

# Analysis on the Influencing Factors of Network Attention of Canton Fair

Zhenjie Liao

South China University of Technology School of Economics and Commerce  
Guangzhou, Guangdong, Peoples R China

liaozenjie\_2002@163.com

**Keywords:** Baidu index; Network attention; China Import and Export Fair (Canton Fair); Influencing factors

**Abstract.** As an important tool and way of data acquisition, Baidu Index plays an important role in the analysis and research of related issues. Based on Baidu Index System, the data of network attention of Canton Fair in different regions of China are collected, and the influencing factors are studied. The study found that: the network attentions to Canton Fair were influenced by the network developed, regional economic development, the exhibition industry development and population size factor.

## Introduction

With the development of social economy and the progress of science and technology, the penetration rate of Internet has been increasing year by year. Internet is becoming an indispensable tool and means in people's production and life. As of December 2017, the number of Internet users in China reached 772 million, and the Internet penetration rate was 55.8% [1]. Search engine is an important network information platform for the public. According to their interests and needs, people use its retrieval service function to query and obtain information. During this period, huge network search data were recorded by network search tools, which not only provided resources for large data computing, but also provided new research ideas and network survey tools for scientific research. Baidu is the largest Chinese search engine in the world. Baidu Index is a data sharing platform based on Baidu's massive behavior data of Internet users, which provides a strong guarantee and support for the research of network attention. As an important way and tool for data acquisition of network attention, scholars have applied Baidu index to research such as geographical network [2], public opinion review [3], real estate bubble [4], stock market trading [5], journal research [6] and other subjects. In contrast, the network attention analysis of the exhibition as the research object is relatively lacking. The Canton Fair was founded in the spring of 1957. It is held in Guangzhou in spring and autumn every year. It is jointly sponsored by the Ministry of Commerce and the People's Government of Guangdong Province. It is sponsored by the China Foreign Trade Center. It is a comprehensive international trade event with the longest history, the highest level, the largest scale, the largest variety of commodities, the largest number of buyers and the widest distribution in different countries and regions, and the best trading effect. It is known as "China's First Exhibition". In this study, we input the key word "Canton Fair" on Baidu Index and use Python crawler technology to crawl the displayed values one by one on Baidu Index page to obtain the daily data of users' attention of "Canton Fair" from January 2011 to December 2016. Based on the search data of Baidu Index, this paper uses STATA14.0 software tool to analyze the influencing factors of the network attention of the Canton Fair, so as to provide reference for promoting the continuous holding of the Canton Fair and promoting the transformation, upgrading and sustainable development of the Canton Fair.

## Index Selection and Model Construction

**Index Selection.** This research takes the network attention of Canton Fair in 31 provinces (cities, autonomous regions) as the research object, and selects panel data of 31 provinces, autonomous regions and municipalities directly under the Central Government from 2011 to 2016 as the research

sample, establishes a suitable panel data model, and uses STATA14.0 software to empirically analyze the impact of various explanatory variables on the network attention of Canton Fair. By reviewing the relevant literature, it is found that the influencing factors of network attention of Canton Fair involve many indicators. After in-depth exploration and comprehensive consideration of the accessibility and integrity of index data, the selected indicators in this study are listed in Table 1.

Table 1 Influencing Factors and Indicators of Network Attention of Canton Fair

Variables	Categories	Influencing Factor	Symbols
Explanatory variables	Statistical characteristics of social demography	sex ratio (female=100)	X1
	Regional Economic Basis	GDP(RMB 100 million)	X2
		Per capita GDP (RMB)	X3
	Internet Development Level	Internet Pervasion Rate (%)	X4
	Development Level of Exhibition Industry	Exhibition Area	X5
Interpreted Variables	Regional Population Scale	Regional Population	X6
	Network Attention of Canton Fair	Baidu Index of Canton Fair	Y

Data sources: Gender ratio, GDP and GDP per capita are derived from China Statistical Yearbook, 2012-2017, Internet penetration data from the 29th, 31st, 33rd, 35th, 37th and 39th China Internet Development Statistics Report, and exhibition area data from China Exhibition Data Statistics Report(2011-2016).

