Data analysis of factors influencing consumers' payment price of traceable food

Xia Li¹,², Heliang Huang¹

¹School of Management, Fujian Agriculture and Forestry University, Fujian, China
²School of Economics and Trade, Fujian Jiangxia University, Fujian, China

Corresponding author and e-mail: Heliang Huang, hhh370@163.com

Abstract. Taking infant formula milk powder, an important food, as an example, this paper selects a total of 12 explanatory variables from three aspects: personal characteristics, family factors and psychological factors, takes consumers’ willingness to pay for food traceability labels as dependent variable Y, and adopts binary logistic regression model to understand and analyze the influencing factors of consumers' willingness to pay for food traceability labels with big data collection and analysis. The results show that consumers' age, education background, monthly family income, consumers' trust in traceable labels and total family population will significantly affect consumers' willingness to pay for food traceability labels; specifically, younger consumers are more willing to pay for food traceability labels, consumers with higher education are more inclined to pay for food traceability labels, consumers with higher income are more inclined to pay for food traceability labels, consumers with higher education are more inclined to pay for food traceability labels, and consumers with high trust in traceability labels are more inclined to pay for food traceability labels.

1. Introduction

This paper selects 12 explanatory variables from three aspects: personal characteristics, family factors and psychological factors through the literature review and analysis of the actual situation of milk powder consumption to design each explanatory variable and forecast the influence. Then, combined with the objective content of this study, this paper makes a research hypothesis. Based on the types and situations of independent variables and dependent variables, the binary logistic regression model was selected to investigate and analyze the influencing factors of consumers' willingness to pay for food traceability labels [1-3]. Then, we adopt the mode of field survey to ask the respondents to fill in the electronic questionnaire by scanning QR code with mobile phone, and 359 valid questionnaires have been collected as the data source of the study. We make a statistical analysis by means of binary logistic in SPSS statistical software, take consumers' willingness to pay for food traceability labels as the dependent variable Y, and 12 independent variables proposed in the research hypothesis as explanatory variables, so as to specifically analyze the influencing factors of consumers' willingness to pay for food traceability labels.
2. Research methods and theoretical basis
Consumers' willingness to pay for food traceability labels is taken as the dependent variable \( Y \), which is qualitative data, and there are only two results: "yes" and "no", that is, when the dependent variable is a binary variable, it is no longer applicable to the requirements of general linear regression model (it is required to be interval data rather than qualitative data) [4]. In addition, there are both continuous variables and categorical variables in the independent variables. In this case, the binary logistic regression model is suitable for analysis.

With binary logistic in SPSS statistical software, we take consumers' willingness to pay for food traceability labels as dependent variable \( Y \), and 12 independent variables proposed in the research hypothesis as explanatory variables, and use Forward: LR method to make an independent variables screening, which is a forward step-by-step screening strategy based on maximum likelihood estimation, so that some independent variables that may not be significant can be eliminated, and some more significant independent variables can be included in the regression model, making the results more effective and reliable. The specific results are as follows.

The payment mechanism of consumers for traceable food is shown in the figure below:

![Figure 1. Payment response mechanism of consumers to traceable food.](image)

3. Results of regression model
3.1. Coding categorical variables
For binary logistic regression model, first of all, we need to deal with the types of their variables, and distinguish between categorical variables and continuous variables. Among the explanatory variables, these independent variables "occupation, marriage and gender" are categorical variables, which need to be assigned in a quantitative way, i.e. changed into dummy variables (dumb variables) in the operation of SPSS. The result output is as follows:
Table 1. Categorical variables codings.

| Parameter coding | Frequency | (1)  | (2)  | (3)  | (4)  | (5)  | (6)  | (7)  |
|------------------|-----------|------|------|------|------|------|------|------|
| Occupation       | 1         | 18   | .000 | .000 | .000 | .000 | .000 | .000 |
|                  | 2         | 103  | .000 | 1.000| .000 | .000 | .000 | .000 |
|                  | 3         | 114  | .000 | .000 | 1.000| .000 | .000 | .000 |
|                  | 4         | 17   | .000 | .000 | .000 | 1.000| .000 | .000 |
|                  | 5         | 29   | .000 | .000 | .000 | .000 | 1.000| .000 |
|                  | 6         | 1    | .000 | .000 | .000 | .000 | .000 | 1.000|
|                  | 7         | 31   | .000 | .000 | .000 | .000 | .000 | 1.000|
|                  | 8         | 46   | .000 | .000 | .000 | .000 | .000 | .000 |
| Marriage         | 0         | 108  | 1.000|      |      |      |      |      |
|                  | 1         | 251  | .000 |      |      |      |      |      |
| Gender           | 0         | 236  | 1.000|      |      |      |      |      |
|                  | 1         | 123  | .000 |      |      |      |      |      |

Notes: The data are from the questionnaire.

