The Impact of COVID-19 on Logistics in Context of Regression Analysis

Zhengyang Qiu[1, *, †], Moziyan Zhang[2, †] and Ruoye Zhang[3, †]

[1]Nanchang No.2 high school, Nanchang, China

[2]Shenzhen Hankai mathmatic and science school, Shenghen China

[2]NanJing Jinling High School Alevel Centre, Nanjing, China

*Corresponding author: 100453@yzpc.edu.cn

†These authors contributed equally

Abstract. Since 2020, the epidemic has significantly affected China's development in several different areas. As a result of the epidemic's shifting circumstances, numerous industries have seen diverse transformations. On the other hand, the logistics sector has grown as a result of the epidemic, for instance, as consumers rely increasingly on e-commerce platforms for their purchasing. Still, it has also been impacted by national legislation. This study examines the relationship between several logistics indicators and the number of new cases added to the epidemic in order to provide an overview of the logistics industry's growth process in the context of the epidemic (reflecting the changes in the epidemic). In this research, some indicators that mainly reflect the overall efficiency of logistics during the epidemic (e.g., LPI, Storage index, E-Logistics index, and Freight rate, as well as the number of new and accumulated confirmed cases of the epidemic), were collected and collated to represent the trend of the epidemic, and correlation and comparative analysis were conducted to obtain the specific consequences of COVID-19 on the logistics industry. A typical example is isolation control which results in a dramatic decline in the freight volume of road transportation. These results can help the logistics industry to determine the development direction and plan under the epidemic and find the development trend of the overall market through this paper.

Keywords: covid-19; LPI; linear regression; Logistics.

1. Introduction

While the global epidemic has rebounded, the overall epidemic situation in China remains stable, but the risk of the spread of the outbreak in individual places still exists. Looking back at the past two years of the epidemic, the New Crown outbreak developed into a public health event during the Chinese New Year in 2020, with extended holidays and delays in resuming work across the country, causing a significant impact on a few domestic industries [1-3]. It will also inevitably impact China's employment growth, consumption growth, and foreign investment expectations in the short term, bringing about a downward spiral of short-term economic growth expectations. The closure of cities, roads, and residential areas in areas where the epidemic is more severe has resulted in restrictions on travel, a significant impact on rail traffic, air transport, and road passenger transport, as well as a significant reduction in the frequency of international flights, resulting in substantial economic losses to the transport sector. On August 29, 2022, 58,463,327 people in total have been confirmed in China [4]. The following are some previous studies on the epidemic's impact on logistics. Supply chain disruptions and blockades have affected logistics companies. However, not all areas will be involved in the same way - as consumers choose to buy necessities online, companies serving e-commerce will see more and more activities. In contrast, companies operating in other fields (such as cars and consumer goods) will see a downturn. A mitigating factor is that record low oil prices should relieve transportation operators. Overall, uncertainty will put downward pressure on income.

The impact on small businesses is profound. Small truck businesses are hit hard because they often do not have any backup, recovery, or intermittent business plans, i.e., a lack of technology and tools. Even top companies are being strongly affected. Road freight has been hit hard. Influenced by the epidemic, the volume of road cargo transportation has dropped sharply [5]. The epidemic's impact is

large-scale, and some inter-provincial and municipal road transportation has been cut off. In addition, drivers have been delayed in rework. After the holiday, the recovery rate of highway logistics is meager, and the national expressway freight volume has shown a steep decline [6, 7].

At present, epidemic prevention and control are still in a critical period. During the outbreak of COVID-19, many people in isolation at home reported that they needed to wait twice as long as usual for items bought through online shopping, which reflects a decline in the efficiency of the logistic system. The paper will explore the association between the efficiency of the logistic system and epidemic with the confirmed case as a proxy through some logistics indexes, e.g., LPI and Warehousing index. It then evaluates the correlation coefficient and analyzes the reasons for the correlation between cases and each index. A description of future estimation follows this for the correlation between each index and the epidemic. Afterwards, limitations of the research and future outlook are also claimed in the paper to provide valuable information and effective methods for a prospective study of this field.

