EasyChair Preprint

№ 8463

# Influence of Traceability Information Display on Consumption Behaviour in Cross-Border E-Commerce

Ruiqian Yang, Rong Du, Yimeng Qi and Minna He



July 12, 2022

# Influence of Traceability Information Display on Consumption Behaviour in Cross-border E-commerce

**Ruiqian Yang**
Xidian University
ruiqian-yang@stu.xidian.edu.cn

**Rong Du**
Xidian University
durong@mail.xidian.edu.cn

**Yimeng Qi**
Xidian University
yimengqi@stu.xidian.edu.cn

**Minna He**
University College London
minna.he.20@ucl.ac.uk

## Abstract

*The introduction of traceability information on cross-border e-commerce platforms reduces information asymmetry and improves consumers' perception of trust. In this study, we combined the data-driven and model-driven methodologies, created a new index of traceability information display degree, and constructed econometric models to analyze the influential effect and marginal effect of traceability information display degree on consumer consumption behaviour about cross-border e-commerce products. Furthermore, we examined the moderating effect of platform on the display degree of traceability information and conducted a robustness test using propensity score matching (PSM). This paper provides a new perspective on traceability information and makes recommendations for how to improve the construction of traceability systems on cross-border e-commerce platforms.*

**Keywords**: Traceability information display, Cross-border e-commerce products, Consumption behaviour, PSM

## 1. Introduction

With the strong support of Internet marketing and national policies, cross-border e-commerce has developed rapidly. Consumers urgently need a real, reliable, and credible mechanism to provide a strong basis when purchasing. The quality traceability system based on blockchain technology addresses this issue by enabling the full cycle monitoring of product production, transportation, and consumption. Tmall.com and JD.com, the two Chinese e-commerce platforms, have introduced traceability information into product display pages. JD.com has created a full-process anti-counterfeiting traceability system through joint branding, monitoring agencies, government supervision, and other departments.

Traceability information includes product-related production, transportation information, etc., which can improve supply chain transparency (Sodhi and Tang 2019), reduce information asymmetry（Liu *et al.* 2018）, and improve consumers' trust perception (Suhandoko *et al.* 2021). Scholars have demonstrated that disease, pests, and input traceability information can increase consumers' trust in food safety (Matzembacher *et al.* 2018). And more and more scholars are increasingly focusing on the influence of food traceability information on consumers' purchase behaviour, such as fresh agricultural products (Yu *et al.* 2021) and pork products (Xu *et al.* 2019).

Scholars have demonstrated that the perceived quality of traceability information has a beneficial influence on purchase intention toward organic food (Wu *et al.* 2021), and the food traceability system positively affects purchase intention (Yuan *et al.* 2020).

However, the traditional traceability information mainly focuses on the production chain, and maybe there exists information fraud. The traceability information of cross-border e-commerce products can display the information of production, transportation, and other chains' information, which is more abundant. What's more, this traceability information on cross-border e-commerce platforms is guaranteed by blockchain technology, providing a trustworthy and difficult to counterfeit quality signal to consumers (Connelly *et al*. 2011).

At present, the studies are mainly carried out through questionnaire surveys which lacked objective data to verify the actual effect of product traceability on consumers' purchase behaviour. Therefore, we intend to verify the actual effect and boundary conditions of traceability information on consumers' purchase behaviours by obtaining real data from cross-border e-commerce platforms.

Information exposure is the number of times information is repeated (Campbell and Kevi 2003). We discovered that there are various ways to display traceability information on cross-border e-commerce platforms, and the quantity of information display will affect consumers' recognition and attitude towards products, but there is no consistent conclusion on the effect. Short video e-commerce information display has a significant positive influence on purchase intention (Guo *et al.* 2021). Li and Zheng have demonstrated that the frequency of brand information presentation and consumers' brand attitude is an inverted "U" shaped relationship (Li and Zheng 2021). In terms of research methods, most scholars conducted research using questionnaire survey, qualitative research, and other methods to carry out research, without truly realizing quantitative research of information display. To fill in this gap, we attempt to further excavate the traceability information, complete the measurement of the display degree of traceability information, and explore the causal relationship between the display degree of traceability information and consumption behaviour.

For the above purpose, in this study, we combine the data-driven method and model-driven method. First, we integrated the structured data and unstructured data to measure the traceability information display degree. Second, we used econometric models to examine the relationship between traceability information and cross-border e-commerce consumption behaviour. Finally, we discussed the causal relationship by using the propensity score matching method.

