Research on Pricing Model Based on Internet "Making Money by Taking Photos"

Xiaolan Lu[1,*], Kexin Zhang[2]

[1]School of Business, Jianghan University, Wuhan, China

[2]School of Statistics and Mathematics, Guangdong University of Finance & Economics, Guangzhou, China

[*]luxiaolan@jhun.edu.cn

Keywords: Making Money by Taking Photos; Influencing Factors; Pricing Law; Logistic Model

Abstract: With the development of Internet technology, "taking photos to make money" work has become a popular Internet self-help mode, and the rationality of task pricing and task completion is particularly important. Based on the data of Question B of the 2017 National Mathematical Modeling Contest for College Students, this paper established a task pricing linear model and a task completion Logistic regression model, and studied the law of task pricing and the influencing factors of task completion. The results show that the average distance of potential members to the task site and the average salary of potential members in the region are the main factors affecting the pricing. The main reasons why the task cannot be completed are the low price of the task, the low reputation of the members and the long distance between the potential members to the task point.

1. Introduction

With the rapid development of mobile Internet, crowdsourcing business develops rapidly. In this context, the specific work is no longer limited to a part of the people, but for the voluntary participation of the public, to maximize the power of the public to improve the efficiency of some traditional work. "Making money by taking photos" is a crowdsourcing method for the public. Users download the APP, register as members of the APP, and then receive the task of taking photos on the APP, complete the task and earn commission. The task pricing of "making money by taking photos" is the core, and the rationality of the pricing directly affects the completion of the task.

Question B of the 2017 National Mathematical Contest in Modeling mainly focuses on the pricing rules of the task of "making money by taking photos". Appendix 1 shows the task data for an ended project, including the location, pricing, and completion of each task ("1" means completed, and "0" means not completed). Appendix 2 provides member information data, including the location and reputation value of each member. Based on the data, this paper established task pricing law model and task completion Logistic model respectively, and further improved the model on this basis to analyze the main influencing factors of pricing and completion.

2. Variables and data

The data are from Appendix 1 and Appendix 2 of The 2017 National College Students Mathematical Modeling Contest.

2.1 Influencing factors of task pricing

Task density (*NumA*).Take each task point as the center of the circle to calculate the number of tasks within a radius of 20km. The greater the task density around a task point, the greater the competitive pressure of the task point, and the higher the pricing to attract members.

Member density (*NumB*).Take each task point as the center of the circle to calculate the number of members within a radius of 50km. The setting of the 50km radius is based on the comparison between the average hourly wage in the region where the member lives and the remuneration for the task. After many attempts, it is more reasonable to calculate the density of members with a radius of

50km. The greater the density of members around a task point, the less competitive pressure the task point will be and the pricing will be low.

The average distance between potential members and the task point (*DistanceB*).With each task point as the center of the circle and 50km as the radius circle, members are the average distance between potential members and this task point. The average distance between potential members and task points is a supplement to the density of members. In the case of the same density of members, the longer the average distance between potential members and task points is, the higher the price will be in order to attract more bidders.

Firstly, the distance function in Stata16.0 software is used to calculate the distance among members and tasks and between tasks. Then calculate the task density and member density respectively with the radius mentioned above. In addition, other factors should be also considered.

Average salary in mission area (*IncomeA*). The higher the average annual wage in the region where the task is located, the higher the level of economic development in the region, and the higher the pricing of the corresponding task.

Average pay for potential members in their region (*IncomeB*). Take each task point as the center of the circle to calculate the average salary of each potential member in the region within a radius of 50km. The higher the average salary of potential members, the higher the economic development level of the region, the higher the expected profit value of members, and the higher the pricing of the corresponding task.

The data of these two variables are from EPS database. To reduce the possible heteroscedasticity, natural logarithms such as *lnNumA, lnNumB, lnDistanceB, lnIncomeA* and *lnIncomeB* are included in the regression model. The pricing price is obtained from Appendix 1, and the natural logarithm *lnprice* is taken as the dependent variable. The regression model is as follows:

$$\ln price = \alpha_0 + \alpha_1 \ln NumA + \alpha_2 \ln NumB + \alpha_3 \mathrm{lnDistanceB} + \alpha_4 \ln IncomeA + \alpha_5 \ln IncomeB + \mu_{ijt}$$

2.2 Influencing factors of task completion

Task marked price(*price*). From the perspective of members, the higher the task price, the more attractive it is to members, and the more likely it is to choose to complete the task. Conversely, the task is less likely to be completed.

Member credibility(*possible*).The correlation analysis of credit value, reservation start time and reservation task limit shows that there is a significant negative correlation between credit value and reservation start time, with a correlation coefficient of -0.217, that is, the higher the credibility, the earlier the reservation task start time. In addition, there was a high positive correlation between the credit value and the scheduled task quota, with a correlation coefficient of 0.648, that is, the higher the credit value, the more the scheduled task quota. Therefore, the starting time of the reservation task and the credit value variables are selected to construct a comprehensive index to measure the credibility. Assuming that the scheduled task end time is 10:00, the duration (minutes) from start to end is calculated. Make the variable possible=min*Q (where Q stands for the original value of member's credibility) to represent the comprehensive index of member's credibility. The smaller possible is, the greater the likelihood of unfinished tasks will be.

