Using linear regression model to investigate the relationship between people's average expenditure on food consumption and GDP per capita in China under pandemic across cities

Zhuoyu Zhang

Anderson Serangoon Junior College, Singapore

Abstract. The Food and Beverage (F&B) industry in China has always been a hot industry and it has experienced a continuing rise in revenue across years before the pandemic hits China's cities. Given the large scale of the industry, its performance is closely related to people's everyday lives. However, ever since China entered the pandemic and placed restrictions on citizens' mobility, significant drops in the revenue have been seen in the F&B industry. A considerable number of physical restaurants had to close, and people's food consumption has been largely affected. Moreover, in view of the different population size of provinces, it is also important to analyse the relationship between average expenditure on food consumption and GDP per capita across time. Therefore, it is important to analyse the factors related to the F&B industry and to understand the extent of damage on the F&B industry by the pandemic. In this paper, we recognized the large extent of damage and measured it using the linear regression model, after proving the strong linear correlation between F&B industry revenue and the GDP of the city. We also analysed the linear relationship between average expenditure on food consumption and GDP per capita. The Analysis of Covariance Test is used to measure the closeness of this relationship between the years 2019 and 2020. Due to the complexity of the F&B industry, we considered another possible factor; the average age of the workforce, and introduced the multiple linear regression model.

Keywords: Expenditure, Food and Beverage (F&B) industry, GDP, linear regression model, average age of workforce.

1. Introduction

Ever since the covid-19 breakout, almost every government has established rules and regulations to curb the spread of the virus, such as social distancing and home isolation. Some rules have greatly restricted people's movement in order to alleviate and control the situation. For example, in China, some cities experienced periods of lock down. There are restrictions on the dates that workers are permitted to return to work and people have to stay at home for a long time and can only go out to purchase and stock up goods on permitted days. As a result, the restaurants have suffered considerably and the supply chain in the F&B industry could have been broken. Such impacts greatly affected those working in the industry. Many restaurants faced a sharp decrease in the dining in service and continuously increasing cost of production of labour, especially for the middle and small-sized restaurants, making it even more difficult for them to survive under the damage of the pandemic.[1]

However, as China gradually kept the pandemic situation under control, the F&B industry entered a recovering phase with restaurants beginning to resume some normal activities around April 2020 (Figure 1). Although the restaurants were still earning less than before the pandemic, many expanded their business into online delivering. Together with the easing government measures and the general publics' increasing demand for good tastes, there have been forms of revival of the F&B industry in China. In view of these results, I will investigate, in this essay, the relationship between the revenue in F&B industry and the GDP per capita across cities in China since 2016, using a linear regression model.