## 2. Data and variable selection

This data sample originated from the International Channel of JD.com Mall. This channel is a one-stop consumption platform for imported products operated by JD.com, which is mainly engaged in the cross-border import product industry. It is the first consumption platform in China that fully focuses on large import businesses, bringing consumers a better and richer shopping experience with imported products.

We gathered the information on the product pages using the platform's API. Three important categories of cross-border e-commerce products were gathered including milk powder, cosmetics, and health care products. Following a series of selections, a total of 1536 products were reserved. Text preprocessing was performed on the obtained text data, including product introduction and Q&A text. The Jieba Python package was used to segment the text content, and then the segmentation results were compared with the self-built traceability dictionary.

Given the important impact of brand value on product sales, we measured product brands according to the China Brand Power Index (CBPI), an independent and unbiased study based on the feedback of Chinese consumers on products or services. CBPI can be used to measure the brand power index that affects consumer purchase behaviour. For the products that were not on the CBPI list, we assigned a brand value based on the influence of the products on JD.com.

There are six major variables in this study, and the definitions and measurement methods of each variable are shown in Table 1 and Table 2. We cannot directly obtain the sales volume of the products on the platform. Considering that consumers will comment after purchasing, the system will give default praise to those who have not commented. So, we selected the number of comments as the proxy index of product sales. In order to compress the dimension of variables, and weaken the collinearity and heteroscedasticity, log transformations were made to price, the total number of reviews, brand power index, and sales to make the model results more robust.

**Table 1**. Description About Traceability Information Display Degree

| Variable | Secondary Variable | Definition and Measurement |
|---|---|---|
| Trace<br>(Range 1-9) | Trace_1<br>JD.com international label<br>Trace_2<br>Quality traceability label<br>Trace_3<br>JD.com international label and introduction<br>Trace_4<br>Traceability information in the product introduction<br>Trace_5<br>Assured purchase quality traceability label and introduction<br>Trace_6<br>Bonded warehouse information<br>Trace_7<br>Quality traceability label and special introduction | Traceability information display degree<br>1= This product has the corresponding part of the traceability information<br>0= This product does not have the corresponding part of the traceability information<br>$Trace = \sum_{i=1}^{9} Trace\_i$ |

| | |
|---|---|
| Trace_8<br>JD.com international label and<br>detailed introduction | |
| Trace_9<br>Traceability information in the<br>Q&A content | |

**Table 2.** Key Variables Description

| Variable | Definitions and Measurement | Adds |
|---|---|---|
| Price | Products price at the point in time of data collection | *Ln* change |
| Self | Whether are platform self-operated products<br>1= The product has this label<br>0= The product does not have this label | 0/1 variable |
| all_num | The total number of reviews for all products contained on the search page, which will influence the next behaviour of the consumer | *Ln* change |
| C_BPI | China Brand Power Index and Platform Brand Power Index which can reflect the brand value. The higher the power, the less brand influence | *Ln* change |
| Sales | The number of reviews for this product can reveal the sales of the product | *Ln* change |

## 3. Empirical analysis

### *3.1 Effect analysis of traceability information display degree (TIDD)*

#### *3.1.1 Influence on the sales*

The traceability information display degree (TIDD) is treated as the independent variable and product sales as the dependent variable. The control variables included product price, whether are platform self-operated products, brand power index, and the total number of reviews on search pages. We constructed a multiple regression model as illustrated in Equation (1), to analyze the influence of the TIDD on the sales of cross-border e-commerce products.

$$\ln \_sales = \beta_0 + \beta_1 Trace + \beta_2 \ln \_price + \beta_3 self + \beta_4 \ln \_all\_num + \beta_5 \ln \_C\_BPI + \epsilon \quad (1)$$

The core variable and other variables were added in sequence using Hierarchical Regression, and the results are displayed in Table 3. Only the TIDD is included in the first column's explanatory factors. As can be observed, increasing the TIDD by one-unit results in a 41.2% increase in product sales. As the number of control variables grows in columns 2 to 5, the $R$^2 of the model improves steadily, and the explanatory power of product sales becomes more muscular. Additionally, all models pass the F-test ($p$ <0.001).

