due to the preliminary nature of this study, but future research will gradually incorporate them to test the robustness of the results.

3.3. Theoretical framework and research approach

This study first tests the linear and nonlinear effects of discount on order absolute profit. It then compares the differential effects of discount on absolute profit versus profit margin. Finally, it conducts heterogeneity analysis by product category (Furniture, Office Supplies, Technology)

3.4. Model specification

Linear Regression Model

$$\text{Profit} = \alpha + \beta_1 \cdot \text{Discount} + \varepsilon \quad (1)$$

Quadratic Term Regression Model

$$\text{Profit} = \alpha + \beta_1 \cdot \text{Discount} + \beta_2 \cdot \text{Discount}^2 + \varepsilon \quad (2)$$

$$\text{Profit_margin} = \alpha + \beta_1 \cdot \text{Discount} + \beta_2 \cdot \text{Discount}^2 + \varepsilon \quad (3)$$

Category Heterogeneity Model

Model (2) is estimated separately for three sub-samples: Furniture, Office Supplies, and Technology.

4. Empirical results

4.1. Descriptive statistical analysis

As shown in Table 1, the average discount rate is 15.3%, with discounts ranging from 0 to 80%. The average order profit is 24.18 US dollars, with a standard deviation of 73.15, indicating substantial variation in profitability. The average sales revenue is 182.44 US dollars, and the average purchase quantity is 3.75 units.

Table 1. Descriptive statistics

Variable	N	Mean	Std. Dev.	Min	25%	Median	75%	Max
Discount	9794	0.153	0.203	0.000	0.000	0.200	0.200	0.800
Profit	9794	24.175	73.153	-319.19	1.816	8.667	28.704	580.54
Sales	9794	182.44	348.76	0.444	16.77	51.49	195.14	8187.65
Quantity	9794	3.753	2.209	1.000	2.000	3.000	5.000	14.000

4.2. Correlation analysis

As shown in Table 2, the correlation coefficient between discount and profit is -0.345 ($p < 0.01$), indicating a moderate negative correlation. The correlation between sales and profit is 0.546 ($p < 0.01$), indicating a positive correlation. Scatter plots combined with LOWESS smoothing curves show a U-shaped trend in the discount-profit relationship (first decreasing, then increasing), providing preliminary support for the nonlinear hypothesis.

Table 2. Correlation matrix

Variable	Discount	Profit	Sales	Quantity
Discount	1	-0.345	-0.066	-0.001
Profit	-0.345	1	0.546	0.182
Sales	-0.066	0.546	1	0.247
Quantity	-0.001	0.182	0.247	1

4.3. Regression results

4.3.1. Linear regression results

The linear regression results for Model (1) are shown in Table 3. The coefficient for Discount is -124.58 ($p < 0.001$), with $R^2 = 0.119$, indicating that the linear model fits poorly and the negative effect of discount on profit is averaged out, failing to capture nonlinear characteristics.

Table 3. Linear regression results (dependent variable: profit)

Variable	Coefficient	Std. Error	t-value	p-value	95% CI
Constant	43.22	0.87	49.73	0.000	[41.52, 44.93]
Discount	-124.58	3.42	-36.39	0.000	[-131.29, -117.87]

$R^2 = 0.119$, Adjusted $R^2 = 0.119$, $F = 1324$, $p < 0.001$

4.3.2. Nonlinear regression results

After adding the quadratic term, the results for Model (2) are shown in Table 4. The coefficient for the linear term of Discount is -211.78 ($p < 0.001$), and the coefficient for the quadratic term is 133.11 ($p < 0.001$). The quadratic term is significantly positive, indicating a U-shaped fitted curve. The turning point is calculated as: $-\beta_1/(2\beta_2) = -(-211.78)/(2 \times 133.11) \approx 0.795$, meaning that profit reaches its minimum at approximately 79.5% discount. It should be noted that this turning point is close to the sample maximum discount of 80%, and the limited number of high-discount observations warrants caution in extrapolation. This result confirms the presence of a significant nonlinear relationship between discount and profit, thus supporting H1.

