Empirical Analysis of the Factors Influencing Online Retail Sales in China based on all-Subsets Regression Model

Ruijie Cheng[1], Xuecheng Wang[2]*
[1]crjpretty666@qq.com, [2]* wxc@outlook.com

North China University of Technology Beijing, China, North China University of Technology Beijing, China

Abstracts—It selected 20 representative indicators from the macroeconomic, demographic and transportation conditions to analyze the factors influencing e-tailing sales through cross-sectional data from 31 Chinese provinces in 2019. By the All-Subsets Regression method, it screened out several variables to set three multiple linear regression equations. All above the models have passed the basic test. The model results show that urban per capita disposable income, the number of cold chain logistics enterprises, and the number of students colleges and universities have significant positive effects on online retail sales. The results of robustness test model with online retail sales of physical goods also show that the conclusion is relatively reliable.

Keywords-online retail sales; E-Commerce; All-Subsets Regression

1 INTRODUCTION

Currently, China is entering an important stage of consumption upgrade, with Chinese consumers' demand for diversified, personalized and high-quality goods growing. The e-commerce industry has further prospered as various traditional service markets, from catering and tourism to office, education and healthcare, have achieved offline integration due to digital empowerment. 2020 saw e-commerce industry had been explosive as the real economy was hit harder by the epidemic. The online retail sales of physical goods hit new highs, while the legalized market environment for e-commerce continued to improve, contributing to the growth of online retail sales. [2]As an important organizational component of the e-commerce industry, China's e-tailing market has continued to improve steadily these years. Online retail sales are an important economic indicator reflecting changes in the domestic e-tailing market as well as economic prosperity, and according to the data from China National Bureau of Statistics, in the first three quarters, online retail sales reached 9.19 trillion yuan, up to 18.5% year-on-year.[1]

Overall, the development of China's e-commerce industry has been transformed from extensive management to legal compliance and high-quality intensive development. 2019 saw the official implementation of the E-commerce Law of China, which establishment and implementation of a number of e-commerce standards, the gradual optimization of the legalized market environment of e-commerce, the

ICEMME 2022, November 18-20, Nanjing, People's Republic of China

DOI 10.4108/eai.18-11-2022.2326933

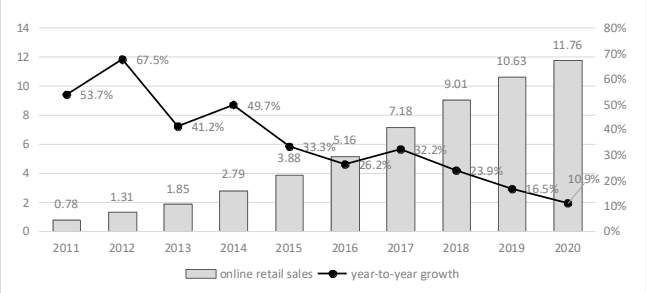


Figure 1. 2011-2020 online retail sales in China

issuance of the Ministry of Commerce's "Guidance on Strengthening the Statistical Monitoring and Analysis of E-commerce", the in-depth promotion of The Ministry and provinces share big data on e-commerce, the process of digital change accelerates, and the national e-commerce statistical monitoring work enters a new stage. 2020, the General Administration of Market Supervision publishes the draft revision of the Anti-monopoly Law, and 2021, the Guide on Anti-monopoly in the Platform Economy is officially promulgated and implemented. Alibaba was fined 18.2 billion and Meituan was fined 3.4 billion, and the e-commerce business environment was further optimized. The General Administration of Market Regulation issued the Interim Provisions on Regulating Promotional Behavior and the Guidance on Strengthening the Regulation of Online Marketing Activities.

Comprehensive domestic and international research is relatively rich in online retail markets, however, few empirical studies about the factors influencing online retail sales. In the relevant literature exploring the factors influencing online retail sales, Lingling Lai and Lifang Peng used two-factor analysis to find that six factors: agricultural products and their prices, seller service quality, e-commerce platform use, trustworthiness, information quality and online store design are the key to influence consumers' willingness to purchase agricultural products online [1]. Huan Xia used a gray correlation model to find that three indicators, the number of mobile phone payment users, the transaction size of B2C market and the transaction size of mobile shopping market, were the most associated with online retailing. [2]

In this paper, we use multiple linear regression models to screen the optimal variables in three areas: income level and prices, social and environmental factors, and logistics and transportation conditions, and explore the factors influencing online retail sales to provide suggestions for the sound development of the e-commerce industry.

2 ANALYSIS OF FACTORS INFLUENCING ONLINE RETAIL SALES

This paper summarizes three main factors that influence online retail sales, namely: macroeconomic factors, demographic factors and logistics and transportation conditions.