3.2. Omnibus tests of model coefficients

Table 2. Omnibus tests of model coefficients.

|        | Chi-square | Degree of freedom (df) | Sig.  |
|--------|------------|------------------------|-------|
| Step   | 27.616     |                        | .000  |
| Step   | 16.679     |                        | .000  |
| Block  | 44.295     | 2                      | .000  |
| Block  | 54.411     | 3                      | .000  |
| Block  | 6.107      | 1                      | .013  |
| Model  | 60.519     | 4                      | .000  |
| Model  | 64.763     | 5                      | .000  |

Wherein, in the model line, the probability P-value is <0.05, which means that the ratio of at least one of the variables included in the model is statistically significant, that is, all the parameters are not all 0. Therefore, there is a significant relationship between all the explanatory variables and Logit P in the model at this time, which shows that the regression model has significance.
3.3. Goodness of fit test of the model
Hosmer and lemeshow fit test can better test the fit degree of regression model. Based on the actual needs of this study, this paper uses Hosmer and lemeshow to test the goodness of fit of logistic regression equation.

Table 3. Hosmer and lemeshow test of the model.

| Step | Chi-square | Degree of freedom (df) | Sig. |
|------|------------|------------------------|------|
| 1    | 9.618      | 3                      | .022 |
| 2    | 12.662     | 7                      | .081 |
| 3    | 11.029     | 8                      | .029 |
| 4    | 7.464      | 8                      | .488 |
| 5    | 3.713      | 8                      | .882 |

Notes: The data are from the questionnaire.

The results in the table show that the value is not less than 0.05, then 0 hypothesis is accepted, indicating that the goodness of fit of the model is high, that is, the model can better fit the data. \( P = 0.882 > 0.05 \)

3.4. Regression results

Table 4. Variables in the equation.

| Coefficient | Standard deviation S. E. | Wald | Degree of freedom df | Sig. | Ratio Exp (B) | 95.0% C.I.for EXP(B) | Lower | Upper |
|-------------|--------------------------|------|----------------------|------|---------------|---------------------|-------|-------|
| Step 1a     | Monthly family income    | 0.996 | 0.224 | 19.805 | 1 | 0.000 | 2.707 | 1.746 | 4.198 |
|             | Constant                 | -.284 | 0.513 | 0.307 | 1 | 0.579 | 0.753 |      |      |
| Step 2b     | Monthly family income    | 0.961 | 0.234 | 16.918 | 1 | 0.000 | 2.614 | 1.654 | 4.131 |
|             | Degree of trust          | 1.103 | 0.283 | 15.233 | 1 | 0.000 | 3.014 | 1.732 | 5.246 |
|             | Constant                 | -3.406 | 0.959 | 12.625 | 1 | 0.000 | 0.033 |      |      |
| Step 3c     | Monthly family income    | 0.962 | 0.252 | 14.610 | 1 | 0.000 | 2.618 | 1.598 | 4.289 |
|             | Family members           | -0.648 | 0.219 | 8.748 | 1 | 0.003 | 0.523 | 0.341 | 0.804 |
|             | Degree of trust          | 0.961 | 0.289 | 11.097 | 1 | 0.001 | 2.616 | 1.486 | 4.605 |
|             | Constant                 | -0.442 | 1.386 | 0.102 | 1 | 0.750 | 0.643 |      |      |
| Step 4d     | Age                      | -0.627 | 0.259 | 5.863 | 1 | 0.015 | 0.534 | 0.321 | 0.887 |
|             | Monthly family income    | 1.113 | 0.268 | 17.300 | 1 | 0.000 | 3.045 | 1.802 | 5.145 |
|             | Family members           | -0.746 | 0.231 | 10.397 | 1 | 0.001 | 0.474 | 0.301 | 0.746 |
|             | Degree of trust          | 0.939 | 0.289 | 10.543 | 1 | 0.001 | 2.558 | 1.451 | 4.508 |
|             | Constant                 | 1.051 | 1.544 | 0.463 | 1 | 0.496 | 2.860 |      |      |
| Step 5e     | Age                      | -0.644 | 0.264 | 5.972 | 1 | 0.015 | 0.525 | 0.313 | 0.880 |
|             | Education background     | 0.384 | 0.185 | 4.285 | 1 | 0.038 | 1.467 | 1.021 | 2.110 |
|             | Monthly family income    | 1.018 | 0.270 | 14.217 | 1 | 0.000 | 2.767 | 1.630 | 4.698 |
|             | Family members           | -0.767 | 0.236 | 10.584 | 1 | 0.001 | 0.465 | 0.293 | 0.737 |
|             | Degree of trust          | 1.049 | 0.303 | 11.970 | 1 | 0.001 | 2.854 | 1.576 | 5.169 |
|             | Constant                 | -0.313 | 1.683 | 0.035 | 1 | 0.853 | 0.731 |      |      |