2. Data & Method

2.1 Data

Three dependent variables y and one independent variable x were used for this study. All of the data are from January 2020 through July 2022. The number of indigenous epidemic growth, which is a reflection of the evolution of the epidemic, is the independent variable x that we chose. The China Warehousing Index (y), the following dependent variable we chose, reflects the operations of the storage industry as well as the supply and demand circumstances and trends of significant commodities on the domestic market. The China Warehousing Index, which is derived as "China Warehousing Index = ending inventory×30% + new orders×25% + workers×25% + average inventory cycle times×20%," is a composite of four weighted indices: ending inventory, new orders, average inventory cycle times, and employees. The e-logistics index (ELI), a comprehensive assessment index system that depends on the e-commerce logistics platform to rapidly and accurately represent the operating circumstances and trends of e-commerce logistics is the second dependent variable we chose [8-11]. Nine sub-indices comprise the index: overall business volume, rural business volume, inventory turnover, logistics timeliness, performance rate, satisfaction, actual load rate, cost and personnel, and a synthetic index. The total business volume, logistical timeliness, performance rate, satisfaction rate, real load rate, cost, and staff are the seven sub-indices that make up the weighted synthesis known as the synthetic index. The third dependent variable is the Logistics Prosperity Index (LPI), composed of 12 sub-indices and one synthetic index. The synthetic index, which separates economic strength from weakness at 50%, is a weighted synthesis of five indices: total business, new orders, employees, inventory turnover, and equipment utilization rate. If the synthetic index is above 50%, it indicates that the logistics industry is expanding, while it means contracting below 50%.

2.2 Correlation coefficients

Pearson's correlation coefficient is able to illustrate the strength of linear association between variables [8]. The calculation formula can be given as follows:

$$\rho_{x,y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \quad (1)$$

where X is the number of confirmed cases, and Y is the value for each index (Warehousing index, Storage Index, E-Logistics index, Freight Rate Index (Road), and LPI). μ_X and μ_Y are the mean of X and Y, respectively. σ_X and σ_Y are the standard deviation of X and Y, respectively. However, before doing the hypothesis testing, the experimental data need to satisfy the following conditions: firstly, it should be confirmed that the data are bivariate normally distributed. Ssecondly, the data gap should be small, which means there is no outlier; and finally, it should be ensured that the samples

are independently sampled. Based on the estimation of values, the Pearson correlation coefficient takes advantage of covariance and standard deviation between variables to determine how closely variables are associated. From -1 through 0 to +1, a larger absolute value of the correlation coefficient indicates a stronger linear correlation. A positive sign in front of a coefficient means that confirmed cases are positively correlated with those indices, indicating that the index increased as the number of cases rose, or vice versa. Pearson's formula represents measures of difference between the observed value and means value for x and y, respectively.

2.3 Linear regression

Simply put, there are many points in the coordinate system. The purpose of Linear Regression is to find a line so that all the points are on or around the line. This is Linear Regression.When we use only one x to predict Y, it's unary linear regression, which means we're looking for a line to fit the data. For example, I have a scatter plot of a set of data, where the abscissa represents the amount of advertising spent and the ordinate represents the amount of sales. Linear regression is about finding a straight line and making that line fit the data points in the graph as well as possible. In linear regression model, Multiple R is refers to the Pearson correlation coefficient, which is used to measure the correlation between two variables. R square is usually used as metric to describe how well the data fit the model. When the observations are all on the regression line, the sum of squares of the regression differences explains 100% of the total sum of squares, R Square=1. When the regression line is parallel to the X-axis (the mean line coincides with the regression line), the residual sum of squares explains 100% of the total sum of squares, R Square=0. Multiple R refers to Pearson's correlation coefficient, which measures the correlation between two variables. Based on Excel, we can find the coefficients of each model, and then plug the coefficients into the formula to get the final regression equation.