Considering that the influence of the TIDD on sales may be different for different products, we divided the products into three groups: milk power, cosmetic, and health care products. We conducted a Grouped Regression, and the results are shown in Table 4. The TIDD has been

demonstrated to have a beneficial influence on product sales in all three product categories. However, the coefficient of influence and platform effect is smaller in health care products than in milk powder and cosmetics, which may be a result of the unique nature of health care products, where the key concern of customers is about the product function.

There is a platform self-operated in JD.COM, which means that the platform directly procures and sells cross-border products, and provides a series of guarantees in terms of quality, transportation, and after-sales service. So, consumers' perceptions about the quality of whether the products are platform self-operated also differ, and the traceability information of these products will also have a different influence on consumers. We, therefore, introduced a platform self-operated dummy variable and constructed an interaction term with the TIDD to investigate the possible platform effect. The regression model is shown in Equation (2).

$$\ln\_sales = \alpha_0 + \alpha_1 Trace + \alpha_2 self + \alpha_3 Trace * self + \alpha_4 \ln\_price + \alpha_5 \ln\_all\_num + \alpha_6 \ln\_C\_BPI + \delta \quad (2)$$

The results are shown in the corresponding positions in Table 3 and Table 4. It can be seen that the regression coefficients are significant for both overall and specific products, which means that when the products belong to the platform self-operated, the TIDD is more likely to promote the sales of cross-border products. So, the platform effect exists.

**Table 3.** The Results of Parameter Estimates

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Trace | 0.412*** | 0.415*** | 0.455*** | 0.424*** | 0.454*** | 0.472*** |
| Ln_price | | -0.157*** | -0.270*** | -0.195*** | -0.259*** | -0.287*** |
| self | | | 2.336*** | 1.612*** | 1.711*** | 1.511*** |
| Ln_all-num | | | | 0.710*** | 0.708*** | 0.705*** |
| Ln_C-BPI | | | | | -0.214** | -0.252** |
| self*Trace | | | | | | 1.572*** |
| Constant | 5.194*** | 6.041*** | 6.454*** | 1.441*** | 1.450*** | 1.770*** |
| Observations | 1536 | 1536 | 1536 | 1536 | 1536 | 1536 |
| $R^2$ | 0.098 | 0.101 | 0.410 | 0.659 | 0.696 | 0.704 |

Dependent Variable: Ln_sales

Note: *** Significant at the 0.1% level; ** Significant at the 1% level; *Significant at the 5% level.

We further understood the marginal effect of different levels of traceability information display on product sales by dividing the sample into three groups based on the degree of traceability information display, namely the high display group (HTrace), the medium display group (MTrace), and the low display group (LTrace). Then set dummy variables with the low display group as the base group and conducted regression analysis. The results showed the MTrace regression coefficient of 0.549 ($p$<0.001) and 0.880 for Htrace ($p$<0.001). Compared to the low display group

of traceability information, the marginal effect of medium and high display is gradually increasing. However, the overall display degree of traceability information on the platform is currently not high enough and may not have reached the peak of the declining marginal effect. So, the marginal effect may level off or even appear to decline in the future as the overall display degree of traceability information increases.

**Table 4.** The Results Grouped Regression

| | Milk Power | Cosmetic | Health Care |
|---|---|---|---|
| Trace | 0.394*** | 0.459*** | 0.219** |
| Ln_price | -0.146** | -0.332*** | -0.276*** |
| self | 2.222*** | 2.566*** | 1.819** |
| Ln_all-num | 0.755*** | 0.658*** | 0.658*** |
| Ln_C-BPI | -0.216** | -0.267*** | -0.225*** |
| self*Trace | 2.848*** | 4.248*** | 0.952* |
| Constant | 1.125*** | 3.311*** | 2.508*** |
| Observations | 460 | 620 | 456 |
| $R^2$ | 0.708 | 0.710 | 0.592 |

Note: *** Significant at the 0.1% level; ** Significant at the 1% level; *Significant at the 5% level.

### *3.1.2 Influence on the scale effect*

We already know that the degree of traceability information display has a positive effect on the sales of cross-border e-commerce products via linear OLS regression and regression coefficient testing. But we have yet to understand the trend and change in the degree of traceability information display's influence on sales. As a result, we used Quantile Regression which indicated in Equation (3), to gain a better understanding of the differences and patterns of change in the influencing factors for products with varying sales volumes.

$$Q_{(q)}(y_{sales}|x_i) = \beta_q x_i' + \mu_q \quad (3)$$

$Q_{(q)}(y_{sales}|x_i)$ denotes sales at quantile q conditional on the explanatory variable being $x_i$, $\mu_q$ is the random disturbance term, and $\beta_q$ is the regression coefficient at quantile q. As q changes, the conditional distribution of $y_{sales}$ on $x_i$ can be obtained. Table 5 shows the results of Quantile Regression.

