Industrial Engineering and Innovation Management (2021) 4: 18-25
Clausius Scientific Press, Canada
DOI: 10.23977/ieim.2021.040204
ISSN 2522-6924

An empirical study on the influence mechanism of commodity sales on e-commerce platform

Yue Bao

Department of economics and management, East China Normal University, Putuo District, Shanghai

Keywords: influence mechanism, Apple pencil, e-commerce, factors

Abstract: On the basis of sorting out the influencing factors of online commodity sales studied by scholars, combined with the Apple pencil sales and related factor values selected by Taobao, the best regression model is obtained through regression test, and the important influencing factors on Apple pencil sales are analyzed by using the numerical value of the regression model. Finally, according to these influencing factors, appropriate suggestions are given.

1. Introduction

According to the data of China Internet Information Center, as of December 2016, the number of Chinese netizens reached 731 million, with a total of 42.99 million new netizens added throughout the year. The Internet penetration rate was 53.2%, 2.9 percentage points higher than that at the end of 2015. The e-commerce platform has developed rapidly in the past decade, and the e-commerce field has entered the Red Sea stage. With the increasingly fierce competition, the flow dividend is gradually reduced. Taobao, as the largest shopping platform in China, was established in May 2003 and is a leader in the e-commerce industry. However, with the rapid development of the e-commerce industry, e-commerce platforms such as JD and PDD have successively joined the competition. The commodity sales of some sellers on Taobao, especially small and medium-sized sellers, continue to be depressed, but the operating cost continues to rise. In the long run, they will face the risk of bankruptcy. Studying the influence mechanism of Taobao commodity sales is of great significance for merchants to choose and optimize the way of commodity display information. This paper selects the apple pencil on Taobao as the research object, and uses the signal theory to study the influencing factors of Taobao users' choice to buy Apple pencil on the platform.

2. Literature review

There are many factors affecting online consumers' shopping decision-making. At present, domestic scholars mainly focus on online reviews in terms of content, timeliness, type of reviewers, number of reviews, commodity prices, favorable comments and poor reviews.

Yan Qiang et al. (2013) did research on mobile phone products on Exun website, which proves that there is a clear correlation between online reviews and commodity sales. The results show that the more reviews there are and the higher the ranking of sales, the better the sales of products. [1] Li

Jian (2012) did study of Amazon mobile phone products, it is concluded that brand score, comment usefulness and product attention are positively correlated with product sales, while product price and commentator rank are less correlated with product sales. [2] Guo Gongxing (2013) studied the top 50 electric kettles sold by Taobao. The results show that the browsing volume, rating and other related information of the products will not significantly improve the sales of the products, while the number of ratings and rating will significantly improve the sales of the products. [4] Hao Yuanyuan (2010) and other scholars have found that positive emotions, higher mixing of positive and negative emotions and longer sentences in online film reviews have significant positive effects on the usefulness of film reviews. [5] Research by Chen et al. (2008) shows that the total number of online reviews has a significant impact on the sales of products. The more reviews there are, the more goods they sell. In the process of shopping decision-making, the quality of reviews will significantly affect consumers' shopping decision-making. According to a survey of 15 consumers who often have online shopping behavior, good and bad reviews of goods will have an important impact on their shopping decisions. [3] Hao Yuanyuan (2010) and other scholars believe that customer's online comment emotion has a significant impact on product sales. Consumers will refer to the comments of previously purchased goods when shopping, especially negative comment will more affect consumers' decision-making. Zhang Geng (2012) and other scholars believe that negative online comment has a significant negative impact on product online sales through empirical research on Taobao. [6]

Signal theory was first proposed by Spencer, an American economist. It is widely used to study the transaction scenario of asymmetric information between buyers and sellers, mainly including signal sender, signal receiver, feedback and other elements. The process of signal transmission is accompanied by the generation of cost, and the cost and benefit of the sender are negatively correlated. The commonly used signal theory research variables in online shopping include product sales, repurchase intention, online comments and so on.