Explanatory variables: (1) Social demographic characteristics. The index of sex ratio of residents in different regions was used. According to the unified division of China Statistical Yearbook, gender ratio refers to the ratio of male to female population, that is, the corresponding number of male population per 100 female population. Jiang Meihua and other research points out that there are differences in the orientation and degree of attention between men and women on Internet information [7].

(2) Regional economic base. The level of economic development is an important factor affecting the spatial distribution of tourists' network attention [8]. This study chooses GDP and GDP per capita to measure the level of regional economic development, and expects that GDP, GDP per capita and regional Canton Fair network attention have a significant positive impact.

(3) Internet development level. The development level of the Internet is mainly reflected by the number of Internet users and the penetration rate of the Internet. Based on previous literatures and conceptual connotations [9], this study chooses Internet penetration rate as an index to measure the level of Internet development in various regions of China, and considers that the number of Internet users will have an impact on the network attention of Canton Fair.

(4) Development level of exhibition industry. Because of the availability of data, this study chooses exhibition area as an index to measure the development level of exhibition industry in different regions of China. It is expected that the development level of exhibition industry in different regions and the attention of Canton Fair network will have a significant positive impact.

(5) Population size of the region. The network attention index of regional Canton Fair reflects the users' attention to the Canton Fair in the area where the Internet is located based on the IP address of the region. Considering the actual calculation logic of Baidu Index, the study considers that the regional population size is one of the important factors of the difference in the network attention of regional Canton Fair. Therefore, in contrast, the Canton Fair network of regions with a large population base will have a higher degree of concern.

Model Construction. The basic model adopted in this paper is:

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + \beta_5 X_{5it} + \beta_6 X_{6it} + \mu_{it}$$

$Y_{it}$  is the dependent variable,  $X_{it}$  is the independent variable,  $I$  is the first province,  $t$  is the year, parameter  $\alpha_i$  is the intercept term of panel data model,  $\beta_i$  is the coefficient of explanatory variable, and  $\mu$  is the error term.

### Empirical Analysis

Unit Root Test. In order to prevent the problem of "pseudo-regression" of time series, stationarity test of variables is needed first. In this paper, LLC, IPS, ADF-Fisher, PP-Fisher test and other four methods are used to test the unit root of all series in panel data. The results show that all variables pass two or more unit root tests and are stationary series, which can be directly regression analysis.

Descriptive Statistical Analysis. The descriptive statistics of interpreted variables and interpreted variables are shown in Table 2. The observed samples of all variables are  $31 \times 6 = 186$ , and there are no missing samples.

Table 2 Descriptive statistical analysis results

Variable	Observation	Mean	Std. Dev.	Min	Max
Y	186	46216.22	46477.95	1317	304393
X1	186	105.4182	5.101898	95.77	150.71
X2	186	21072.72	16604.79	605.83	80854.91
X3	186	48469.53	22369.98	16413	118198
X4	186	46.87796	12.46196	24.2	77.8
X5	186	327.9224	372.8131	0	1773.78
X6	186	4385.296	2761.061	303	10999

correlation analysis. Thomas's research shows that when the correlation coefficient of independent variables is not more than 0.65, there is no collinearity between variables. In this paper, the Pearson correlation coefficient test of variables is carried out before regression analysis. The test results show that the correlation coefficient is more than 0.65, indicating that there is no serious multi-collinearity between variables, and regression analysis can be carried out. The results are shown in Table 3.

Table 3 Relevance analysis results

	Y	X1	X2	X3	X4	X5	X6
Y	1.0000						
X1	0.2750*	1.0000					
X2	0.7777*	0.0260	1.0000				
X3	0.4053*	0.0649	0.4398*	1.0000			
X4	0.5441*	0.1894*	0.3768*	0.8346*	1.0000		
X5	0.7747*	0.1126	0.7623*	0.5996*	0.6161*	1.0000	
X6	0.5938*	-0.0149	0.8307*	-0.0190	-0.0588	0.4861*	1.0000

F test and Hausman test. As shown in Table 4, the results of F test and Hausman test show that the fixed effect model is more suitable.