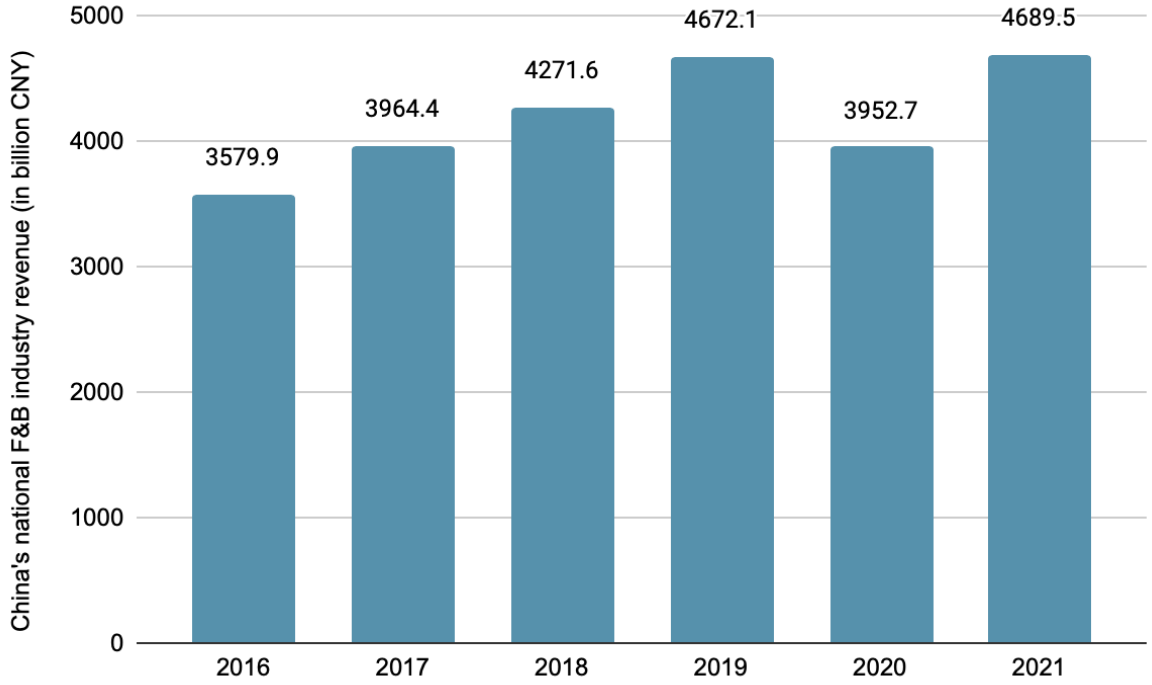


Figure 1. National Food and Beverage industry revenue (in billion CNY) in recent years in China. [2]

2. Introduction of Linear regression model

The linear regression model analyses the relationship for a set of bivariate data, particularly helpful in investigating their linear correlation. Given a set of scattered points, the best fit line is the least squares regression line, such that the sum of the squares of the vertical distances between the line and the points ($\sum e_i^2$, where e_i is the error) is a minimum.

As shown in Figure 2, this is an example of error and best-fit line in the linear regression model:

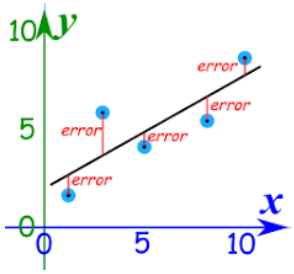


Figure 2. The black line is the best-fit line and the distance from the blue dot to the black line indicates the error.

The Pearson's product moment correlation coefficient, r is a measure of the fit of a scatter diagram to a linear model. It indicates the strength and direction of a linear relationship between two variables and the formula for r is:

$$r = \frac{\Sigma(x-\overline{x})(y-\overline{y})}{\sqrt{[\Sigma(x-\overline{x})^2][\Sigma(y-\overline{y})^2]}}$$

If r is close to 1, i.e. $0.8 \leq r \leq 1$, there is said to be a high positive linear correlation between the two variables x and y.

The best-fit line (y=mx+b) is given by:

$$m = \frac{N\Sigma(xy)-\Sigma x\Sigma y}{N\Sigma(x^2)-(\Sigma x)^2} \quad b = \frac{\Sigma y - m\Sigma x}{N}$$

3. Analysis of data

As illustrated in Figures 3 and 4 are the revenues from the Food and Beverage industry in 31 provinces (including 4 municipality cities) in China in 2019 and 2020 respectively. The least-squares regression line is drawn according to the data. According to calculation: $r^2 = 0.844,0.826$, $r = 0.9187,0.90885$ close to 1. This shows that there exists a strong positive linear correlation between the F&B industry revenue and the level of GDP in cities.