Below are some important influence factors by scholars in table 1.

Table 1: Important influence factors of monthly sales

Influence factor	Main opinion	Scholar
Price	Obvious negative effect	Wu Desheng (2013) [7]
Applause rate of store	Positive effect	Guo song, Gao Baojun (2011) [8]
Praise number	Buyer's reputation and rating have a positive impact, while bad reviews have a negative effect	Cui Xiangmei, Huang Jinghua (2010) [9]
Historical evaluation number	Obvious positive effect	Zhao Zhanbo, Sun Luping, Su Meng (2013) [10]

This paper takes the sales volume of apple pencil in these stores as the dependent variable Y, price, applause rate of store, praise number, historical evaluation number as X1, X2, X3 and X4 respectively.

3. Empirical study

3.1 Correlation test

Firstly, we study the correlation between these independent variables. If the correlation between the two variables is too high, we may repeatedly calculate the influence of these two variables on the

dependent variable, thus affecting the accuracy of the model. Based on the correlation matrix below, it can be seen that both X_3 and X_4 are strongly correlated with Y, and the correlation between them is relatively high. However, they do not exceed our predetermined values. We will keep these variables for the time being and conduct VIF test to further determine their correlation.

Table 2: Correlation matrix of model

Matrix correlation				
1.00000	-0.31363	0.15685	0.87027	0.82518
-0.31363	1.00000	0.08615	-0.25297	-0.24253
0.15685	0.08615	1.00000	0.11666	0.14558
0.87027	-0.25297	0.11666	1.00000	0.83421
0.82518	-0.24253	0.14558	0.83421	1.00000

In VIF test, VIF values of several variables are 1.1, 1.0, 3.3 and 3.3 respectively, which are all less than our predetermined value of 5, so all variables can be retained. Below are the VIF values of several variables.

Table 3: VIF values of variables

Predictor	Coef	SE Coef	T	P	VIF
Constant	-3117	4501	-0.69	0.492	
X_1	-1.1396	0.7478	-1.52	0.133	1.1
X_2	3878	4661	0.85	0.399	1
X_3	0.13972	0.02634	5.3	0	3.3
X_4	2.501	0.8943	2.8	0.007	3.3

3.2 White test

After determining that all variables passed the correlation test, we conducted white test to check whether the data were homoscedastic or heteroskedastic. At first we list the residual and fits values (table 4), square the two columns, and use the data in another regression to get a new F value of 1.57.

Table 4: White test result

Analysis of variance					
Source	DF	SS	MS	F	P
Regression	2	383938	191969	1.57	0.216
Residual Error	57	6955427	122025		
Total	59	7339365			

We compare this F value with the F value we got from the F test. The result is shown in table below.

Table 4: F distribution value

F distribution with 4 DF in numerator and 56 DF in denominator	
P (X<= x)	x
0.95	2.53658

We can see that F value 1.57 is obviously less than 2.536. From this result, we can know that our data is homoscedastic. This is consistent with the values built into the data analysis software we use, so we don't have to recalculate the SE.

3.3 F test

After the white test, we will conduct an F test. F test result is shown below in table 5.

Table 5: F test result

Analysis of variance					
Source	DF	SS	MS	F	P
Regression	4	7339365	1834841	54.92	0.000
Residual Error	55	1837402	33407		
Total	59	9176767			

The F value we calculate is 54.92, which is much higher than 2.53. Therefore, we pass the F test and can carry out the next test.

3.4 T test

The next test is T test. The T values required for the T test are already visible in the previous table, as X_1 — 1.52, X_2 — 0.85, X_3 — 5.3, and X_4 — 2.8, respectively. And since our data is homoscedastic variance, we don't need to recalculate. With 95% of the significance level, the value of t-actual is 1.96. After comparing the values of these t's, only X_3 and X_4 pass the significant value. And then we're going to test the rest of these variables again.