Notes: The data are from the questionnaire.

As Forward: LR method is used to screen the independent variables, the explanatory variables and parameters listed in the above table are the final screened independent variables and their regression coefficients. Based on the results summarized through five step-by-step screening, we can see that
the last step (step 5), the Sig. values of five explanatory variables: the consumer's monthly family income, trust degree, total number of family members, age and education are less than 0.05 significance level, indicating that these five explanatory variables are all variables that significantly affect consumers' willingness to pay, which should be put into regression formula [5-6]. However, Sig. value of constant is greater than 0.05 significance level, it will not be put into the formula. The final model expression is as follows:

\[
\text{Logit } P = -0.644X_1 + 0.384X_2 + 1.018X_3 - 0.767X_4 + 1.049X_5
\]

(1)

Wherein, \(X_1\) represents age, \(X_2\) represents education background, \(X_3\) represents monthly family income, \(X_4\) represents total number of family members, and \(X_5\) represents consumers' trust in traceability labels.

It can be seen that consumers' age, education background, monthly family income, consumers' trust in traceability labels and the total number of family members will significantly influence consumers' willingness to pay for food traceability labels. Next, we will further analyze the degree of influence and direction.

According to the regression results in the table, the partial regression coefficient B value of the consumer age explanatory variable is equal to -0.644, which means that Logit P decreases by an average of 0.644 units for each grade increase of the consumer age group, for example, the grade of age group increases from 26-35 to 36-45. The ratio Exp (B), i.e. OR value is equal to 0.525, which means that for each grade increase in the consumers' age group, the incidence of the older consumers being willing to pay for food traceability labels is 0.525 times higher than that of the younger consumers when other independent variables remain unchanged, indicating that the younger consumers are more willing to pay for food traceability labels;

The partial regression coefficient B value of the consumer education background variable is equal to 0.384, which can be explained that Logit P decreases by an average of 0.384 units for each grade increase of the consumer education degree, such as from middle school to high school. Exp (B) is equal to 1.467, which means that for each grade increase of consumer education degree, the incidence of consumers with higher education to pay for food traceability labels is 1.467 times higher than that of consumers with lower education when other independent variables remain unchanged, indicating that consumers with higher education are more inclined to pay for food traceability labels;

The partial regression coefficient B value of the monthly family income variable is equal to 1.018, which can be explained that Logit P decreases by an average of 1.018 units for each grade increase of the monthly family income, such as the increase of monthly family income level from RMB 3,000-8,000 to RMB 8,000-12,000. Exp (B) is equal to 2.767, which means that the incidence of consumers with higher income to pay for food traceability labels is 2.767 times higher than that of consumers with lower income when other independent variables remain unchanged, indicating that consumers with higher income are more inclined to pay for food traceability labels;

The partial regression coefficient B value of the total number of family members is equal to -0.767, which can be explained that Logit P decreases by an average of 0.767 units for each grade increase of the monthly family income, such as the increase of the total number of family members from 2 to 3. Exp (b) is equal to 0.465, which means that the incidence of consumers with more family members to pay for food traceability labels is 0.465 times higher than that of consumers with less family members when other independent variables remain unchanged, indicating that consumers with less family members are more inclined to pay for food traceability labels;