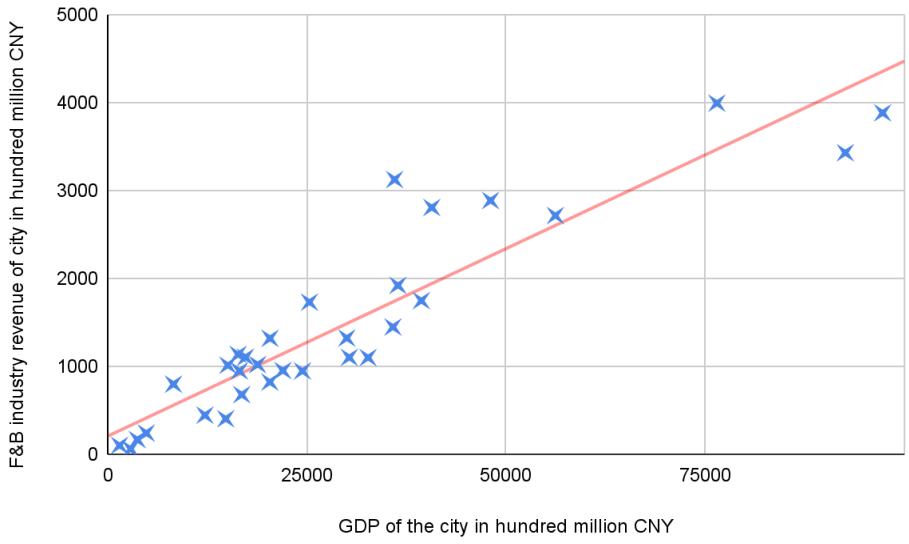


Figure 3. Food and Beverage industry revenue (in hundred million CNY) of cities in China in 2019 (financial year). [3]

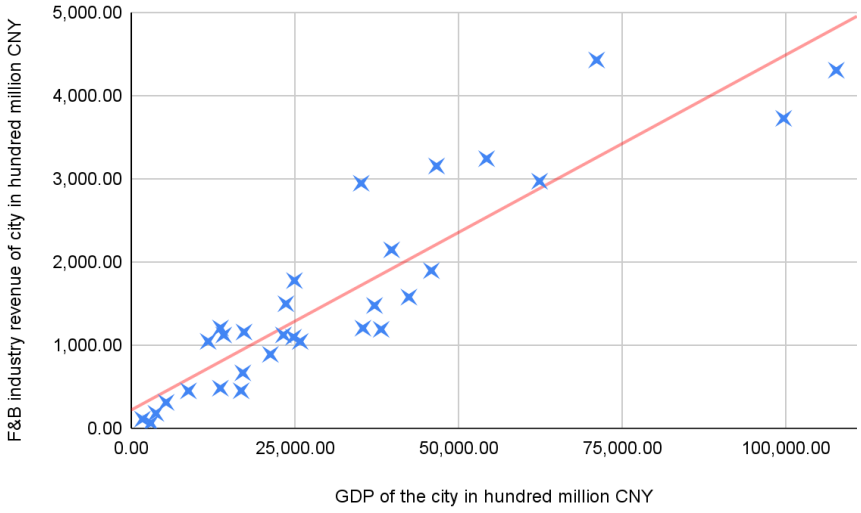


Figure 4. Food and Beverage industry revenue (in hundred million CNY) of cities in China in 2020 (financial year). [3]

In 2019, the calculated best-fit line is $y = 0.0427x + 206$

In 2020, the calculated best-fit line is $y = 0.0427x + 221$

To test the significance of the difference between the two sets of data, we conduct an Analysis of Covariance Test (ANCOVA test). It examines the influence of an independent variable on a dependent variable while removing the effect of the covariate factor. According to the test, as shown in Figure 5, the p-value is 0.863093. This implies there is no significant difference in the correlation between people's expenditure on food consumption and GDP in 2019 and 2020.

Test for homogeneity of regressions:

Source	SS	df	MS	F	P
between regressions	7220.42	1	7220.42	0.03	0.863093
remainder	13508323.69	58	232902.13		
adjusted error	13515544.11	59			

Figure 5. The ANCOVA test result for the F&B industry revenue in relation to GDP across cities in 2019 and 2020.