3.5 A New Linear Model

Then, we should get matrix correlation of the new linear model after tests before. The correlation result is shown below (table 6). The VIF test result is shown in table 7.

Table 6: Matrix correlation of the new linear model

Matrix correlation		
1.00000	0.87027	0.82518
0.87027	1.00000	0.83421
0.82518	0.83421	1.00000

Table 7: VIF test of the new linear model

Predictor	Coef	SE Coef	T	P	VIF
Constant	4.41	28.9	0.15	0.879	
X_3	0.14336	0.0264	5.43	0	3.3
X_4	2.6491	0.8945	2.96	0.004	3.3

The new model also apparently passed the correlation test. After that, we'll do a white test on it (table 8). The newly calculated value of F is shown below in table 9, which is definitely less than the

F statistic.

Table 8: White test of the new linear model

Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	2	300663	150332	1.23	0.229
Residual Error	57	6946438	121867		
Total	59	7247101			

Table 9: New F distribution value

F distribution with 2 DF in numerator and 58 DF in denominator	
P (X<= x)	x
0.95	3.15593

After verification, the remaining variables are determined to be homoscedastic after passing white test, so we can directly use the t values in the above table for comparison. The T values of X_3 and X_4 are 5.43 and 2.96. They are also greater than t actual 1.96, so the remaining two variables can pass the t test.

And then we have a preliminary model:

$Y = 4.4 + 0.143 X_3 + 2.65 X_4$.

Where X_3 means historical evaluation number and X_4 means product likes.

And its SER, R-Square, adjusted R-Square value respectively are

S = 183.994; R-Sq = 79.0%; R-Sq(adj) = 78.2%.

3.6 Nonlinear model analysis

After completing the analysis of linear regression, we select the independent variable of X with stronger influence to conduct the nonlinear regression analysis according to the current model. From the correlation matrix obtained in the second time above, the correlation between X_3 and X_4 and Y is 0.87027 and 0.82518, respectively. Therefore, X_3 is selected as the main independent variable in the following analysis.

After a simple regression analysis of nonlinear models, we can get two models, they are quadratic model and cubic model.

Model 1:

$Y = 44.58 + 0.1792 X_3 + 0.000004 X_3**2$

S = 196.086 R-Sq = 76.1% R-Sq(adj) = 75.3%

Model 2:

$Y = 61.89 + 0.09531 X_3 + 0.000040 X_3**2 - 0.000000 X_3**3$

S = 195.652 R-Sq = 76.6% R-Sq(adj) = 75.4%

We can see that no matter the quadratic model nor cubic model, their adjusted R Square's values are not higher than the linear model's we made before. Even after the detailed analysis of the model, the model extended by the new model is established and can be used, its performance will not be better than the linear model obtained above. So we will no longer do a concrete analysis of several nonlinear model.

3.7 Logarithmic Model

Finally, we redo a few logarithmic model, hoping to find a better model. Logarithmic model basically has three kinds, respectively is linear - log model, the log - linear model and the log - log model, we made three models in this performance (59 cases used, 1 cases contain missing values), the three models' results are shown below in three tables (table 10, table 11, table 12).

Model 1:

$Y = - 470 + 127 \ln X_3$

S = 321.939 R-Sq = 35.4% R-Sq(adj) = 34.2%

Table 10: Model 1 result

Predictor	Coef	SE	Coef	T	P
Constant	-470.2	132.8	3.54	0.001	
$\ln X_3$	126.68	22.68	5.59	0.000	

Model 2:

$\ln Y = 3.77 + 0.000615 X_3$

S = 1.14165 R-Sq = 44.4% R-Sq(adj) = 43.5%

Table 11: Model 2 result

Predictor	Coef	SE	Coef	T	P
Constant	3.7733	0.1710	22.06	0.000	
$\ln X_3$	0.00061494	0.00009032	6.81	0.000	