2.1Economic development

As a major component of total retail sales of social goods, online retail sales are an important indicator of economic prosperity, and economic prosperity promotes the growth of online retail

sales, while economic recession makes online retail sales shrink. In macroeconomics theory, we use the price level of a region to analyze the economic situation of the region and predict its potential consumption capacity, and its representative statistic is the consumer price index (CPI). The per capita disposable income of a region is the most important determinant of consumer spending, reflecting to a certain extent the purchasing power and consumption level of the region's residents, which affects online retail sales.

Due to the uneven income of urban and rural residents and lagging infrastructure construction and other reasons, the development scale of rural e-commerce is far lower than urban e-commerce, and online sales in rural areas are also far behind urban e-commerce. The urban-rural gap has become a problem for e-commerce enterprises to expand online sales. The urbanization rate is the proportion of the urban population to the total population, and the urbanization rate of a province reflects the urban-rural structure of the province; a higher urbanization rate means that the province has a higher level of urbanization, which may also mean higher online sales.

Based on this, the statistics of urban per capita disposable income (UCDI), per capita disposable income of rural residents (CCDI), consumer price index (urban and rural) (CPI, UCPI), and urbanization rate (UR) for each province in 2019 are selected as explanatory variables in this paper. Data from the National Statistics Office and where not stated below, the data are taken from the National Statistics Office.

2.2Factors of demographic characteristics

Ultimately the factor that determines how high or low online sales are people. Different people have a different impact on online sales. The gender and age of the population all have different effects on online sales. In Beijing, for example, in 2019 gender grouping, women accounted for 64.9% of online purchases, while men accounted for 35.1%. In terms of the distribution of education, highly educated groups are fonder of online shopping. The proportions of online shopping users with junior high school and below, high school, college, and undergraduate and above education were 7.6%, 19.1%, 23.0% and 50.3%, and the percentages of online shopping amount were 3.3%, 11.8%, 20.3% and 64.6%. Data from Beijing Municipal Bureau of Statistics.

In China, young people love online shopping and are the main force of online spending. In 2019, online shoppers aged 19 to 40 accounted for 70.5%, while older people prefer to spend in physical stores, with only 2% of online shoppers aged 51 and above. Data from China Electronic Commerce Association.

While men and women and transgender people have different purchasing preferences when it comes to products sold online, the impact of such preferences is not reflected in the sub-provincial data due to the small differences in gender proportions. In contrast, differences in the age of the population between provinces and the distribution of students in higher education are more pronounced.

Based on this, this paper has selected the percentage of population aged 65 and above (POP), the number of higher education institutions (NEC), and the number of students in colleges and universities (NUC) in each province in 2019 as explanatory variables for the social environment factor.

2.3Logistics and transportation construction

When shopping online, the delivery time of goods is an important factor for customers to choose whether to buy the product or not. Logistics and e-tailing are closely related, and there are significant synergies between the two developments. [3] The booming e-tailing market also provides huge market space and continuous business growth for the logistics industry, on the other hand, express logistics is the key foundation and important link for the continuous rapid growth of the e-tailing market, which also in turn affects the online retail sales. In the logistics sector this paper combines relevant studies and selects express business volume (EBV), freight volume (total freight FTT, rail freight FTL, road freight FTD), year-end actual road length (LCR), and year-end actual road area (ARA) as explanatory variables for each province in 2019. In particular, the development of the cold chain logistics industry is used as a significant explanatory variable in this paper.

The cold chain infrastructure ensures the freshness of fresh produce through refrigeration, storage and transport functions, thus extending the transport time and range. The cold chain infrastructure offers the possibility of transporting fresh agricultural products, blood products, vaccines and biologicals over long distances, enriching the range of products sold online and increasing the number of sales, which affects online sales.

Cold chain infrastructure an integral and important part of the logistics industry, there is a paucity of research related to the impact of cold chain infrastructure on online retail sales. Combining the background knowledge of cold chain logistics and related studies, the following variables are selected in this paper as explanatory variables. Cold storage capacity (CSC) and the number of cold chain vehicles (CCT), direct indicators of the level of cold chain infrastructure in

TABLE 1. SUMMARY OF MODEL RESULTS

models	**M1**	**M2**	**M3**	**M4**
Observations	31	31	31	31
Adjusted R^2	0.933	0.932	0.937	0.923
Overall p-value	<0.0001	<0.0001	<0.0001	<0.0001
BP test p-value	0.385	0.323	0.068	0.318
Explained variables	Y_1	Y_1	Y_1	Y_2
constant	-49.840*** (4.018)	-23.918*** (5.593)	-45.481*** (4.998)	-54.856*** (6.387)
UCDI	3.963*** (0.385)	2.432*** (0.504)	3.681*** (0.426)	4.288*** (0.545)
FTL	-0.082 (0.073)	-0.160*** (0.075)	-	-
FTD	-0.095 (0.171)	-0.032 (0.171)	-0.191 (0.150)	-0.080 (0.192)
NUC	1.246*** (0.171)	-	1.091*** (0.188)	1.192*** (0.241)
NEC	-	1.200*** (0.241)	-	-
CCLE	-	0.522*** (0.519)	0.287* (0.161)	0.259 (0.206)