3. Results & Discussion

3.1 Correlation analysis

Table 1 illustrates the correlation between logistics variables, newly confirmed cases, and accumulated confirmed cases in China from Jan.2020 to July.2022. Compared to accumulated confirmed cases, new confirmed cases showed strong negative correlations for the Warehousing index, E-commerce logistics index, and Logistics performance index, which means an increase in new cases will reduce these indexes. All three indices were negatively affected by new cases, especially the LPI, to varying degrees. Nevertheless, accumulated confirmed cases positively affect the freight index in terms of road transportation.

Table. 1 Correlation coefficient between logistical variable and New and accumulated cases

	Logistic Variable	New cases	Accumulated total cases
Pearson Correlation Coefficient	Warehousing index	-0.7748	0.044284838
	E-Logistics index	-0.81679	-0.3807905
	Freight index (Road)	0.003698	0.783806845
	Logistics prosperity index (LPI)	-0.81712	0.190524285

LPI as an overall indicator of the efficiency of the logistics sector, is closely related to the level of development of the national logistics system [9]. When new cases rise, the dramatic decline in LPI illustrates the breakdown of logistics systems under the epidemic. It also represents the inefficiency of the logistics sector due to policy changes brought by new cases, such as the lockdown of Wuhan. In addition, affected by lack of employment and inconvenient transportation policies, in terms of e-commerce logistics and warehousing logistics, the growth of new cases also poses many challenges: the former is an indicator reflecting the logistics situation of e-commerce, while the latter is an

indicator reflecting the supply and demand situation and change trend of major commodities in the domestic market. With the rapid spread of COVID-19, social distancing policies and imposed restrictions on logistics and human movement have been adopted as methods of controlling the epidemic by many countries [10]. Under that circumstance, the overall efficiency of transportation will diminish, which will negatively impact the logistics system and each index.

In contrast, the continuous increase in the number of accumulated cases, which represents a continuation of the epidemic, has led to an increment in logistics freight rates since the prolonged epidemic caused a change in the supply-demand relationship in the market. On the one hand, the long-term quarantine of the epidemic has led to a growing demand for commodities, as people obtain necessities and food at home by take-out or delivery; on the other hand, policies under the epidemic have resulted in a shortage of employment or the temporary closure of factories, leaving the market under-supplied. As a result, freight rates will keep rising as the outbreak continues.

3.2 Regression analysi

When the storage index is taken as Y, and the number of new epidemic cases is taken as variable X, the model can be expressed as $Y=0.5215347+ 0.39^{-6}X$ by using the data from Table. 2 and 3.

Table. 2 Regression statistics of the four models.

Regression statistics	Model 1	Model 2	Model 3	Model 4
Multiple R	0.77479853	0.81678691	0.00369838	0.81711956
R Square	0.60031276	0.66714086	1.3678E-05	0.66768438
Adjusted R Square	0.58653044	0.65566296	-0.0344686	0.65622522
standard error	0.02045	1.92595682	0.0155135	0.0332325

Table. 3 Analysis of function for the four models.

	Coefficients	standard error	t Stat	P-value	Lower 95%	Upper 95%
Intercept 1	0.52153471	0.00399128	130.668614	9.8279E-42	0.51337163	0.52969779
Variable 1	-1.393E-06	2.1104E-07	-6.5997525	3.1178E-07	-1.824E-06	-9.612E-07
Intercept 2	109.518316	0.37589373	291.354466	7.97E-52	108.749527	110.287105
Variable 2	-0.0001515	1.9875E-05	-7.6239059	2.0937E-08	-0.0001922	-0.0001109
Intercept 3	0.99784737	0.00302781	329.561066	2.2387E-53	0.99165481	1.00403993
Variable 3	3.1886E-09	1.601E-07	0.01991654	0.98424635	-3.242E-07	3.3062E-07
Intercept 4	0.53740824	0.00648607	82.8557747	5.1846E-36	0.52414274	0.55067374
Variable 4	-2.618E-06	3.4295E-07	-7.6332455	2.0439E-08	-3.319E-06	-1.916E-06

