Jingwen Ni

Rough draft

<https://github.com/macs30200-s22/replication-materials-nijingwen>

***Title****：The study of covid19's impacts of unemployment in China on provinces' level in 2020 by using regression and time series model.*

***Abstract:***

Unemployment is a social problem that affect the social stability and economic development. The existing of Covid19 leads unstructured changes of unemployment rate in the worldwide. In UK, covid19 promotes the unemployment rate of the nationwide. However, the political and social condition in China is different from it in UK while China has much larger land area. Covid19 causes the economic and geographically locked down in many provinces in China in 2020. Because of the crisis leaded different policies in provinces, the change of unemployment rate of provinces should vary. My result is consistent with literature in UK, but I use different method to measure and analysis the covid impact on unemployment rate. This paper is going to run a regression model and time series model to analyze How covid19 affects unemployment rate in China on provinces’ level in 2020? Data of control variables and dependent variables are scraping from Chinese government. The data of covid19 was download from Harvard University. After train-test the model and visualizing the result, the study finds that Covid will promote unemployment in the nationwide model but in provinces level, covid19 promoted unemployment in 22 provinces and 9 provinces have decreasing unemployment rate in 2020. Covid impacts will be divided geographically into two parts: West-Southern and East-Northern.

The study is Supplement of covid19 impacts of China because there are lacking study of provinces level in China. The limitation is data from Chinese government could not be persuasive because Chinese government sometimes post fake data to keep social stability and avoid citizens anxiety and the covid dataset was recorded by province instead of region like community. In addition, the regression model is not accuracy because of the limited control variables selection. To conclude, Covid19 will promote unemployment rate in China on province level generally, but the impacts is various in different provinces.

**Introduction**

Covid19, later renamed as Coronavirus, gripped China since the beginning of 2020(Qiu et al.2020). Wuhan in Hubei province was the first city where was found covid19 cases in China. Covid19 spread rapidly over the world and brought unignorable crisis in both economics and medical systems and by the end of November in 2021, there were more than 63 million reported cases and 1.4 million deaths over the world (Brodeur, Gray, Islam, Bhuiyan 2021). Because Covid19 is a respiratory infectious disease, countries published several policies about social distance, quarantine to reduce covid19 infections. This contagious disease changed the world from economics and politics throughout the history (Ceylan, Ozkan, Mulazimogullari, 2020). In China, government took enforce measures that locking down the cities which has covid cases which slowdown the economics activities considerably.

The cost of the covid19 was burden on China’s economics. The control measures to prevent the covid19 leaded 2.7% loss of China’s annual gross domestic product. From BBC news, From January 23rd, Wuhan was locked down for 72 days which means no one can exit and enter this city. Citizens in Wuhan cannot get out of their home. They can only stay and wait for necessities delivery which was arranged by government. To the nationwide, China also took actions: for instance, The risk classification for each country is based on the principle of formulating guidelines for classification and implementing the strategy of "internal prevention of spread and external prevention of import" to reduce the possible impact of imported cases on China. Establish a joint prevention and control mechanism at ports of entry, organize the Civil Aviation Administration, customs, public security, health, foreign affairs, border control, airports, and other departments to coordinate the information, identity registration, health monitoring, emergency response and other related work. Due to covid19, both exports and imports were limited which leaded many businesses got into trouble. They had to cut the number of employees to reduce operations expenses. Hence in this case, the unemployment rate should increase.

In another case, the unemployment could decrease because of more remote jobs were provided during the covid19 period. Many women who were a housewife and gave up their jobs to take care of their families, could have a chance to balance their work and families. In addition, because of the remote option, people did not need to worry about the relocation problem so that they can have more options for applying jobs which would decrease the unemployment rate.

China, a socialist country, which labor markets reformed and transformed into a market-driven market system. After the reformation, people were not limited by “Hukou”, a certification of their birthplace and they were not allowed to work in outside of the birthplace before the reformation, so that there were a lot of migrant-workers those affected by Covid19, the quarantine and lock-down policies. Unemployment was a historical problem and become more serious with time in China because China is a populous country. From United Nations, the world population was 7.6 billion and the population of Mainland was 1.39 billion which was 18.3% of the world while China only has 7.059% land area share of the world. From January 1996 to September 2002, the unemployment of urban residents increased from 6.1% to 11.1% (Giles, Park, Zhang, 2004). In economics, unemployment was a significant factor which would lead to poverty and income disparity in China. In 1999, urban unemployment was a major cause of urban poverty, and the growing urban poverty became a significant factor which worsen the urban inequality. This inequality in China has increasing influence on migrant households (Xue, Zhong, 2003). In addition to economic perspective, unemployment would also cause psychological and health problems. The unemployed group could have greater symptoms like depression and anxiety than employed group. Moreover, as central of policy debate and aggregate resource utilization, unemployment was also an important indicator (Gali, Smets, Wouters 2022). Furthermore, unemployed group was more likely to visit a physician than employed group (Linn, Sandifer, Stein, 2011). Accordingly, unemployment in China was worth to putting efforts on.