Model 3(59 cases used, 1 cases contain missing values):

$\ln Y = 0.935 + 0.619 \ln X_3$

S = 1.01331 R-Sq = 56.9% R-Sq(adj) = 56.1%

Table 12: Model 3 result

Predictor	Coef	SE	Coef	T	P
Constant	0.9354	0.4179	2.24	0.029	
$\ln X_3$	0.61878	0.07138	8.67	0.000	

When doing the logarithmic model analysis, there is a data in X_3 is 0, so part of the model analysis using only the remaining 59 set of data. But just missing a set of data will not cause too much error in the results of data analysis. By the analysis of the results before we can see the new model is far from good comparing to the nonlinear model, not even better than a nonlinear model can show these data, so we will not furtherly analyze these few logarithmic model for finding a better model.

After comparing multiple models, we can find that the best performance model is the linear model:

$Y = 4.4 + 0.143 X_3 + 2.65 X_4$

The standard errors of each coefficient in the model are SE (b_0) =28.9, SE (b_1) =0.0264, SE(b_2) = 0.8945

With the SE of these coefficients, we can also compute the confidence interval for b_1 and b_2 at the 95%significance level

B_1's confidence interval:

(b_1-SE(b_1)*1.96, b_1+SE(b_1)*1.96)

Substitute the Numbers we can get
(0.143-0.0264*1.96, 0.143+0.0264*1.96)
After calculating it we can get
(0.091256, 0.194744)
In the same way we can get the b_2's confidence interval
(0.89678, 4.40322)

4. Conclusion

Through our final model, we can know that when people buy the product apple pencil, they will refer more to the historical evaluation of the product and the number of times people praise the product. Every time a product has seven more historical comments, there may be one more person willing to buy the product, and every time the product gets two more likes, there will be five more increase the desire to buy it from this shop.

The research results of this paper show that the number of commodity comments and praise have a significant positive impact on the commodity sales of e-commerce platform. For Taobao sellers, consumers can be encouraged to publish positive comments, enrich the words and pictures in positive comments as much as possible, and consumers' experience can be more convincing. Taobao sellers can also optimize the selection of goods. High quality goods can bring more praise. Merchants can establish a good reputation and attract more consumers to buy. For businesses selling similar goods, if they can accumulate more praise, they can still obtain higher sales even if the price is slightly higher than other sellers. Businesses can strive to obtain more high-quality comments as soon as possible and find a breakthrough in the fierce competition.

References

[1] Yan Qiang, Meng Yue. Influencing factors of perceived usefulness of online reviews—— An Empirical Study Based on online film reviews [J]. China management science, 2013 (S1): 126-131

[2] Li Jian. Research on the impact of online commodity reviews on product sales [J]. Modern intelligence, 2012 (1): 164-167

[3] Yubo Chen, Jinhong Xie. (2008). Online Consumer Review: Mouth as a New Element of Marketing Communication Mix, Science, 54: 477-491.

[4] Guo Gongxing. Analysis on Influencing Factors of consumers' online shopping decision——An Empirical Study Based on online sales information of electric kettle [J]. Consumer economy, 2013 (4): 52-57

[5] Hao Yuanyuan, ye Qiang, Li Yijun. Study on Influencing Factors of online review usefulness based on film review data [J]. Journal of management science, 2010, 13 (8): 78-89

[6] Zhang Geng, Guo Ning. The impact of negative online comments on product sales: An Empirical Study Based on Taobao [J]. Consumer economy, 2012 (6): 86-89

[7] Wu Desheng, Ren Xingyao. Research on the effectiveness of online auction trading mechanism -- Evidence from Taobao panel data [J]. Nankai management review, 2013, 16 (1): 122-137

[8] Guo song, Gao Baojun. Empirical analysis on Influencing Factors of store seller credit on Taobao [J]. Luojia management review, 2011 (02): 162-169