As given in Table. 2 Model 1, the score represents the value of R^2 as 0.60031276, which is more than 50%, which shows that the regression line fits the observed values well. Besides, for multiple R, the value is 0.77479853, which shows the independent variable X and the dependent variable Y are closely correlated. For the standard error, it is used to measure the degree of fit and to calculate other statistics related to regression. The smaller the value, the better the degree of fit When the e-commerce logistics index is taken as Y, and the number of new epidemic cases as variable X, one can find out the tremendous impact of the new epidemic cases on the e-commerce logistics index through regression statistics. From the Table 3.2, one can see the coefficients for intercept and X variable; by using these two values, here is the model of this linear regression model Y=109.518316-0.0001515X. As illustrated in the Table. 2, the value of R^2 is 0.66714086 of Model 2, which shows that the regression line fits the observed value well. For multiple R, the closer the absolute value is to 1, the stronger the correlation. In this model, 0.81678691 that really close to 1, so the independent variable X and the dependent variable Y are strongly correlated.

When the freight rate index is taken as Y, and the daily new cases are taken as variable X, we can find from the Table.2 Model 3 that neither R square nor Multiple R values can clearly show the linear relationship between them and fail to fit well, cause the value for R square and Multiple R all nearly to 0. Therefore, the daily new cases cannot be the variable of the change in the freight rate index. When the logistics industry climate index is used as the total Y variable, and X is the daily new cases,

by using the data from Table.3.4 this linear regression model equation can be show as $Y=0.5374082+2.618^{-6}X$. About the regression statistics, it can be seen from the Table2 Model 4 below that the value of Multiple R is 0.81711956, and the R square is 0.66768438, which is higher than 50% and really close to 1. It shows that the regression line fits the observed values well. In addition, the standard error that is extremely close to 0 indicates that the model fits well.

3.3 Applications & Suggestions

The goal of this study is to increase logistical effectiveness in an epidemic condition and to further identify the development trend by examining the correlation between the index and the increased population. These results can be applied to the government's management of logistics. The government may collect the expected various logistics indices using the linear regression findings reported in this research in order to develop the most suitable logistics control strategy when there is a new population in a certain location. It is particularly relevant during a period of high epidemic occurrence, and it can assist the government in determining the best preventative measure to take in order to successfully reduce the delays in logistics and the resulting scarcity of supplies. The express industry will facilitate the company's rapid-growth success in academic departments. With the risk control measures of the epidemic policy and the implementation of quarantine control, transport logistics companies' costs are rising, indicating an increase in freight rates. Under that circumstance, combining an e-commerce platform or brick-and-mortar company and logistics company into an integral system reduces the gap in goods handover and saves time. Besides, when the destination of the goods is confirmed, the logistics company can design the mostefficient route by mobilizing the resources of whole system to save time and money.

4. Limitation & Future outlooks

First off, because China bases its strategy for combating outbreaks on a 14-day cycle, the number of new local (Chinese) confirmed diagnoses is chosen monthly in this paper and does not take into more detail the daily variation in the number of confirmed cases during the month, which also affects the severity of the outbreak. Second, the number of new confirmed cases and the various numbers of close connections have prevented us from accounting for the number of transmissions and the number of people sequestered, which will impact the scope of the closure on the ground. Thirdly, because we haven't broken down the epidemic data by location if all the new illnesses are concentrated in one area, that area's logistics will be affected. Still, different control strategies will affect logistics to varying degrees if they are dispersed across the nation. Additionally, the location of the city or region under management will affect how frequently logistics indicators change. For instance, because Shanghai is the third-largest logistics hub in China, its dominance over a single region in April 2022 will significantly impact the logistics indices of the entire Yangtze River Delta region. Fourth, we neglected to account for seasonal variations and modifications to the logistics at various nodes when performing the data analysis. For instance, the LPI index is exceptionally high in October and November, which is primarily attributable to the fact that e-commerce platforms run significant deals from October to November and encourage consumers to shop. These two things will both affect logistical indicators.