The partial regression coefficient B value of the variable of consumers' trust in the traceability labels is equal to 1.049, which can be explained that Logit P increases by an average of 1.049 units for each grade increase of the consumers with higher trust in the traceability labels compared with the consumers with lower trust in the traceability labels. Exp (B) is equal to 2.854, which means that the incidence of consumers with higher trust in the traceability labels to pay for food traceability labels is 2.854 times higher than
that of consumers with lower trust in the traceability labels when other independent variables remain unchanged, indicating that consumers with higher trust in the traceability labels are more inclined to pay for food traceability labels.

From the eliminated variables in the following table "variables not in the equation", it can be seen that the Sig. values of gender (1), marriage (1), occupation, number of children in the family, degree of concern, risk perception and cognitive level are all greater than 0.05, which means that these independent variables are not significant influence variables and have no statistical significance, so they are not screened into the model.

### Table 5. Variables not in the Equation.

| Variables                  | Score | Degree of freedom (df) | Sig. |
|---------------------------|-------|------------------------|------|
| Gender (1)                | 0.000 | 1                      | 0.994|
| Marriage (1)              | 0.020 | 1                      | 0.889|
| Occupation                | 11.851| 7                      | 0.106|
| Occupation (1)            | 0.029 | 1                      | 0.865|
| Occupation (2)            | 2.060 | 1                      | 0.151|
| Occupation (3)            | 7.631 | 1                      | 0.006|
| Occupation (4)            | 2.146 | 1                      | 0.143|
| Occupation (5)            | 1.124 | 1                      | 0.289|
| Occupation (6)            | 0.017 | 1                      | 0.897|
| Occupation (7)            | 2.749 | 1                      | 0.097|
| Number of children in the family | 0.384 | 1                      | 0.536|
| Degree of concern         | 2.174 | 1                      | 0.140|
| Risk perception           | 0.205 | 1                      | 0.650|
| Awareness degree;         | 0.449 | 1                      | 0.503|
| Overall Statistics        | 14.584| 13                     | 0.334|

Notes: The data are from the questionnaire.

### 4. Conclusions

According to the above table, the main factors, significance and influence degree that influence consumers' willingness to pay for traceability labels of infant milk powder are specifically analyzed and summarized as follows:

1. The age has a significant influence on the customers' willingness to pay for food traceability labels at the level of 0.05 and the regression coefficient is -0.644, indicating that the younger the consumers are, the stronger the willingness to pay for food traceability label is, which is consistent with the conclusions of previous scholars. Generally speaking, young people are more likely to accept new things and new products, have a higher expectation of future income, and their consumption level is relatively high. Therefore, young people are more willing to pay for food traceability labels.

2. The education background (education level) has a significant influence on the customers' willingness to pay for food traceability labels at the level of 0.05, with a regression coefficient of 0.384, indicating that consumers with higher education have a stronger willingness to pay for food traceability label. This may be because consumers with higher education have higher cognitive ability, higher degree of acceptance of new things, higher concern for their own health, and better economic strength. Therefore, consumers with higher education are more willing to pay extra price for food traceability labels.

3. The monthly family income level has a significant influence on the customers' willingness to pay for food traceability labels at the level of 0.05, with a regression coefficient of 1.018, indicating...
that consumers with higher monthly family income are more willing to pay for food traceability labels. This is because consumers with low monthly family income can't buy food in a higher attribute level just when their income can only meet the lower material needs, so they will give up the more expensive traceable food when they buy food; For consumers with higher monthly family income level, there will be more disposable income left. At this time, they start to pursue higher-level food safety and other needs. Therefore, consumers with higher income can accept a certain degree of additional payment price in the selection of infant milk powder, and prefer safer traceable food. In addition, the sensitivity of different groups of monthly family income to price is also different. Relatively speaking, consumers with lower monthly family income are more sensitive to price changes. The increase of traceable food price will reduce or even give up their demand. The higher the traceable food price is than the ordinary food price, the stronger their sensitivity is. When they feel that the price exceeds their income tolerance, they will be more sensitive Around, they will choose to give up traceable food; Consumers with higher monthly family income are less sensitive to the high price of traceable food. In the final process of determining whether to pay for traceable infant milk powder, income level still plays a role.