The results show that the influence of TIDD on cross-border e-commerce product sales is basically significantly positive, but there exist situations where TIDD fails to work. When product sales are between the 0.4 and 0.5 quartiles, traceability information has a negligible effect on product sales, which may be because other influencing factors interest customers more than traceability information. The influence of traceability information presentation on product sales is

strong for niche products and products with a large market share, demonstrating that customers care about the quality traceability when purchasing these products. The impact coefficient is most significant for products in the 0.1 quartiles, demonstrating that the degree of traceability information display is a significant factor impacting sales of highly niche products.

**Table 5.** The Results of Quantile Regression

| | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
|---|---|---|---|---|---|---|---|---|---|
| Trace | 0.554*** | 0.381*** | 0.348*** | 0.064 | 0.076 | 0.360*** | 0.396*** | 0.401*** | 0.447*** |
| Ln_price | -0.229*** | -0.197*** | -0.220*** | -0.241*** | -0.238*** | -0.239*** | -0.228*** | -0.169** | -0.112* |
| self | 1.292*** | 1.244*** | 1.199*** | 1.253*** | 1.227*** | 1.309*** | 1.295*** | 1.188*** | 1.161*** |
| Ln_C-BPI | -0.226*** | -0.216*** | -0.238*** | -0.201*** | -0.263*** | -0.208*** | -0.234*** | -0.221*** | -0.218*** |
| Ln_all-num | 0.626*** | 0.668*** | 0.643*** | 0.652*** | 0.679*** | 0.707*** | 0.761*** | 0.818*** | 0.837*** |

Note: *** Significant at the 0.1% level; ** Significant at the 1% level; *Significant at the 5% level.

### *3.2 Propensity Score Matching (PSM)*

While the display of traceability information has an effect on cross-border e-commerce product sales, the degree to which it is displayed is not random and is determined by a number of factors, including national policy regulation, improved e-commerce platform governance, and product characteristics. To further investigate the effect of the degree of traceability information display on product sales, control groups must be established for comparison. The control groups that are as similar to the treated group as possible are identified using propensity score matching (PSM). This method allows for some control of observable heterogeneity factors and reduces selectivity bias.

In this section, we regarded the samples with TIDD greater than or equal to 7 as the treated group, considered them as products with high TIDD, and categorized the other samples as the control group. We set product price, whether are platform-owned products, brand power index, and the total number of reviews on search pages as covariates, established a logistic regression model, and estimated propensity scores. Propensity score matching was then performed, and the matching success rate was 99.676%, which was a good matching result.

The findings of the PSM parallel hypothesis test, which may be used to further evaluate the effectiveness of PSM, are shown in Table 6. As can be observed, the absolute values of standardized bias were all less than 20% following matching, indicating a significant reduction in standardized bias. The t-tests unmatched were all significant ($p<0.05$), whereas the t-tests conducted after matching were not significant ($p>0.05$), indicating that the characteristics of the treated and control groups remained essentially identical after matching. All in all, the matching effect was generally good.

The average treatment effect of traceability information display on product sales was calculated based on the matched treatment and control group samples. The difference between the treated and

control groups before matching was 1.764 and statistically significant ($p$ <0.001). After propensity score matching, the difference was 1.487, which is less than that before matching but is still significant at the 0.1% level, indicating the robustness of the findings.

**Table 6.** The Results PSM Parallel Hypothesis Test

| Variable | Unmatched/Matched | Treated | Control | % bias | % reduct \|bias\| | *p* |
|---|---|---|---|---|---|---|
| Ln_price | U | 5.571 | 5.452 | 14.65% | 25.44% | 0.017 |
| | M | 5.567 | 5.648 | -10.93% | | 0.175 |
| self | U | 0.703 | 0.143 | 137.60% | 99.49% | 0.000 |
| | M | 0.702 | 0.699 | 0.71% | | 0.930 |
| Ln_all-num | U | 10.810 | 8.584 | 92.84% | 94.72% | 0.000 |
| | M | 10.801 | 10.915 | -4.90% | | 0.543 |
| Ln_C-BPI | U | 2.415 | 2.846 | -42.52% | 66.52% | 0.000 |
| | M | 2.423 | 2.572 | -14.24% | | 0.077 |