There are 31 provinces in China, the policy about covid were various and depend on their own situation. In addition, because of the geographic differences, the policy could also be affected. In previous study, scholars analyzed the covid19’s effects on unemployment in different races, and gender group in United States. And they found the effects of covid 19 on unemployment were significant different on races and genders (Gezici, Ozay, 2020). In China, there were scholars studied the covid impacts on nationwide level in 2020 because it is obvious that the unemployment rate increased in 2020 compared with previous years. In Brazil, researchers found the burden of covid 19 is greater in areas with high social deprivation. Until August 6th 2020, Bahia has 179139 confirmed cases, and 3767 deaths. There was a spatial association between the epidemiological indicators and SDI been observed. 22 municipalities had the priority for incidence which was 1.6 times higher than state rate. And there were 40 cities which had 1.2 times higher death rate than the state one while they also had 4.1 times higher than the state rate (Souza, Carmo, Machado, 2020). It motivated me to study about the covid19 impacts on unemployment on province level in China instead of study the impacts on nationwide level. The unemployment rate of provinces has different trends, and some fluctuations were caused by seasonal, structural etc and could be affected by other factors. Covid19 would not disappear in China and in the world rapidly. Even though only one country over the world has the disease, the covid 19 never ends. Accordingly, to face, solve, and predict the social problems from Covid, the study needs to analyze how much fluctuations was caused by covid19.

To study the research question, how covid19 affected unemployment rate in China on province level in 2020, I utilized both regression and time series model to make prediction and analyze the results with the dataset scrapping from National Bureau of Statistics of China. Before starting the work, the hypothesis I had is covid19 will promote unemployment rate in China by province level in 2020. The regression model can show the feature importance and the causal effects between covid19 and unemployment. From regression model, one unit of confirmed increasing will lead to 0.066 unit increase in unemployment, which prove the hypothesis that covid will promote unemployment rate of provinces in China. The time series can predict unemployment by avoiding using control variables’ data which could be affected by covid. I found the covid impacts were various based on geographical reasons and then lead to different impacts on unemployment rate of provinces.

The limitation of the dataset also cannot be ignored. In the regression model, the selection of control variables was not comprehensive. The dataset of control variables was scraping from Chinese government which could not be the true value. Besides, the time series makes prediction by trends over time so it cannot consider the special cases in the model like policy impact. The strengths of the research are using two different models for studying the impacts of covid and clearly show the initial finding by data visualization on Tableau.

Because covid19 will not disappear fast, this research can provide guideline to Chinese government and policy makers that which region should be put more effort on to reduce unemployment rate. In addition, the model can make government to predict future unemployment condition and do a better preparation. For academic field, although every country has different policy and condition but this work can still provide them some ideas about analyzing covid19 regionally.

***Data and Method:***

The goal of the project is to figure out the impact of covid 19 on employment rate of provinces in China. The hypothesis is Covid19 has negative impact on employment of provinces in 2020 which means it will increase the unemployment rate. There are many factors which will impact employment besides covid 19, so I choose them as control variables. For example, the development level of provinces, the price level of production, education level, the change of population of the province (Zeng 2020). To measure these concepts, I will use GDP to measure general development level, CPI to measure price level, the graduation number university, and the college (Note: college is for students who failed entered the university in China) to measure education level, the death rate and born rate to measure the change of province’s population and residence population as control variables, the first industry increased income, second industry increased income and third industry increased income, total retail income to measure economic development. To measure the impact of covid19, I will use the number of covid19 confirmed, the number of deaths, the number of recoveries of each province in 2020 to as my dependent variable.