[9] Cui Xiangmei, Huang Jinghua. Empirical Study on the impact of credit evaluation system and related factors on buy it now online trading [J]. Management, 2010, 7 (01): 50-63

[10] Zhao Zhanbo, Sun Luping, Su Meng. Comparative study on Influencing Factors of product browsing and sales in C2C [J]. Management science, 2013, (1): 58-67

Appendix (data)

Store name	Monthly sales	Price	Store praise rate	Historical evaluation number	Product likes
Top digital	337	628	0.9984	591	66
Ping'an digital technology	2270	610	0.9984	10315	282
Taopu tide Technology	891	608	0.9979	187	45
Leap digital	1355	618	0.9962	4067	145
Qichuang Technology	496	610	0.9973	1834	68
I Aigo digital	845	626	0.9995	1135	46
Wanglong digital communication	653	610	0.9992	1681	111
Yousheng mobile tablet	27	650	0.9979	85	15
Fast e digital network	860	608	0.983	4791	72
Hangzhou messenger digital	187	610	0.995	2578	37
Genuine American goods purchased from overseas	164	610	0.9972	58	15
Rocky digital mall	520	625	0.9998	2857	80
My embarrassing digital	190	630	0.9919	1359	29
Pacific physical store	45	625	0.9968	207	9
Grape digital intelligence	54	598	0.9962	2015	42
Mike McCaw	158	625	0.9955	749	19
Fast e mobile network	688	626	0.9996	2993	42
BOJIANG digital store	113	628	0.9726	220	9
Mike regiment	115	630	0.9981	1135	10
Boat of wealth	131	610	0.9994	1728	12
Bentley discount store	94	608	0.9806	12	10
Shurui Technology	47	606	0.9893	159	28
Hongyue liangpin	191	700	0.998	1322	22
machome	55	658	0.9972	429	20
Yantao apple	87	648	1	277	6
Pingjia	21	636	0.9943	58	3
a little	272	640	0.9959	2310	31
Imagination technology digital	125	580	0.9988	19	11
Apple zone	962	608	0.9995	2796	70
Shenguan Digital Museum	21	658	0.998	15	3
Thursday digital store	17	709	0.9939	229	11
Jinhua communication	107	600	0.9991	231	29
Shanghai Kai Technology Co., Ltd	35	639	0.9967	173	14
Boar digital technology	25	668	0.998	136	8
Guangzhou Xinyan Technology Co., Ltd	9	670	0.9964	11	7
TENGWEI Huanyu Technology	136	620	0.9916	1431	176
Erniu Digital City	2	680	0.9963	43	1
Super digital Shenzhen store	24	660	0.9917	317	8
Si Niu electronic digital	50	648	0.9972	437	21
Yonghui Xinde apple	246	680	0.9976	794	52
Icing studio	24	555	0.9945	39	12
Monkey	205	619.8	0.9912	207	28
Brother digital	17	599	0.9988	5	1
Post zero studio	96	650	0.9827	88	9
Qunxing digital mall	45	616	0.9839	0	5
Changhong Digital Museum	48	640	0.9936	59	17
Zhejiang Apple source	11	640	0.9896	151	7
Xinlong Diantong	34	670	0.9975	118	3
Frank digital	8	699	0.9901	15	0
Xuanhe Technology	25	680	0.994	16	6
Mood digital mall	23	668	0.9963	56	23
Jingpin	41	650	0.9925	178	13
Zhonggangtong	35	700	0.9946	68	11
Hong Kong gateway digital	60	658	0.9942	1317	19
Yifan digital pioneer	23	688	0.9966	122	10
Dongsheng and fruit powder	25	680	1	38	7
Huaxiaxing digital technology	7	650	0.9999	41	3
Toyo bastard	36	665	0.9949	926	12
Zhuo Chen	6	680	0.9963	15	7
Top digital	433	699	1	2386	123