Analysis and evaluation of relevant influencing factors based on the big data of Douyin live broadcast sales

Faqiang Cui1 2*
1 Shandong Institute of Commerce and Technology, Jinan, China
2 North-Chiang Mai University, Chiang Mai, Thailand
Corresponding author and e-mail: Faqiang Cui, g636501003@northcm.ac.th

Abstract. In order to explore the influence of live broadcasters’ own factors on the effect of live delivery, we conducted field surveys of no less than 1680 minutes for live broadcasters and audiences participating in live events on the Douyin platform for 14 consecutive days. We also collected 488 valid questionnaires through online surveys, and used Eviews10.0 software to analyze the collected primary data. The results proved that the optimization of the live broadcaster’s image and the improvement of the interactive effect of the live broadcaster can better stimulate the consumers who participate in the live broadcast and continuously increase their consumption quota and consumption times. Through research, the live broadcaster has finally achieved the goal of improving the ability to carry goods by optimizing himself.

1. Introduction

1.1. Research Background
When most scholars talk about the perspective of studying live video sales, they focus on the head effect. Through the sinking of the research perspective, we pay attention to and understand the current status of the most extensive ordinary live video anchor live video sales. A further exploration of its own problems and solutions restricts the sales effect of live video anchors. The research results can help general live video anchors to better understand their own problems through comparison, and adjust their live video strategies in time to optimize the effect of live video sales.

1.2. Literature review
Live streaming is a relatively new concept, which is mainly manifested in the activity process of selling goods (services) through live streaming. According to the attributes and characteristics of the live broadcast platform, one type is the form of e-commerce live broadcast (Tan Yuli, 2017)[1], such as Taobao live broadcast, Jingdong live broadcast, and Pinduoduo live broadcast; the other is the form of live broadcast e-commerce (Kim DJ, 2014)[2], such as Douyin Live, Kuaishou Live and Momo Live, etc. Although the two platforms have different live broadcast services, they share the same live broadcast elements: live broadcasters and consumers. (Jia Xiaofeng, 2019)[3].

The characteristics of the live broadcast platform have common influences on the live delivery of goods, such as common experience, perceived efficiency and synchronization characteristics (Bründl S, Matt C, Hess T, 2017)[4], value compatibility, consumer experience delivery and good products Exhibitions and others (Chen Z, Benbasat I, CenfetelliRT, et al. 2017)[5], all have a positive stimulus effect on promoting consumer behavior. At the same time, combined with the theory of media richness,
the platform also shows the characteristics of multiple clues and real-time interaction (Du Yanwu, 2016) [6]. In short, the study of platform characteristics is the basis of this research, but it is not the focus of our research. We pay more attention to the evaluation of the influence of the broadcaster’s own factors on the choice of the broadcaster’s own factors.

In the research on live broadcasters, the obvious public personality and a high degree of familiarity with the product help to improve the effect of live streaming (Chan K K, Misra S. 1990) [7]. At the same time, the higher the degree of trustworthiness of the broadcaster and the stronger the professional characteristics, the better the attributes that stimulate consumption (Sussman SW, Siegal W S. 2003) [8]. Regardless of the characteristics of Internet celebrity leaders in the head effect, the live broadcast methods of ordinary broadcasters should be diverse and flexible (Zhong Xujun, Wang Yanrong, 2016) [9], authentic and reliable (Sussman SW, Siegal WS. 2003) [8], and also demonstrate effective communication and intimacy (Mengfei. 2012) [10]. The language personality and innovation of the broadcaster (Zhao Yan, 2018) [11] have a significant role in promoting the effect of long-term delivery. In summary, research scholars have interpreted the product professionalism, behavioral diversity, trust authenticity, and communication effectiveness of live broadcasters during the live broadcast process, and further study the influence of individual broadcasters’ own factors on live broadcast delivery. Lay a solid foundation.

Researchers also conducted a lot of research on live streaming from the perspective of consumers. They proposed that consumers actively watch the live broadcast of the broadcaster because they need to purchase goods on the one hand, and watch the live broadcast for entertainment (Cai J, Wohn DY. , Mittal A, et al. 2018) [12] and produce purchase behavior. At the same time, the main influencing factors that consumers are willing to consume also include: the authority of the live broadcast platform (Diao Leiyu, Wang Xicheng, Lu Xiaozhen. 2010) [13], and the awareness of commodity brands (Lu Qiang, Fu Hua. 2016) [14], The influence of the credibility of the merchant (Koufaris M, Hampton-Sosa W. 2004) [15] and the commodity price advantage (Dai Jianhua, Ma Haiyun, Wu Yingying. 2020) [16]. In summary, the researchers gave consumers a systematic analysis of the impact of merchants, products and purchase motivation on the effect of live broadcast delivery, and solved the paradigm research problem of our research on individual broadcasters. It promotes the research of individual live broadcasters from the perspective of consumers.