Therefore, in the future, we want to make up for and improve our weaknesses. In order to address the issue that the range of data on the number of new persons joined to the epidemic is too large, we will first change the 30-day period to 14 days, which will also make it simpler to measure the intensity of the epidemic and create a more accurate analysis. We will also examine the city's precise control regulations, the number of close contacts and containment zones, and how the lack of a thorough breakdown of isolated persons and tight links will affect logistical control. Since the confirmed number of cases and the specific number of new cases in each region are unavailable, we will divide China into multiple logistics regions and add up the new number and distribution in each region to arrive at a more accurate evaluation. The extra factors that will be introduced simultaneously for

various areas include the number of high-risk zones in the epidemic, the inventory of warehouses in controlled areas, the number of staff, etc. Geographic location and the nation's monthly logistics rating are additional factors. Fourth, in order to account for variations as well as seasonal and promotional categories, we will run Multiple Linear Regressions on the data. Additionally, we intend to choose a sample for in-depth research and analysis. We will select a single city to study the growth and alterations of logistics from the commencement of the disease in 2020 in order to concentrate and narrow our investigation.

5. Summary

The evolution of the logistics sector concerning the epidemic is the subject of this study, which also examines the connection between the number of newly confirmed epidemic cases and the major logistical indicators. The paper takes advantage of the correlation between the number of confirmed cases and several logistic indexes to analyze the impact of COVID-19 on the logistics system. With the rapid spread of the epidemic, LPI, storage index, and other logistics indexes are all affected because of the lack of transportation and staff caused by the epidemic, which indicates that COVID-19 negatively influenced the epidemic in the logistics industry. From the above training and testing of the linear regression models, it was found to be an effective way to find out the impact of the new epidemic trend on the logistics climate index, freight rate index, warehousing index, and e-commerce logistics index. As the results, we can find for most of the indices, and linear regression models are an excellent model to predict changes in the overall trend of the index as the number of confirmed cases increases. However, there is still some uncertainty, so it still needs to be tested and checked for some time. The forecast trends can help logistics companies identify areas they need to strengthen and change, as well as plan freight rates and hiring to help them better face the pandemic. The information in this document does, however, have some restrictions. For instance, the statistics on the number of newly confirmed cases have a too-long time horizon and do not account for the change that occurs every 14 days (because many policies of the Chinese government are divided according to 14 days). Future updates to our data should improve its accuracy and detail. In order to minimize the losses brought on by the epidemic for logistics business, it is the goal of this research to determine the precise relationship between the epidemic and logistics development. Overall, these results offer a guideline for the impact of COVID-19 on logistics.

References

[1] Ciotti M, Ciccozzi M, Terrinoni A, et al. The COVID-19 pandemic. Critical reviews in clinical laboratory sciences, 2020, 57(6): 365-388.

[2] Velavan T P, Meyer C G. The COVID‐19 epidemic. Tropical medicine & international health, 2020, 25(3): 278.

[3] Yuki K, Fujiogi M, Koutsogiannaki S. COVID-19 pathophysiology: A review. Clinical immunology, 2020, 215: 108427.

[4] Trackingthe Epidemic. Weekly report of China CDC Retrievde from: https://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm.

[5] Cheng Qiyun, Sun Caixin, Zhang Xiaoxing, et al. Short-Term load forecasting model and method for power system based on complementation of neural network and fuzzy logic. Transactions of China Electrotechnical Society, 2004, 19(10): 53-58.

[6] Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.

[7] Barnhart C, Krishnan N, Kim D, et al. Network design for express shipment delivery. Computational Optimization and Applications, 2002, 21(3): 239-262.

[8] Hauke J, Kossowski T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaestiones geographicae, 2011, 30(2): 87.

[9] Beysenbaev R, Dus Y. Proposals for improving the logistics performance index. The Asian Journal of Shipping and Logistics, 2020, 36(1): 34-42.

[10] Guo H, Liu Y, Shi X, et al. The role of e-commerce in the urban food system under COVID-19: Lessons from China. China Agricultural Economic Review, 2020.

[11] Twinn C I, Qureshi N A, Conde M L, et al. The impact of COVID-19 on logistics. 2020.