Table 4. Nonlinear regression results (dependent variable: profit)

Variable	Coefficient	Std. Error	t-value	p-value	95% CI
Constant	47.98	0.98	48.89	0.000	[46.06, 49.90]
Discount	-211.78	9.17	-23.09	0.000	[-229.76, -193.80]
Discount ²	133.11	13.00	10.24	0.000	[107.63, 158.59]

$R^2 = 0.128$, Adjusted $R^2 = 0.128$, $F = 721.4$, $p < 0.001$

4.4. Product category heterogeneity analysis

Model (2) was estimated separately for the three product category sub-samples, with results shown in Table 5. For Furniture and Office Supplies, the quadratic term coefficients are positive but not statistically significant ($p > 0.05$), indicating no significant nonlinear relationship for these two categories. For Technology products, the quadratic term coefficient is -265.84 ($p = 0.001$), which is significantly negative and theoretically suggests an inverted U-shape. However, the calculated turning point is approximately -29%, which falls outside the definition domain of discount rate [0,1]. Therefore, within the observable discount range, the profit of Technology products actually exhibits a monotonically decreasing trend and cannot be interpreted as having meaningful economic significance as an inverted U-shape. The finding that the nonlinear relationship varies across product categories supports H3.

Table 5. Category-wise regression results

Category	N	Discount Coef.	Discount ² Coef.	p-value (Quadratic)	Pattern
Furniture	2121	-355.83	280.75	0.083	Not significant
Office Supplies	6026	-150.69	31.42	0.512	Not significant
Technology	1647	-270.48	-265.84	0.001	Monotonic decreasing (negative turning point)

4.5. Robustness checks

4.5.1. Alternative dependent variable test

Replacing the dependent variable with profit margin (Profit/Sales) and re-estimating Model (2), the results are shown in Table 6. The quadratic term coefficient is -1.314 ($p < 0.001$), indicating an inverted U-shape. This result is opposite in direction to the main model (which shows a U-shape for absolute profit), suggesting that discount affects absolute profit and profit margin through fundamentally different mechanisms: discount may increase absolute profit by boosting order size, but it compresses per-unit profitability. The contrasting patterns observed between absolute profit and profit margin as dependent variables support H2, indicating that the choice of profit metric fundamentally alters the conclusion.

Table 6. Regression results with profit margin as dependent variable

Variable	Coefficient	Std. Error	t-value	p-value
Constant	0.343	0.002	155.19	0.000
Discount	-0.687	0.022	-31.31	0.000
Discount ²	-1.314	0.036	-36.34	0.000

$R^2 = 0.721$, Adjusted $R^2 = 0.721$, $F = 12260$, $p < 0.001$, $N = 9482$

4.5.2. Extreme discount sample exclusion test

After excluding observations with discount rates greater than 70% (resulting in 9,515 remaining observations), Model (2) was re-estimated. The results are shown in Table 7. The quadratic term coefficient remains significantly positive at 115.55 ($p < 0.001$), indicating that the U-shaped relationship is robust to the exclusion of extreme discount observations.

Table 7. Regression results after excluding extreme discounts (>70%)

Variable	Coefficient	Std. Error	t-value	p-value
Constant	47.81	0.99	48.16	0.000

Table 7. (continued)

Discount	-205.42	9.85	-20.85	0.000
Discount ²	115.55	16.14	7.16	0.000

$R^2 = 0.115$, Adjusted $R^2 = 0.115$, $F = 619.8$, $p < 0.001$, $N = 9515$

5. Discussion

5.1. Economic interpretation of the U-shaped relationship

The main model reveals a U-shaped relationship between discount and absolute profit, meaning that profit first decreases and then increases as the discount rate rises. A possible explanation is that in the low-discount range (0-20%), the discount intensity is insufficient to significantly stimulate sales volume, but directly compresses profit margins, leading to profit decline. In the high-discount range (>50%), steep discounts may trigger clearance-style purchasing, substantially increasing order size (quantity and sales revenue) and thereby driving total profit recovery. This interpretation requires further data verification.