## 4. Discussion

This study examined the influence of the traceability information display degree on consumers' consumption behaviour for cross-border e-commerce products using a combination of data-driven and model-driven methods and empirically analyses data for 1,536 products. In the data-driven aspect, some textual data were analyzed and combined with structured data to create a variable for traceability information display degree and to improve the quantification of the degree of traceability information display. In the model-driven aspect, OLS regression and Grouped Regression were used to verify the significant influential effect and the marginal effect of the TIDD on consumers' consumption behaviour of cross-border e-commerce products. Quantile Regression was used to verify the scale effect of the TIDD on consumers' consumption behaviour. Finally, the robustness test was completed by propensity score matching.

For cross-border e-commerce platforms, it is necessary to increase the degree of display and publicity of traceability information. In order to encourage consumers to pay attention to and participate in discussions about traceability information, thereby increasing the platforms' influence. Simultaneously, the display of traceability information should be differentiated by product type and sales volume. For example, for niche products and products with a large market share, the display of traceability information should be increased further. For products in the middle of the market, consider introducing additional strategies to increase product sales.

There are still some unresolved issues in this study. Firstly, only three product categories were analyzed and the research findings may not yet be applicable to other cross-border e-commerce products. Secondly, the traceability information in reviews and product detail pages was not considered. In the future, natural language processing technology and image analysis technology

are considered to improve the construction of traceability information display index. Finally, the degree of traceability information display is not weighted, but different traceability information indicators should have different weights in practice.

**References**


1. Campbell, M. C., and Keller, K. L. "Brand familiarity and advertising repetition effects," *Journal of Consumer Research* (30:2), 2003, pp. 292-304.
2. Connelly, B. L., Certo, S. T., Ireland, R. D., and Reutzel, C. R. "Signaling theory: A review and assessment," *Journal of Management* (01), 2011, pp. 39-67.
3. Guo, H., Wang, H., Pang, L., and Chang, Q. "Research on the Influence of Short Video E-commerce Information Display on Customer Engagement Behavior," *Price: Theory & Practice* (09), 2021, pp. 155-158.
4. Li, X., and Zheng, C. "The Influence of Frequency of Brand Information Presentation on Consumers' Brand Attitude: Based on Social Media Narrative Advertising," *Social Science Front* (01), 2021, pp. 250-255.
5. Liu, C., Li, J., Steele, W., and Fang, X. "A Study on Chinese Consumer Preferences for Food Traceability Information Using Best-Worst Scaling," *PloS one* (13:11), 2018, e0206793.
6. Matzembacher, D. E., do Carmo Stangherlin, I., Slongo, L. A., and Cataldi, R. "An Integration of Traceability Elements and Their Impact in Consumer's Trust," *Food Control* (92), 2018, pp. 420-429.
7. Sodhi, M. S., and Tang, C. S. "Research Opportunities in Supply Chain Transparency," *Production and Operations Management* (28:12), 2019, pp. 2946-2959.
8. Suhandoko, A. A., Chen, D. C. B., and Yang, S. H. "Meat Traceability: Traditional Market Shoppers' Preferences and Willingness-to-Pay for Additional Information in Taiwan," *Foods* (10:8), 2021, 1819.
9. Wu, X., Xiong, J., Yan, J., and Wang, Y. "Perceived quality of traceability information and its effect on purchase intention towards organic food," *Journal of Marketing Management* (37:13-14), 2021, pp.1267-1286.
10. Xu, L., Yang, X., Wu, L., Chen, X., Chen, L., and Tsai, F. S. "Consumers' Willingness to Pay for Food with Information on Animal Welfare, Lean Meat Essence Detection, and Traceability," *International Journal of Environmental Research and Public Health* (16:19), 2019, 3616.
11. Yu, M., Yan, X., and Wang, J. "Research on the Influence of Traceability Information of Fresh Agricultural Products on Consumers' Purchasing Behavior," *Open Journal of Business and Management* (9:5), 2021, pp. 2370-2388.
12. Yuan, C., Wang, S., Yu, and X. T. "The Impact of Food Traceability System on Consumer Perceived Value and Purchase Intention in China," *Industrial Management & Data Systems* (120:4), 2020, pp. 810-824.

## Notes

The work described in this paper was supported by the National Natural Science Foundation of China (NSFC: 72171187).

Yimeng Qi and Minna He are joint third authors.