(4) The total number of family members has a significant influence on the customers' willingness to pay for food traceability labels at the level of 0.05, with a regression coefficient of -0.767, indicating that consumers with less family members are more willing to pay for traceable food, whereas consumers with more family members are not willing to pay additional price for food traceability labels. This is because, for families with more members, including those with many children, and with many members to be supported, the per capita income of the family is relatively low, which leads to a high sensitivity to the price change of food, a large total burden on the family, so these families may not be willing to pay for the food traceability labels. However, consumers with less family members have a relatively high disposable income and light family burden, and they have a higher willingness to pay for food traceability labels.

(5) The degree of consumer's trust in food traceability labels has a significant influence on consumer's willingness to pay at the level of 0.05, with a regression coefficient of 1.049, indicating that the more consumers trust food traceability labels and the information they find, the more willing they are to pay for food traceability labels. On the contrary, if consumers do not trust food traceability labels or the information they carry, they are not willing to pay extra price. The possible reason is that the more consumers trust the food traceability labels, the more they will believe that they can find the real and reliable food information they need through the food traceability labels, and believe that the traceable food is safer. However, as food traceability label is still a new thing in China, and the relevant laws and regulations are not perfect, many consumers do not know or understand food traceability labels. In the context of frequent dairy safety events, there is a crisis of trust of consumers in the government supervision and governance, and have a sense of distrust of dairy enterprises, which leads to consumers' doubts about the food traceability labels and the inquired information, so the consumers who are willing to pay for food traceability labels are not the majority at present. If the government can strengthen the propaganda of food traceability labels, so that more consumers can understand and trust food traceability labels, and the government can strengthen the supervision of food traceability system construction, so that consumers can trust the government's supervision and policy guidance on food traceability labels, more consumers will trust and choose traceable food.

(6) The results show that the factors that do not significantly influence the willingness to pay are consumer gender, age, marital status, the number of family children, the degree of consumer's concern about food safety, the level of consumer's risk perception of food safety and the degree of consumer's cognition of infant milk powder traceability labels. This result only shows that to some extent, the data processing results of this survey result show that these factors are not significant factors that influence the willingness of consumers to pay for food traceability labels, but they are not absolute results, which may also be caused by incomplete data, or by the effectiveness of problem setting. However, generally, the survey results and analysis show that these influencing factors are not included in the scope of main influencing factors.
Acknowledgement
About the author: Xia LI (1981-), female, Doctoral Candidate, School of Management, Fujian Agricultural and Forestry University, Working unit of the author: Lecturer, School of Economics and Trade, Fujian Jiangxia University. Corresponding author: Heliang HUANG (1965-), male, Professor, Doctoral Supervisor, School of Economics, Fujian Agricultural and Forestry University.

References
[1] ANGULOA M, GIL J M, TAMBURO L. Food Safety and Consumers ‘Willingness to Pay for Labeled Beef in Spain [J]. Journal of Food Products Marketing, 2015, (3):89-105.
[2] Linhai WU. Consumer preference analysis of traceability origin attribute and traceability information attribute of traceable pork [J] Chinese Rural Economy, 2015 (6): 47-73.
[3] Shijiu YIN, Yingjun XU, Yusheng CHEN. How food quality information labels affect consumers' preferences -- a selection experiment based on 843 samples in Shandong Province [J] China Rural Survey, 2015, (01): 39-49+94.
[4] Linhai WU, Shuxian WANG, Wuyang Hu. Consumers' preference and willingness to pay for traceable food attributes: Pork case [J] Chinese Rural Economy, 2014 (08): 58-75.
[5] Xiaoshuan ZHANG, Tieyan ZHANG, Changyang MA, etc Guangdong consumers' cognition and willingness to pay for the logo of edible agricultural products [J] Journal of China Agricultural University, 2015,20 (01): 268-274.
[6] Baichen JIANG, Yunsheng MI, Qiaoyan ZHU Study on consumer selection preference and willingness to pay for quality characteristics of high-quality agricultural products based on Hedonic model [J] Journal of Nanjing Agricultural University (Social Sciences Edition), 2017, 17(04): 128-137+160.