However, this is not in line with the above-mentioned impact of the pandemic on the F&B industry. One possible reason for this is that in some areas, the local government may have implemented effective precautions so that there could be few or no positive cases and people could continue their way of life. In addition, provinces and cities have adopted fiscal policies to encourage population consumption and to support firms. This could stimulate the local economy, such as lowering taxes and increasing government expenditure. [4]

According to my 2020's model, I predict that the revenue for the F&B industry in 2021 is 4905.7 billion CNY, given the real GDP of China is 114.37 trillion CNY in 2021. In comparison to the real revenue of 4689.5 billion CNY, my predicted value only differs by 4.61%. Hence, I think this model well fits the linear correlation between the F&B industry revenue and the GDP in China.

4. Comparison of average consumption expenditure in the first-tier cities before and since China entered the pandemic

Figure 6 illustrates the average expenditure on food consumption per capita in the top 10 cities in China from 2015 to 2021 (in financial year). The best-fit line is calculated based on the data from 2016 to 2019 and the data in 2020 and 2021 are afterwards added onto the graph.

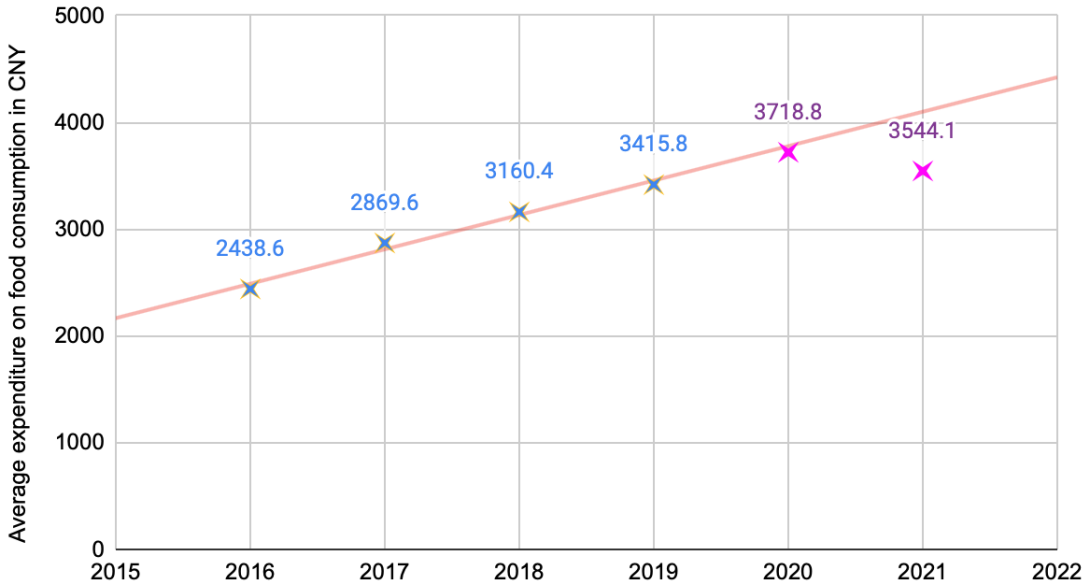


Figure 6. Average expenditure (in CNY) on food consumption in Guangdong, Hebei, Jiangsu, Shandong, Zhejiang, Sichuan, Henan, Anhui, Hunan, Fujian. [2,3,5]

The calculated r-value for the trend line is 0.992. This indicates that there is a strong linear correlation between the average expenditure on food consumption per capita and the year. The calculated gradient of the trend line is 322. According to this model, I predict that in 2020, the average expenditure on food consumption in these 10 cities is 3437.8 CNY. This only differs from the actual value by 0.5%. Hence the model well fits the correlation between the two variables.

If there had not been the pandemic, under this model, the predicted average expenditure on food consumption would be 4059.8 CNY in 2021. It differs from the true value by 515.7 CNY. Hence the perceived actual damage of pandemic on the average expenditure on food consumption in the 10 cities is 12.7%. In view of the large population in the 10 cities at 803.5 million in total in 2021, the total loss in F&B industry revenue is calculated to be 414.36 billion CNY. This shows the large extent of impact of the pandemic on the F&B industry. [2]

5. Further analysis

The multiple linear regression model analyses the relationship between one dependent variable and two or more independent variables. The calculated equation for the relationship is $y = c_0 + c_1x_1 + c_2x_2$, where c_0, c_1,c_2 are constants, x_1 and x_2 are independent variables. In this case, we consider other than GDP, whether the average age of the working force also affects the F&B industry revenue in the city.