The average distance between potential members and the task point(*DistanceB*).The average distance between a potential member and the task will reduce the attraction to the member, which may cause the potential member to abandon the task.

Member distribution density(*NumB*).Too low member distribution density means that the number of members around the task site is small, so members may find more cost-effective tasks to complete, leading to the abandonment of the task.

The completion of the task in Appendix 1 is a dummy variable. There are only two cases: completed and incomplete, which are "1" and unfinished become "0". Therefore, Logistic regression model is needed to fit the completion of the task. In the Logistic regression model, *lnprice, lnpossible, lnDistanceB, lnNumB* were taken as independent variables, and task *finish* was the dependent variable.

3. Basic situation analysis

3.1 Distribution of task points

According to the latitude and longitude of the task points in Appendix 1, the distribution map of the task points was drawn with ArcGIS software, and it was found that the task points were distributed in Guangzhou, Foshan, Shenzhen, Dongguan and Qingyuan, and there were 835 task points in total. Among which Guangzhou had the most task points, accounting for 38.2%.At the same time, the distance between each task pair was calculated and the descriptive statistics were shown in the first row of Table 1. The average distance between tasks was 57.1km and the maximum distance was 188km.

Table 1. Statistical table of distance between task points

VarName	Mean	SD	Skewness	Kurtosis	Min	Max
Distance between tasks	57.104	33.769	0.441	2.292	0.005	187.987
Distance between members and tasks	60.748	53.220	9.859	190.684	0.010	1272.235

3.2 Distribution of members

According to the longitude and latitude positions of members in Appendix 2, ArcGIS software was used to draw the administrative region distribution map of members, and it was found that task points were distributed in 14 cities including Guangzhou, Foshan, Shenzhen, Dongguan, Qingyuan, Huizhou, Heyuan and Shanwei. The total number of members was 1,877, among which Guangzhou had the largest number of members, accounting for 35.8%.At the same time, the distance between members and tasks was calculated. As shown in the second row of Table 1, the average distance between members and tasks was 53.9km, and the maximum distance was 1272km.

4. The empirical analysis

4.1 Analysis of influencing factors of task pricing

Table 2. Regression results of pricing influencing factors

lnprice	M1	M2	M3
lnNumA	-0.00440	-0.0156**	-0.0216***
	(0.00673)	(0.00684)	(0.00691)
lnNumB	-0.0107	-0.0101	0.00886
	(0.00878)	(0.00860)	(0.00956)
lnDistanceB	0.109***	0.0524***	0.0453**
	(0.0168)	(0.0189)	(0.0188)
lnIncomeA		-0.0683***	-0.0859***
		(0.0113)	(0.0118)
lnIncomeB			0.294***
			(0.0675)
Constant	3.968***	4.974***	1.746**
	(0.0675)	(0.179)	(0.762)
Observations	835	835	835
Adjusted R-squared	0.187	0.221	0.237
F	65.11	60.13	52.94

Note :***,**, and * are significant at the levels of 1%,5%, and 10%, respectively

The regression results of the pricing rule are shown as Model1-Model3 in Table 2.Take model M3 as an example, the regression coefficient of task density (l*nNumA*) is -0.0216 and the symbol is negative, which is not consistent with the expectation. The regression coefficient of the member density (*lnNumB*) is 0.00886 and the symbol was positive, which was not consistent with the expectation. The regression coefficient of the average distance between potential members and the task point (*lnDistanceB*) was 0.0453, the sign was positive and significant at the statistical level of 5%, which was in line with the expectation and had a promoting effect. The regression coefficient of the average wage in the area where the task point is located (*lnIncomeA*) is -0.0859, and the sign is negative, which is not consistent with the expectation. The average wage regression coefficient in the region of potential members (*lnIncomeB*) is 0.294, the sign is positive and significant at the 1% statistical level, which is in line with expectations and has a promoting effect. Thus, the average distance between potential members and the task point and the average salary in the region of potential members are the main factors affecting the pricing.

4.2 Analysis of influencing factors for task completion

Logistic regression results are shown in Table 3.In Logistic model 1, *lnprice* and *lnpossible* were taken as independent variables. In this model, *lnprice* coefficient was positive and significant, indicating that the higher the task price was, the more likely it was to be completed. *Lnpossible* coefficient is positive and significant, indicating that the higher the credibility of potential members, the more likely the task will be completed. The model is used to judge the completion status, and the data of each factor is substituted into the model to get P value. If P>0.5, the task can be regarded as completed, and if P<0.5, it can be regarded as an unfinished task. The statistical table 4 describes the completion situation and the actual completion situation, and the correct rate is, and the prediction effect is fair.