2. Research design

2.1. Data Sources

The research data mainly comes from the 15-day field survey of the research group. We first recruited 15 live broadcasters who volunteered to participate in the research on the university campus. According to the requirements, the live broadcasters who participated in the survey (10 randomly selected) From 20:00 to 22:00, a 120-minute live-streaming event was held, and members of the research team were arranged to record and supervise each broadcaster’s live-streaming process. There are no strict requirements for the products that are shipped live, and there are no restrictions on the product category. After the live broadcast, the supervising team members will promptly report the income of each live broadcast to the deputy team leader, who is responsible for statistical data and forming a statistical table.

At the same time, members of the research team distributed online questionnaires to customers who participated in the live broadcast of broadcasters, and collected 488 valid questionnaires. In the course of this research, one live broadcaster violated the rules of live broadcast. Therefore, we finally selected 14 ordinary live broadcasters who met the requirements, and the 14-day average data was used as the research basis. The specific data is shown in Table 1.
### 2.2. Variable selection
The study selected the income of the live broadcaster from each live broadcast as the main explanatory variable. At the same time, the image indicators of the live broadcaster are selected, and the average value is calculated by the audience who participated in the online survey; the average age of the live broadcaster; the interactive rating of the live broadcaster, the average value is calculated by the viewers participating in the online survey; and the live broadcast The gender coefficient of the host refers to the percentage of the number of female broadcasters in the total number of broadcasters. The above is used as an explanatory variable to construct a multiple regression evaluation model for the effect of live broadcast to better evaluate the effect of live broadcasters.

### 2.3. Model design
Let $Y_i$ be the income of the live broadcaster each time they bring the goods, $X_{1i}$ is the host image indicator, $X_{2i}$ is the average age of the host, $X_{3i}$ is the host’s interactive rating coefficient, and $X_{4i}$ is the host’s gender coefficient, $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ is a parameter, and $\mu_i$ is a random disturbance term. The establishment of the initial multiple regression model is mainly based on the fact that the model has not yet been subjected to parameter estimation and model testing, and there may be adjustments to the explanatory variables. For this reason, we call the formula (1) the initial model.

$$Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{3i}+\beta_4X_{4i}+\mu_i \quad (1)$$

### 3. Measurement inspection and result analysis

#### 3.1. Model estimation results and analysis

Table 2: (1) The estimated output results of the initial model

| Variable | Coefficient | Std. Error | t-Statistic | Prob.   |
|----------|-------------|------------|-------------|---------|
| C        | -100.2410   | 27.53598   | -3.640364   | 0.0054  |
| $X_{1i}$ | 1.569947    | 0.483499   | 3.247050    | 0.0100  |
From the estimation results in Table 2, it can be seen that the image of the live broadcaster and the interaction of the live broadcaster have a significant impact on the live broadcaster’s delivery effect, while the live broadcaster’s age and the live broadcaster’s gender have no significant impact on the live broadcast delivery effect. As a result, there may be multicollinearity between variables. Therefore, it is necessary to continue to use the method of stepwise regression analysis to examine the possible multicollinearity problems among explanatory variables, and to eliminate and replace related variables based on the economic significance of each explanatory variable.

3.2. Stepwise regression analysis

The test results of the initial model (1) fully illustrate that the age of the live broadcaster and the gender of the live broadcaster have no significant influence on the effect of live broadcast delivery. Comprehensive considerations, live broadcasters usually turn on live broadcast functions such as beauty during live broadcasts to help broadcasters of different genders and ages tend to be more in line with the audience's evaluation criteria, so there may be multiple collinearity. Therefore, the method of eliminating one by one is adopted, the model (2) is established, and the secondary regression analysis is performed.

\[ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \mu_i \]  

(2)

The same analysis method as the initial model (1) is adopted to obtain the results of the quadratic regression analysis, as shown in Table 3.