***Data:***

There are two parts of data collection of the research: Covid19 data, and control variables’ data. Firstly, I use the data of residence population, CPI, GDP, Graduation number of university, graduation number of college, natural population increasing rate, first industry increased income, second industry increased income and third industry increased income, total retail income, on province level from Chinese National Bureau of Statistics from 2008 to 2020. For this data source, I will run a data crawler to scrap the data from National Bureau of Statistics because it is impossible to download and combine by hand, the data of each province are separate in different pages and csv on the website and the URLs of website does not include the page number so when I wrote the script, I needed to find the “Next Page” button instead of the page number in URLs . When I scrape the website, it blocked me serval times because there are thousand pages of data need to scrape one by one. To solve this problem, I firstly added some random break in the scraping code and then use three computers with different IP to scrape the data. After collecting the data, I used terminal to combine all csv document.

This dataset of residence population is updated yearly which has 4 columns: Year-end population, region for province name, time for data collection, and value for total residence population. The mean value of population is 4407.40 million. The minimum value is 292.33 million while the maximum population is 12489 million and the median value is 3823.52. The data of CPI of province level, including 4 columns: CPI for province, name of province, time of collection, value of CPI, is updated half year. Hence, this dataset needs to apply the mean value of the first half and the second half of CPI of each province before utilizing the CPI data. The mean CPI is 102.67, the median CPI is 102.4, the minimum is 97.65 and the maximum is 110.09. The dataset of graduation number of university and college also includes four columns while the first column becomes predicted graduation number, and the unit of data is ten thousand. The average number of graduations from university has the unit of million which is 11.28, the maximum value is 30.83, the minimum value is 0.34 and the median value is 11.02. The average number of graduations from college has the unit of million which is 10.98, the maximum value is 35.11, the minimum value is 0.31 and the median value is 9.13. For the GDP, the average value is 20265.25, the maximum value is 107986.92. The minimum GDP is 398.2 and the median value is 14580.35. The mean value of total retails is 8218.86, the maximum value is 42951.8, the minimum value is 130 and the median value is 5761.8. For the first industry increased income, the average value is 1671.15. The maximum value is 5116.99, the minimum value is 57.6 and the median value is 1546.15 with the unit billion. For the second industry increased income, the average value is 8790.42. The maximum value is 43507.53, the minimum value is 110.8 and the median value is 6133.7 with the unit billion. For the third industry increased income, the average value is 9803.7. The maximum value is 60268.1, the minimum value is 229.8 and the median value is 6630 with the unit billion. The average natural population increase rate has the average 5.34, the maximum value is11.47, the minimum is -1.01 and the median value is 5.54. The data of unemployment rate in China by province from 2008-2020 is found on Chinese National Bureau of Statistics either. The average unemployment rate has the average 3.34, the maximum value is 4.57, the minimum is 0 and the median value is 3.41. The dataset includes 31 provinces in China. The average value is 3.34, For most provinces, the unemployment rate increased rapidly in 2020 while a few of provinces in the southern part of China have decreasing trend.

Second, the covid19’s data is collected from Harvard and the dataset is called China covid-19 cases which includes the covid19 cases from 2020.01.15-2021.01.15. This dataset of covid19 was recorded by daily and data of confirmed, death, recovery are separate in three worksheets, so I need to count the total number of confirm, death, recovery number of each province. After collecting all needed data, I combine them into one csv file. The columns include provinces name, unemployment rate, residence population, CPI, GDP, Graduation number of university, graduation number of college, natural population increasing rate, first industry increased income, second industry increased income, third industry increased income, total retail sales, and row is year. The dataset has no NA value, and it is recorded in detail, so it is no needed to do extra data cleaning. However, because the scales of variables are very different, it is necessary to do data scaling to avoid unbalancing.

In addition, here is a limitation of the data of residence population, CPI, GDP, Graduation number of university, graduation number of college, natural population increasing rate, total retail sales, first industry increased income, second industry increased income, third industry increased income on province level from Chinese National Bureau of Statistics are lacking authenticity because Chinese government may conceal real data to maintain social stability.

***Method：***

To figure out the impact of covid 19 on unemployment rate on province level in China, I would like to use regression model and time series model for this research. Regression model is commonly used in interpreting causal effect by making comparison. To make the method work, I assume the selected control variables are complete and the correlation between them are low, there is a linear relationship between covid19 and unemployment rate, and all independent variables in equation are uncorrelated with error term. In this research, the covid19 is the causal variable and unemployment rate is the influenced variable.