GDP	Average workforce age	total revenue
76469.7	39.4	3995
97277.8	37.3	3884.6
92595.4	39.2	3430.5
36010.3	39	3125.6
48055.9	38	2887
40678.1	39	2807.4
56197	39.5	2716
36425.8	39.5	1921.6
39366.5	39.2	1747.7
25315.4	40.3	1730.8
35804	38.6	1447.5
30006.8	38.4	1322
20363.2	39.4	1318.7
16361.6	40.5	1135
17289.2	39.6	1103.9
30320	38.1	1101.8
18809.6	38.8	1020
15074.6	40.2	1012.7
16531.3	38	947.8
24438.3	38.1	947
20352.5	38.2	821.3
8246.1	38	797.2
16818.1	38.4	678.2
12199.1	37.7	444.5
14806.5	36.7	403.6
4832.1	37	240.6
3705.2	37.6	166.3
1477.6	36.5	100
2865.2	38.1	59.2
32679.9	38.8	1099.9
21984.8	38.5	951.8

Figure 7: the total revenue in F&B industry (in hundred million CNY), GDP (in hundred million CNY) and the average age of the workforce in 31 provinces and municipalities in China in 2019

Using the data from Figure 7: according to my calculation, $r^2 = 0.8716796543$, $r = 0.93364$. Hence, I think the multiple linear regression model fits the relationship between the total F&B industry revenue and average workforce age. The multiple linear regression equation calculated is $R = -7000.213851 + 188.138A + 0.04098490224G$, where R is the total F&B industry revenue, A is the average age of the workforce and G is the GDP of that city. Therefore, I think that we should also include the average workforce age as a factor when calculating the predicted value for F&B revenue other than the GDP of the city.

6. Conclusion

In this essay, the relationship between F&B industry revenue, average workforce age and GDP across cities in China is studied. Covid-19 has placed considerable stress on the F&B industry and forced many firms to set up their own online stores.[6] Fortunately, the pandemic has been well-contained in most parts of China. Those areas affected have also improved their situations. This greatly boosted the confidence of investors and companies in the industry and encouraged and enabled consumers to consume freely. As restrictions on people's mobility may further loosen in well-protected areas, most parts of China will have their F&B industry better in the following years. Aiming at this situation, the following work is carried out in the paper:

Through a large amount of data on the F&B revenue in the 27 provinces and 4 municipalities in China, the linear correlation between GDP and F&B revenue in 2019 and 2020 is studied. The best-fit lines for both years are drawn, and the r-value shows a strong correlation between the two data. The model generated well fits the relationship as the prediction of 2021's F&B revenue does not differ from the true value by only 4.61%.

Through the comparison between average consumption expenditure in the first-tier cities before and since China entered the pandemic, the level of damage on the F&B industry by the pandemic is studied. A monetary value measuring the extent of damage is calculated.

Considering the possible relationship between F&B industry revenue and the average age of workforce, the multiple linear regression model is used. The strong correlation in GDP, average workforce age to the F&B revenue shows that there are possibly more factors and data related to F&B industry revenue.

References

[1] Investigation and analysis report on the impact of the epidemic on the catering industry in Taiyuan, Shanxi, China Quality News

[2] Revenue scale of China's catering industry, The National Bureau of Statistics

[3] Annual Report on China's Catering Industry, China Hospitality Association

[4] Proactive Fiscal Policy Protects Economic Market, Economic Daily

[5] Seventh Census Data, Red and Black Population Pool

[6] Under the epidemic, the catering industry moves forward against the wind, XinHuanet