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|-------|
| C        | -103.8193   | 25.87955   | -4.011634   | 0.0025|
| X_{1i}   | 1.701320    | 0.409822   | 4.151362    | 0.0020|
| X_{2i}   | -0.340782   | 0.345520   | -0.986289   | 0.3473|
| X_{3i}   | 0.754748    | 0.267234   | 2.824299    | 0.0180|

From the results in Table 3, it can be seen that after the explanatory variable of the gender of the live broadcaster is first proposed, the live broadcaster’s image and the interaction of the live broadcaster still have a significant impact on the live broadcaster’s delivery effect, while the live broadcaster’s age has a significant effect on the live broadcaster’s delivery. The impact is still not significant. Therefore, it is necessary to eliminate the age of the live broadcaster and adopt further amendments to the model.

3.3. Revised model

Considering the results of the quadratic regression analysis, there may be a correlation between the live broadcaster’s age and the live broadcaster’s image, and there may be multiple collinearity. For this reason, after excluding the live broadcaster’s gender and the live broadcaster’s age, the
explanatory variables Correlation test between. See Table 4 for details.

| Explanatory variables | R-squared | P value of T statistic | P value of F statistic | Is it significant |
|-----------------------|-----------|------------------------|------------------------|-------------------|
| X_{1i}                | 0.9325    | 0.00                   | 0.00                   | Y                 |
| X_{1i}, X_{3i}        | 0.9593    | 0.00, 0.02             | 0.00                   | Y                 |

In summary, the model (3) is established and the regression analysis is continued.

\[ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_3 X_{3i} + \mu_i \]  

(3)

The same analysis method as the initial model (1) and (2) is adopted to obtain the third regression analysis result, as shown in Table 5.

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|-------|
| C        | -126.7436   | 11.36620   | -11.15092   | 0.0000|
| X_{1i}   | 1.972053    | 0.303935   | 6.488408    | 0.0000|
| X_{3i}   | 0.707785    | 0.262632   | 2.694971    | 0.0208|

From the results in Table 5, it can be seen that the image of the live broadcaster and the interaction of the live broadcaster have a significant impact on the delivery effect of the live broadcaster. Therefore, the regression result of the model is obtained:

\[ Y_i = -126.7436 + 1.972053 X_{1i} + 0.707785 X_{3i} \]  

(4)

From the analysis of the results, it can be seen that the effect of the live broadcaster’s delivery of goods is positively correlated with the live broadcaster’s image and the live broadcaster’s interaction, which is the same as the proportional relationship in the model. The economic significance is in line with the theory, indicating the goodness of fit of the revised model. Good, the constant term and the t-test value of each explanatory variable also passed, without multicollinearity.

3.4. Model checking

3.4.1. Goodness of fit test

\[ R^2 = 1 - \frac{RSS / (n - k - 1)}{TSS / (n - 1)} \]  

(5)

We use the Eviews10.0 calculation to get the result, the coefficient of determination $R^2=0.959343$, and the adjusted coefficient of determination $R^2 = 0.951951$, indicating that the sample regression equation fits the sample's observations well, and the explanatory variables evaluate the effect of the live broadcaster’s delivery of goods. The explanatory power of is 95.19%, and it can be judged that the modified model fits well.
3.4.2. Overall significance test
Propose the original hypothesis and alternative hypothesis: \( H_0: \beta_0=\beta_1=...=\beta_k=0 \), \( H_1: \beta_j \) is not all 0. We have obtained \( F=129.7797 \), given \( \alpha=0.1 \), the degree of freedom \( k-1=2 \) and \( n-k=11 \) obtained by looking up the table is \( F_{\alpha}(2,11)=7.2 \), because \( F>F_{\alpha}(2,11) \), it can be fully explained. At a significance level of 0.1, the model rejects the null hypothesis that the overall is not significant. Therefore, the model is overall significant.

3.4.3. Model fitting regression
Comparing the Fitted curve and the Actual curve, it fully shows that the multivariate linear model of the factors influencing the effect of the live broadcaster's live broadcast has a good fit. Now Figure 1.

![Figure 1: Curve of fitted values](image)

4. Conclusions and recommendations
This paper uses multiple regression model analysis methods and first-hand data obtained from field surveys to conduct a stepwise regression analysis on the influencing factors of the live broadcaster's live delivery of goods, and successfully builds a regression model of the live broadcaster's own influencing factors. Through research, we found that the optimization of the live broadcaster’s own image can attract more consumers to consume through live broadcast. At the same time, the improvement of the live broadcast interactive effect of the live broadcaster in the host process can also further stimulate consumers to watch the live broadcast. Consumption behavior occurs in the process. The research results can better help the majority of ordinary live broadcasters to improve based on the live broadcasters' own factors, thereby quickly improving the ability of live broadcasts to carry goods.