I utilized regression model to analyze the causal effects ((Aronow, Samii, 2015) between covid19 and unemployment rate by checking the p value and coefficient of variables in OLS regression. Time series model is used to make prediction of unemployment rate in 2020 if covid 19 never exists based on the trend line from 2008 to 2019. Then I make a comparison between the true unemployment rate of each province in 2020 and the predicted unemployment rate of the time series model. The reason that I use time series model to make prediction of unemployment rate in 2020 without covid effects instead of regression model is the data of control variables in 2020 which needs to be plot into the regression model have already been affected by covid19, for example, the GDP, CPI, Total retail sales because of the lock down policy.

After making the prediction, there exists a gap between the predicted unemployment rate and true unemployment rate of each province. The gap can be treated as the impacts from covid19 on unemployment rate. Then the gap value become the dependent variable of the regression model, and covid19 variables will also be added as independent variables. To satisfy the assumption that the variables should be mutually independent, so I calculate the correlations and draw the Heatmap (figure1) to make it more visualized. The correlation over 0.9 are removed or added together. To analyze causal effects between covid19 and unemployment rate in China on provinces’ level, this project ran a regression model (Aronow, Samii, 2015). 日历

描述已自动生成

*Figure 1: Heatmap of correlation*

From the heatmap, the second and third industry income increase in 2020 were high corelated with GDP, which is 0.98. The correlation between total retail sales and second industry is 0.96 and third industry is 0.97. The correlation between second industry and third industry is 0.99. So, for those correlation, which is larger than 0.9, I only choose GDP, which is more representative than second, third industry income, and total retail sales of each province as control variable instead of using all these variables.

At this time, we can check whether covid19 is statically significant through the p value and feature significant through variables’ coefficient in OLS regression model. Finally, I choose the province which has the largest gap with the unemployment rate and the one has the smallest gap to do data analysis. The limitation of method firstly on the control variables selection. The economic model for unemployment rate includes many factors but I only pick a few of them in as the control variables so the model could be not comprehensive enough and has low accuracy. For the time series model, the R squares for each province are different. Most provinces have high R square which is over 0.85 but few of them are only around 0.25. Hence, the picked two provinces which have the largest and smallest gap are not the real province which has the largest and smallest gap.

*Alternative:*

Besides, here is an alternative non-computational approach to test the hypothesis: Take a survey for basic information like education level, work status, age, gender, income of household etc. for 100000 people in each province.

Compared to the non-computational way, my approach can save more time and money because the data I used is existing and the non-computational method need too many scholars to help with. Also, it is hard to find a non-biased sample which has the size of 100000 per province. But the data of non-computational way is more updated than my approach because the existing data was updated at 2020.8.16 and researchers can have chance to communicate with participants and improve the research which can help them do data interpretation comprehensively.

After analyzing the data, I will make a dashboard in Tableau to make the data more visualized.

***Results***

Method 1: time series

For the time series method, I only choose confirmed number of covid as the measurement of covid impact. I draw a bar chart in Tableau to make data visible. The x axis is the number of confirmed, and y axis is the provinces’ name. From the graph (Figure 2), Hubei province has the most covid cases which is 23,087,143 and Xizang province has the least which is 351.

图表, 条形图, 直方图

描述已自动生成

*Figure 2: bar chart of number of covid confirmed*

After overview the covid condition of 31 provinces in China. I use time series to predict the unemployment rate in 2020. Figure 6 is the head of the graph. In figure 6, the x axis is time, the y axis is province’s unemployment. The whole sheet will have 31 separate graphs of 31 provinces.

图表

描述已自动生成

*Figure 3: Time series prediction on unemployment rate in 2020*

The blue line is the true unemployment rate in 2020, and the yellow line is the predicted unemployment rate. To analyze the gap between predicted and true unemployment rate, I am wondering it may be caused by geographic reason, so I combine the gap value and map.(Figure 4) The darker color means larger gap and lighter means smaller gap.

地图

描述已自动生成

*Figure4: Time series prediction-true gap on unemployment rate in 2020*

The darkest area which has the value of 2.512 is Beijing and the lightest one with value of -0.022 is Sichuan province. From the graph, we can see the larger gap concentrate in the southern part of China and the northern part has comparative smaller gap between predicted and true unemployment rate. Hence, the covid will also affected by weather, temperature etc. geographical reasons and then lead to different impacts on unemployment rate.