AIC values	37.088	38.308	34.976	50.181
Global Stat	5.869	3.217	5.818	7.388

Note: a) t-values are in parentheses. ***, **, * and - indicate significant at 99.9%, 99%, 95% and 90% confidence intervals. b) Global Stat values are calculated using the gvlma() function to test whether the OLS hypothesis is satisfied significant at the 5% confidence interval. c) Calculation of variance inflation factors using the vif() function to test for the presence of multicollinearity. d) Heteroskedasticity test using bp test () function, P-value results greater than 0.05, then pass the heteroskedasticity test.

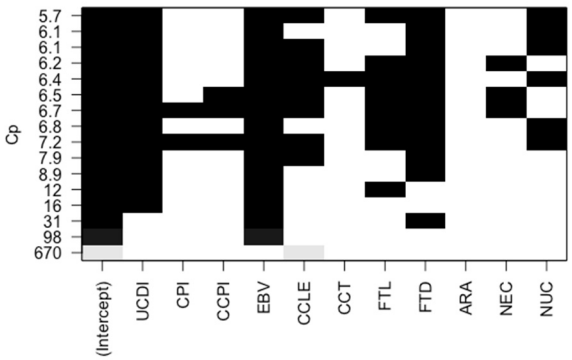
the province. The number of cold chain logistics enterprises (CCLE), which can measure the level of activity of the cold chain logistics industry in the province. Fish production per capita (FPC): Fish products require refrigerated or frozen transportation for preservation, and fish production can reflect the demand for cold chain infrastructure from the population. And the number of fresh food e-commerce enterprises (FEC) reflects the degree of demand for cold chain infrastructure by e-commerce. Except for aquatic products production from China Fishery Statistical Yearbook 2020, the data are from Zhong Guancun Green Cold Chain Logistics Industry Alliance.

3 EMPIRICAL RESEARCH

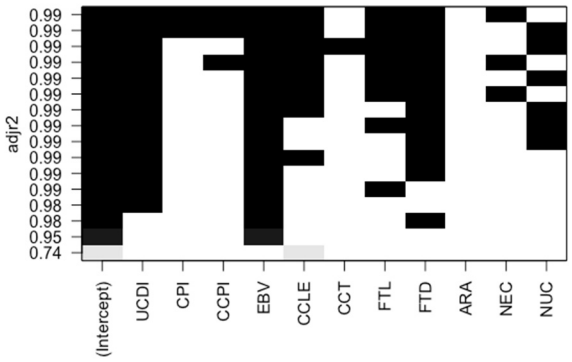
3.1Construction of the model

Multiple linear regression models were developed based on data characteristics:

$$Y_i = X_i\beta + \varepsilon_i$$



(a). results according to the value of adj^2



(b). results according to the value of Cp

Figure 2. All-Subsets Regression results

where Yi is the 2019 online retail sales by province and city, and Xi is the 20 explanatory variables mentioned above. β is the coefficient of the explanatory variable and Ɛi is the residual term.

3.2Parameter estimation

To find the most influential factor on network sales, we performed All-Subsets Regression (All Possible Regression) on the logarithmic values of the variables. The picture of the results generated using R language is as follows.

According to the relevant statistical knowledge, the larger the adjusted R^2 value is, the better, while the Cp value needs to be close to the number of variables. From the figure, when the Cp value is 6.1, the model should contain UCDI, CCLE, FTD, NUC and the intercept term five variables to build the model M1. When the Cp value is 6.2, the model should contain UCDI, CCLE, FTL, FTD, NUC and the intercept term, and model M2 is built. When the Cp value is 6.8, the model should contain UCDI, FTL, FTD, NUC and the intercept term, and model M3 is built. In the following table, UCDI is urban per capita disposable income, FTL is rail freight volume, FTD is the volume of road freight, NUC is the number of undergraduate and college students, NEC is the number of higher education schools, and CCLE is the number of cold chain logistics enterprises.

The model M5 was tested using Y_2 (online retail sales of physical goods) as the dependent variable, using the same methodology as above. The result show that our models are robust. The results are shown in TABLE 1.