Table 4 F Test and Hausman Test Results

Test Type	Statistical	PValue	Model selection results
F Test	40.65	0.0000	Individual Fixed Effect Model
Hausman Test	112.21	0.0000	Individual Fixed Effect Model

Hypothesis Test. Because panel data have the characteristics of time series data and cross-section data, it is necessary to test the correlation of time series and the heteroscedasticity of cross-section data.

#### (1) Intergroup Heteroscedasticity Test

It is assumed that the variance of each province on each section is equal. Because  $2(31) = 3567.25$ ,  $P(X > 2(31)) = 0.0000 < 0.01$ , the original hypothesis that there is no cross-sectional correlation is

rejected at the 1% significance level, that is, there is significant cross-sectional heteroscedasticity.

(2) Sequence Uncorrelation Test

It is assumed that there is no first-order autocorrelation in the sequence. According to Wooldridge autocorrelation test,  $F(1,30)=100.860$ ,  $P(X>F)=0.0000<0.01$  can be obtained. Thus, the original hypothesis that there is no first-order autocorrelation of sequences is rejected at the 1% significant level, that is, there is autocorrelation between sequences.

(3) Cross-section uncorrelated test hypothesis cross-section uncorrelated

Pesaran test of cross sectional independence = 32.729,  $Pr = 0.0000 < 0.01$ , thus rejecting the original hypothesis at a significant level of 1%, indicating that the cross sectional sequence is related.

Estimation of parameters in regression analysis. In order to correct the cross-sectional heteroscedasticity, cross-sectional correlation and intra-group autocorrelation in the fixed-effect panel model, a feasible generalized least squares (FGLS) method is used to correct, which makes the results more reliable. The estimated results are shown in table 5.

Table 5 Estimated results of regression analysis parameters

Y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
X1	66.56098	126.64	0.53	0.599	-181.6489 314.7708
X2	.9661363.	3645389	2.65	0.008	.2516532 1.680619
X3	.0580632	.1322663	0.44	0.661	-.201174 .3173004
X4	422.9189	158.3507	2.67	0.008	112.5572 733.2806
X5	12.44159	5.734381	2.17	0.030	1.202413 23.68077
X6	1.697022	1.499041	1.13	0.258	-1.241043 4.635088
_cons	-19787.72	14649.47	-1.35	0.177	-48500.17 8924.721

Because some variables are not statistically significant, stepwise regression method is used to obtain all significant variables. The non-significant X3 and X1 were eliminated and the FGLS regression model was estimated. After stepwise regression, the remaining variables were statistically significant at the 10% significance level (see Table 6).

Table 6 Stepwise regression results

Variable	Model I	Model II	Model III
X1	66.56(0.53)	73.63(0.51)	
X2	0.966*** (2.65)	0.864*** (3.21)	0.947*** (3.53)
X3	0.0581(0.44)		
X4	422.9*** (2.67)	561.1*** (4.27)	554.4*** (4.27)
X5	12.44** (2.17)	14.55** (2.49)	14.34** (2.52)
X6	1.697(1.13)	2.240* (1.92)	2.060* (1.77)
_cons	-19787.7 (-1.35)	-24900.1 (-1.55)	-17391.2*** (-2.77)
N	186	186	186

t statistics in parentheses

\* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

The example: Tables\*, \*\*, \*\*\*is be include 10%,5%,1%level

Table 7 Multiple Colinearity Test Results

Variable	VIF	1/VIF
X2	18.12	0.055187
X6	10.90	0.091705
X5	5.34	0.187423
X4	3.41	0.293057
Mean VIF	9.44	

Then the multiple collinearity test was carried out: as  $VIF < 10$ , there was no multiple collinearity (see Table 7). Finally, the parameters of the panel model are estimated.

$$Y_{it} = 0.947X_2 + 554.45X_4 + 14.34X_5 + 2.06X_6 - 17391.2$$

## **Analysis of Empirical Results**

Social demographic characteristics. The variable of gender does not accord with the theoretical expectation. From the point of view that gender ratio has no significant impact on the network attention of Canton Fair, in the era of highly developed network, the network has become the basic channel for people to obtain information, regardless of gender. At the same time, it can be inferred that there is no significant difference between men and women in the demand for information on the Canton Fair network.