5.2. Analysis of differential effects on profit vs. profit margin

When the dependent variable is switched from absolute profit to profit margin, the curve shape changes from U-shaped to inverted U-shaped. This indicates that discount affects total order profit and per-unit profitability through different mechanisms: while high discounts may stimulate sales volume and thus increase total profit, the profitability per unit of sales revenue is severely compressed. Firms pursuing total profit maximization may accept a U-shaped strategy, whereas firms pursuing profitability efficiency need to strictly control discount levels.

5.3. Analysis of differential effects on profit vs. profit margin

When the dependent variable is switched from absolute profit to profit margin, the curve shape changes from U-shaped to inverted U-shaped. This indicates that discount affects total order profit and per-unit profitability through different mechanisms: while high discounts may stimulate sales volume and thus increase total profit, the profitability per unit of sales revenue is severely compressed. Firms pursuing total profit maximization may accept a U-shaped strategy, whereas firms pursuing profitability efficiency need to strictly control discount levels.

6. Conclusion

This study empirically examines the nonlinear relationship between discount and order profit using the Superstore retail dataset and OLS regression models. The following conclusions are drawn. First, there is a significant U-shaped relationship between discount and order absolute profit (quadratic term coefficient positive, $p < 0.001$), with the turning point occurring at approximately 79.5% discount. However, because this turning point is near the sample upper bound, the conclusion should be interpreted with caution. Second, when the dependent variable is replaced with profit margin, the quadratic term coefficient becomes significantly negative, indicating an inverted U-shape, which demonstrates that discount affects absolute profit and profit margin through fundamentally different patterns. Third, category heterogeneity analysis shows that the nonlinear relationship is not

significant for Furniture and Office Supplies, while Technology products exhibit a monotonically decreasing trend in profit within the discount range, which cannot be interpreted as a meaningful inverted U-shape.

The theoretical contribution of this study lies in revealing the fundamental impact of dependent variable selection on conclusions regarding the discount-profit relationship, and in validating product category heterogeneity. In terms of managerial implications, firms should monitor both absolute profit and profit margin simultaneously to avoid decision-making biases from single indicators; they should adopt differentiated discount strategies across product categories, and particularly for Technology products, discount intensity should be strictly controlled. Research limitations include: the data are relatively dated (2014-2017); the model has not fully incorporated control variables such as sales revenue, quantity, sub-category, region, and time; discount is not randomly assigned, raising self-selection concerns; the U-shaped turning point lacks sufficient reliability; and the negative turning point for Technology products has no practical economic meaning. Future research should incorporate more complete control variables, employ panel data or instrumental variable methods to mitigate endogeneity, collect more recent e-commerce promotional data, and introduce mechanism variables such as customer ratings, promotion duration, and inventory pressure to more deeply reveal the pathways through which discount affects profit.

References

- [1] Grewal, D., Monroe, K. B., & Krishnan, R. (1998). The effects of price-comparison advertising on buyers' perceptions of acquisition value, transaction value, and behavioral intentions. *Journal of Marketing*, 62(2), 46-59.
- [2] Kalyanaram, G., & Winer, R. S. (1995). Empirical generalizations from reference price research. *Marketing Science*, 14(3), G161-G169.
- [3] Zhang, S. (2023). *A study on the impact of discount promotion on brand value*. Tsinghua University. Beijing.
- [4] Huang, W., & Yu, J. (2019). The dark side of deep discounts: An empirical investigation. *Decision Support Systems*, 125, 113-122.
- [5] Tableau. (2025). Sample - Superstore Dataset. *Kaggle*. <https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>