Method 2: Regression

To see whether covid is statistically significant in unemployment rate in China of different provinces, I utilized OLS regression. (Figure5) 手机屏幕截图

描述已自动生成

*Figure 5: OLS regression results*

From the figure, the R-square of the model is 1.0 which means the model is fit well. The p value of confirmed is 0.501, of deaths is 0.00 and of recovery is 0.00. The p value of deaths and recovery are smaller than 0.05 which means they are significant in the gap of predicted and true unemployment rate. Besides deaths and recovery, the graduations of college 0.003 which is also statistically significant. The hypothesis of the research is: covid19 promote the unemployment rate in China. To test the hypothesis, I divided the test into nationwide part and provinces’ part. I calculate the coefficient of variables of the nationwide by linear regression (Figure 6) and using coefficient of time series model to see the province level one by one.

图表

描述已自动生成

*Figure 6: Feature coefficient of unemployment rate*

The recovery, graduations from university, first industry income, confirmed has positive relation with unemployment rate. And the CPI, GDP, Graduations of college, population and deaths has negative relationship with unemployment. The coefficient of recovery is 7.025, of confirmed is 0.066, of death is -7.098. One unit of recovery increasing will lead to 7.025 unit increase in unemployment rate. One unit of confirmed increasing will lead to 0.066 unit increase in unemployment rate. One unit of death increasing will lead to 7.098 unit decrease in unemployment. From the covid dataset, the covid confirmed is the sum of recovery and deaths. Hence, the number of covid confirmed can be representative to show the covid cases condition. Because of the coefficient of confirmed is 0.66, which prove the hypothesis that covid will promote unemployment rate of nationwide in China.

On the provinces level, from the coefficient of time series model. There are 31 provinces which have different outcomes. The dependent variable is unemployment rate, and covid cases are as independent variables with other control variables. There are 22 provinces have the positive coefficient that means there are 22 out of 31 provinces have increasing unemployment rate and 9 provinces have decreasing unemployment. On provinces’ level, the hypothesis was partially rejected, and it was true on nation level.

***Discussion:***

In the research, the research question that how covid 19 affect unemployment rate in China on province level gets an initial answer that there exists a causal effect between covid 19 and increasing unemployment rate by run the regression model. the hypothesis: covid19 promotes the unemployment rate in China on province level was supported by calculating the feature coefficient of time series model. And the nationwide condition was tested by coefficient of linear regression model. The result of the research is consistent with the similar research that covid 19 economic crisis leads vast increase unemployment and competition between worker in the labor market (Blustein, Duffy.etc 2020).

The score of regression and accuracy of prediction of time series are not high enough. Time series is better than run regression because regression model needs plot the data of control variables in 2020 to make prediction. Although I have not added covid19 impact when I run the regression for the prediction, the data of control variables like GDP, CPI etc (Ke, Hsiao, 2021). have already been affected by covid 19 which could reduce the impact of covid 19 on unemployment rate and make bigger error on prediction and larger gap. However, the time series will not be affected by the data of control variables in 2020. It makes prediction based on the trends over time.

For the regression, the control variables are not enough to do the better regression. The economic model for unemployment rate includes many factors but I only pick a few of them in as the control variables so the model could be not comprehensive enough and has low accuracy. Also, after the feature selection by calculating and eliminating high correlated features which has over 0.9 correlation, there are still some variables have correlation over 0.7. For the time series model, the R squares for each province are different. Most provinces have high R square which is over 0.85 but few of them are only around 0.25. Hence, the picked two provinces which have the largest and smallest gap are not the real province which has the largest and smallest gap.

The limitation of the dataset also cannot be ignored. Firstly, the covid cases were reported on a province instead of the regional level like community level. The data could not be updated and comprehensive enough. The covid hospitalization and death rate are concentrate in people older than 70 years old while the population statistics did not show the observed regional difference in covid19 indicators could be leaded to differences in age distribution (Holmager, Lynge el 2021). The dataset of control variables was scraping from Chinese government. To avoid citizens’ anxiety, Chinese government could hide the real data which may cause the models perform not ideally. Additionally, the time series makes prediction by trends over time so it cannot consider the special cases in the model like policy impact.

The strengths of the research are using two different models for studying the impacts of covid and clearly show the initial finding by data visualization. And analyzing the impacts of covid 19 of unemployment rate in both province and nationwide level. When studying the covid 19 impact on unemployment rate, I think one more step about why covid 19 impacts were different geographically.