4 CONCLUSION

Firstly, there is a significant positive correlation between online retail sales and the number of cold chain logistics enterprises. The greater the number of cold chain logistics enterprises in a province, the better the cold chain infrastructure in that province, the better the cold chain transport services provided, and the more willing consumers will be to buy products delivered

through cold chain logistics online, and online sales will increase. For agricultural products, cold chain infrastructure can preserve freshness thus improving product quality [4], increasing product prices and avoiding grain cheapness; on the other hand, cold chain infrastructure can extend preservation time, expand the sales radius of agricultural products and other products, and extend sales time, thus providing the possibility for residents to buy more fresh agricultural products online. Residents' demand for high quality cold chain food will increase, according to Zhigao Zheng and others [5]. The increase in spending on fish products also boosts the social demand for cold chain infrastructure, further enhancing the impact of cold chain infrastructure on online sales. Simultaneously, the cold chain logistics infrastructure is also very significant for pharmaceutical products because special products such as vaccines, biological products and blood products have special temperature requirements for transportation and preservation [6]. The scale of China's pharmaceutical B2C e-commerce will be close to RMB 150 billion in 2020, and major e-commerce platforms are actively laying out their e-commerce business, and pharmaceuticals may become a new breakthrough for major e-commerce companies to increase their online sales.

Secondly, high correlation between online retail sales and per capita disposable income of urban residents. The higher the per capita disposable income of urban residents, according to Fang Fuzian [7], the per capita disposable income of Chinese residents is highly correlated with per capita consumption expenditure, and the consumption function of Chinese urban and rural residents is relatively stable during 1995-2005. As a major component of residents' consumption, e-tailing accounts for nearly one-fourth of total retail sales of consumer goods, and it is self-evident that it is influenced by disposable income.

Thirdly, online retail sales are highly correlated with the number of students colleges and universities and the number of higher education institutions. In 2019, the group of undergraduate students is a generation growing up under the wave of the Internet, they are naturally close to the Internet, their lives are closely related to the Internet, and their consumption methods will choose to buy online, young people are also the most purchasing power on the Internet. young consumer groups aged 19-25 account for 11.9% of the total number of online shopping users, but their spending power should not be underestimated.

Lastly, based on the conclusions above, this paper gives the following recommendations. Strengthen the construction of cold chain infrastructure and establish a sound system of cold chain infrastructure services. Due to the increasingly high cost of labor in China, coupled with the fact that many operational aspects of the cold chain are in a cold and harsh environment, automation and intelligence of the cold chain enterprises are crucial. However, there are many problems with the existing cold chain infrastructure in China, such as unbalanced regional development, unbalanced supply and demand, unreasonable structure, etc. The gap with developed countries is large. Therefore, cold chain infrastructure is also in urgent need of accelerated upgrading. Focus on the pharmaceutical e-commerce sector and promote the professional development of the pharmaceutical cold chain distribution industry. Rational planning of cold storage layout to improve the utilization of cold chain infrastructure resources. Due to the uneven layout of the cold chain infrastructure, many products that require cold chain transportation can only be transported by ordinary vehicles. However, in developed coastal areas, there are many unused cold storage facilities. The layout of cold storage is in urgent need of improvement. In addition, cold storage capacity, ice production capacity and freezing capacity

can be enhanced to provide a continuous and constant low-temperature storage environment for cold chain products, while reducing idle inventory.

Acknowledgements. This work was financially supported by Beijing Social Science Fund (21JJC021) and Beijing Urban Governance Base Project(21CSZL22).

REFERENCE

[1] Lai Lingling, Peng Lifang. Research on the factors influencing the willingness of agricultural products to purchase online based on two-factor analysis[J]. Price Theory and Practice,2020(12):99-102+163.

[2] Xia Huan. An empirical study on the influencing factors of online retailing in China - based on grey correlation analysis model[J]. Research World,2020(09):36-41.

[3] Liu Dan, Lu Weiwei. The synergistic development path of China's e-commerce industry and express delivery industry[J]. Technology Economics,2014,33(02):45-49.

[4] Yang Han, Qian Jianping, Yu Qiang Yi, Wu Wenbin. Forecast of China's cold storage demand in 2030 against the cold chain circulation rate of developed countries[J]. China Agricultural Information,2020,32(02):65-75.

[5] Zheng Zhihao, Gao Ying, Zhao Yinyu. The impact of income growth on urban residents' food consumption patterns[J]. Economics (Quarterly),2016,15(01):263-288.

[6] Lin Zhenqiang. Current situation and trend of pharmaceutical cold chain development in China [J]. Logistics Technology and Applications,2021,26(S1):42-45.

[7] Fang Fuzian. A study on the causes of insufficient consumer demand in China--based on data from urban and rural China by province[J]. China Social Science,2009(02):68-82+205-206.