Regional Economic Basis. The more developed the regional economy, the higher the GDP, the more active the trade exchanges and the more frequent the import and export trade. Naturally, the more attention is paid to the information of the Canton Fair, which is known as the "vane" of China-foreign trade, the more attention is paid to the Canton Fair network. In the eastern region with more developed economy, the attention of the Canton Fair network is obviously higher than that of the central and Western regions. Therefore, the difference of economic development level is one of the factors that cause the difference of network attention of Canton Fair. Generally speaking, GDP and GDP per capita can reflect the level of economic development of a certain region to some extent, but because of the differences among samples (population, land area and other factors), the level of economic development represented by GDP per capita may differ from the actual situation, so the index of GDP per capita in this study failed to pass the significance test.

Internet Development Level. The level of Internet development has a significant positive effect on the network attention of Canton Fair, which is one of the main influencing factors. From the perspective of the three major regions of the country, the economic base of the central and western regions is less developed than that of the eastern regions, the degree of informatization and networking is not high, the Internet penetration rate is not high, and the attention of the Canton Fair network is relatively low. The scale ranking of the network attention index of 11 regions in the East is higher than that of 8 regions in the middle and 12 regions in the west. The higher the degree of informationization and networking, the higher the Internet penetration rate, the more convenient and quicker access to network information, and the higher the network attention of Canton Fair. Therefore, the difference of network development is the factor that affects the difference of network attention in Canton Fair.

Development Level of Exhibition Industry. The level of Internet development has a significant positive effect on the network attention of Canton Fair. In areas with high level of exhibition industry development, exhibitions and economic and trade activities are abundant every year, and economic exchanges with foreign countries are more frequent. Enterprise groups are more willing to participate in the Canton Fair, and the demand for network information of the Canton Fair is greater. This also shows to some extent that exhibitors and professional audiences need the information of Canton Fair. Therefore, the difference of exhibition industry development level is one of the factors that affect the difference of network attention of Canton Fair.

Area Population Scale. Population size is included in the regression equation. It can be seen that the regional population base is an important factor affecting the attention of the Canton Fair network. The network attention of Canton Fair is based on the index data generated by the local residents' network search records, which is closely related to the size of the local population. The results of this study also verify the previous theoretical analysis that the greater the population of provinces (municipalities, autonomous regions) the greater the network attention index.

## **Acknowledgements**

Theoretical Researches on Event(al) Space & Empirical Researches on Festival & Special Event Space by Emphasizing Qualitative Micro-level/Small-scale Analysis (41571132)

## **References**

[1] [http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201803/t20180305\\_70249.htm](http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201803/t20180305_70249.htm).

- [2] Andrew. P. Davidson, YingMiao YU. The Internet and the occidental tourist:Analyses of Taiwan's tourism websites form the perspective of western tourists [J].Information Technology &Tourism,2005,(07):91-102.
- [3] Andrljic, B,Budic, H. Possibilities of tourism market research via the Internet[J].Annals of Daaam & Proceedings.2010,(02):79-80.
- [4] Lexhagen M. The importance of value-added services to support the customer search and purchase process on travel websites[J]. Information Technology and Tourism, 2005,7(2):119-135.
- [5] Skadberg Y X, Skadberg A N, Kimmel J R. Flow experience and its impact on the effectiveness of a tourism website[J].Information Technology and Tourism, 2005,7(17):147-156.
- [6] TD Vila,NA Vila.The Role of the Internet as a Tool to Search for Tourist Information[J].Journal of Global Information Management, 2018, 26 (1) :58-84.
- [7] X. Li, B. Pan and Law R, et al. Forecasting tourism demand with composite search index[J].Tourism management,2017,59:57-66.
- [8] Choi Y. The effect of social and physical distance on the global communication networks[J].Gazette,1994,54(2):163-192.
- [9] Thomas, A. S. and Williams, G. R. A Strategy to Provide Retirement Benefits for International Transferees in a Global Company, Benefits and Compensation International.1991,21,(3):2-7.