In the future, our research will be based on the foundation of this research to subdivide the influencing factors of live broadcasters themselves, especially in the image segmentation of live broadcasters, and continue to carry out in-depth investigation and research. I hope to better help the live broadcaster optimize his image and improve the ability to bring goods.

Acknowledgement
Finally, I would like to thank Dean Guo Hui of Northern Chiang Mai University in Thailand for giving us guidance in the research process of this subject. At the same time, I would like to thank the members of Shandong Vocational College of Business and Technology for their hard work to ensure that the data is true and valid. I would also like to thank my family for giving me a lot of time to concentrate on research.

References
[1] Tan Yuli. The "live broadcast + e-commerce" model from the perspective of communication science—take Jumei Youpin's live broadcast business as an example [J]. Journal of Beijing
Institute of Graphic Communication, 2017 (1): 20−26.

[2] Kim D J. A Study of the Multilevel and Dynamic Nature of Trust in E—Commerce from a Cross—Stage Perspective [J]. International Journal of Electronic Commerce, 2014(1): 11-64.

[3] Jia Xiaofeng. Research on consumer purchase and integration willingness of e-commerce live broadcast platform [D]. Beijing: Beijing University of Posts and Telecommunications, 2019.

[4] Bründl S, Matt C, Hess T. Consumer Use of Social Live Streaming Services: The Influence of Co—Experience and Effectance on Enjoyment [C/OL] / /The 25th European Conference on Information Systems, 2017. https: //www. researchgate.net/publication/317956955_Consumer_Use_of_Social_Live_Streaming_Services_The_Influence_of_Co-Experience_and_Effectance_on_Enjoyment.

[5] Chen Z, Benbasat I, Cenfetelli RT, et al. "Grassroots Internet Celebrity Plus Live Streaming" Activating IT-Mediated Lifestyle Marketing Services at e-Commerce Websites [C/OL] / /International Conference on Information Systems, 2017. https: / /aisel. aisnet.org/ics2017/EBusiness/Presentations/7/.

[6] Du Yanwu. Research on the influence of media characteristics of webcast on users' willingness to continue participation [D]. Chongqing: Chongqing Technology and Business University, 2018.

[7] Chan K K, Misra S. Characteristics of the Opinion Leader: A New Dimension [J]. Journal of Advertising, 1990(3): 53-60

[8] Sussman S W, Siegal W S. Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption [J]. Information Systems Research, 2003(1): 47-65.

[9] Zhong Xujun, Wang Yanrong. Analysis of the reasons for the popularity of webcasting [J]. Southeast Communication, 2016 (9): 88-89.

[10] Mengfei. Research on the influence of opinion leaders on purchase intention under socialized business environment [D]. Nanjing: Nanjing University, 2012.

[11] Zhao Yan. Research on the operating mode of webcasting [D]. Harbin: Heilongjiang University, 2018.

[12] Cai J, Woh D Y, Mittal A, et al. Utilitarian and Hedonic Motivations for Live Streaming Shopping [C/OL] / /The 2018ACM International Conference on Interactive Experiences for TV and Online Video,2018.https: //doi.org/10.1145/3210825.3210837.

[13] Diao Leiyu, Wang Xicheng, Lu Xiaozhen. Research on the relationship between the availability of B2C websites and the purchase intention of online consumers [J]. Journal of Guilin University of Electronic Technology, 2010 (1): 90-93.

[14] Lu Qiang, Fu Hua. The influence of brand social power on purchase intention——Based on the empirical study of "evaluation-emotion-response" theory [ J]. China's Distribution Economy, 2016

[15] Koufaris M, Hampton-Sosa W. The Development of Initial Trust in an Online Company by New Customers [J]. Information & Management, 2004(3):377-397.

[16] Dai Jianhua, Ma Haiyun, Wu Yingying. Research on the influence of the frame effect of online store information presentation on consumer purchasing decision [J]. Chinese Management Science, 2020(3): 152-161.

[17] Faqiang Cui,Cui Faqiang. Quantitative study on factors affecting the price of residential real estate multiple linear regression model[J]. Journal of physics. Conference series,2020,1629(1).