Table 3. Logistic regression model of task completion probability

finish	Logistic M1			Logistic M2			Logistic M3			Logistic M4		
	Coef.	Std.Err.	P>z	Coef.	Std.Err.	P>z	Coef.	Std.Err.	P>z	Coef.	Std.Err.	P>z
lnprice	9.250	1.613	0.000	3.875	1.332	0.004						
lnpossible	0.553	0.267	0.039	1.483	0.280	0.000	2.224	0.327	0.000	2.071	0.304	0.000
lnDistanceB				-1.603	0.464	0.001	-2.107	0.673	0.002	-1.953	0.615	0.001
lnNumB				3.068	0.404	0.000	-1.127	0.770	0.143			
lnprice_hat							40.176	5.728	0.000	32.842	2.891	0.000
cons	-46.040	8.438	0.000	-35.064	6.121	0.000	-181.461	23.802	0.000	-153.101	13.455	0.000
Prob>chi2	0.0000			0.0000			0.0000			0.0000		
Pseudo R2	0.0470			0.1180			0.1544			0.1525		

Table 4. Accuracy of logistic regression model

Classified	Logistic M1			Logistic M2			Logistic M3			Logistic M4		
	finished	unfinished	Total	finished	unfinished	Total	finished	unfinished	Total	finished	unfinished	Total
>0.5	492	259	751	412	177	589	402	122	524	414	128	542
<0.5	30	54	84	110	136	246	120	191	311	108	185	293
Total	522	313	835	522	313	835	522	313	835	522	313	835
Accuracy judgment	65.39%			65.63%			71.02%			71.74%		

Logistic model 2 was constructed by adding factors such as potential member density (*lnNumB*) and the distance of potential member from task point (*lnDistanceB*). The value of Pseudo R2 in Logistic model 2 was 0.1180, which was better than that in Logistic model 1.The coefficient symbols of *lnprice* and *lnpossible* are consistent with those in Model 1.The sign of the *lnDistanceB* coefficient is negative and significant, indicating that the further the potential member is from the task point, the less likely the task will be completed. *LnNumB* coefficient is positive and significant, indicating that the higher the membership density, the higher the possibility of the task. In addition, the prediction accuracy of the model is $(412+136)/835=65.63\%$, and the prediction effect is fair.

Considering that the predicted Price obtained from the pricing law model contains other factor information except independent variable, it may be more effective to replace the actual Price with *price-hat* in the task completion probability model, so Logistic model 3 was constructed and its prediction effect was checked. The value of Pseudo R2 in Logistic model 3 was 0.1544, and the accuracy rate is $(402+191)/835=71.02\%$, which has a good prediction effect.

However, the coefficient of *lnNumB* in this model was not significant, and this variable could be considered to be deleted to obtain the following Logistic model 4.Each coefficient of Logistic model 4 is significant and its economic practical significance is consistent. The regression equation is:

$$\ln\frac{p}{1-p}=\text{-153.1+2.07lnpossible-1.953lnDistance}+32.842\ln price\_hat$$

The value of Pseudo R2 was 0.1525, and the accuracy rate is $(414+185)/835=71.74\%$. The prediction effect was good, and the model could be used for subsequent research and determination of new task pricing and possible task completion.

5. Conclusion

In this paper, the linear regression model of task pricing and the Logistic model of task completion probability were constructed respectively. The results show that the average distance of potential members to the task site and the average salary of potential members in the region are the main factors affecting the pricing. Task price, member credibility and the average distance potential members travel to the task site are main factors that determine whether the task can be completed.

The “making money by taking photos” task pricing model and task completion probability model in this paper can be extended to the scenarios of other crowdsourcing platforms, such as takeout APP and driving service, so as to effectively improve the operation performance of crowdsourcing platform.

References

[1] Shi Qingqing. 2020. Research on Task Pricing and Failure in "Making Money by Taking Photos" APP [J].*Technology and Markets*, Vol.27,No.1.

[2] Cui Yueqi, Liu Yanxuan, Liu Ying, et al. 2019. Nonlinear Fitting Analysis of Factors Influencing Photo-earning Pricing [J]. *China Collective Economy*, No.34.

[3] Xie Liying, Fang Limin, Chen Qiang. 2019.Task Publication and Pricing Method Based on Curve Fitting Crowdsourcing Platform [J]. *Journal of Guangdong Second Normal University*, Vol.39,No.5.

[4] Fang Yue, Xu Ganchen, Gao Zhen.2019.Task Pricing Model of "Making Money by Taking Photos" Based on Multiple Linear regression [J]. *Economic Research Guide*, No.24.

[5] Yang Feifei, Yuan Chenhui, Tang Shixing, et al.2008. Modeling and Calculation of the Task Pricing Problem of "Making money by Taking Photos" [J].*Value Engineering*, Vol.37, No.29.

[6] Feng Yunqiao, Yan Lingyu.2012.A New Method for Task Pricing on Crowdsourcing Platform [J]. *Industrial Engineering and Management*, Vol.23, No.4.