The reason that the impacts were various geographically is temperature and humidity. Based on previous study, there exists a homogeneity regarding the effect of temperature and humidity on the seasonal viability and transmissibility of covid19. The cold and dry condition make covid19 be more likely to spread than warm and wet weather (Mecenas, Bastos, Normando, 2020).

In southeast part of China, the temperature and humidity are higher than northwest part. From the previous graph, it is obvious that, the gap value between predicted unemployment rate and true unemployment in northwest part is much larger than that in southeast part which is consistent with the virus characteristics. 地图

描述已自动生成

Because of the difference of weather condition in northwest and southeast parts. The virus spread much faster and broad in northwest part of China. The China government started the lock down policy in these provinces to keep social distance. Accordingly, the covid19 has better effects on economic and society in northwest part. More companies in southeast part were shut down or reducing employees to relief economic pressure. Hence, unemployment rate in northwest are increase significantly which is consistent with that the gap value between predicted unemployment rate if covid19 never exists and true unemployment rate with covid19 impacts is larger in northwest provinces.

For future study, we should put more detailed information. The dataset of covid19 cases should be recorded daily in community level. The age, gender, recovery time should also be considered in studying the covid19 impacts. The unemployment rate is still too broad if we want to focus on. Different groups will also have different effects from covid19. What kind of group is more vulnerable? Does elder people are more likely to lose their job? Does first year graduated students cannot find a job? After we decompose the covid19 impacts in detail, it can help government improve the condition fast and efficient.

***Conclusion:***

The research question was answered by the prediction of unemployment rate by time series model and the p value of OLS regression model. The p value of covid19 deaths cases and recovery cases are less than 0.05 which mean statistically significant in the gap of predicted and true unemployment rate that covid 19 has large impacts on unemployment rate in China on provinces’ level. In addition, to test the hypothesis that the covid 19 will promote the unemployment. I used linear regression model to test the nationwide level and time series to test for the provinces’ level. On nationwide level, the covid19 promoted the unemployment rate and on provinces level, covid19 promoted unemployment in 22 provinces and decrease the unemployment rate in 9 provinces.

***Reference:***

1. Fang Cai & Kam Wing Chan (2009) The Global Economic Crisis and Unemployment in China, Eurasian Geography and Economics, 50:5, 513-531, DOI: 10.2747/1539-7216.50.5.513
2. GILES, John, et al. “What Is China's True Unemployment Rate?” *China Economic Review*, vol. 16, no. 2, 2005, pp. 149–170., https://doi.org/10.1016/j.chieco.2004.11.002.
3. Xue, Jinjun, and Wei Zhong. “Unemployment, Poverty and Income Disparity in Urban China.” *Asian Economic Journal*, vol. 17, no. 4, 2003, pp. 383–405., https://doi.org/10.1111/j.1467-8381.2003.00192.x.
4. M W Linn, R Sandifer, and S Stein, 1985:[Effects of unemployment on mental and physical health.](https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.75.5.502)American Journal of Public Health 75, 502\_506, [https://doi.org/10.2105/AJPH.75.5.502](https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.75.5.502)
5. Gezici, A., Ozay, O. An Intersectional Analysis of COVID-19 Unemployment. *J Econ Race Policy* **3,**270–281 (2020). https://doi.org/10.1007/s41996-020-00075-w
6. [Niraj Poudyal](https://www.journals.uchicago.edu/author/Poudyal%2C+Niraj), [Aris Spanos](https://www.journals.uchicago.edu/author/Spanos%2C+Aris) [Model Validation and DSGE Modeling](https://doi.org/10.3390/econometrics10020017), Econometrics 10, no.22 (Apr 2022): 17.
7. Ceylan, R.F., Ozkan, B. & Mulazimogullari, E. Historical evidence for economic effects of COVID-19. *Eur J Health Econ* **21,**817–823 (2020). https://doi.org/10.1007/s10198-020-01206-8
8. https://docs.google.com/document/d/14nEfaKG8eyPWPbRk3YAoa5N-NLJqtHCPriuMqF2QhoY/edit?usp=sharing
9. Aronow, Peter M., and Cyrus Samii. “Does Regression Produce Representative Estimates of Causal Effects?” *American Journal of Political Science*, vol. 60, no. 1, 2015, pp. 250–267., <https://doi.org/10.1111/ajps.12185>.
10. Aronow, Peter M., and Cyrus Samii. “Does Regression Produce Representative Estimates of Causal Effects?” *American Journal of Political Science*, vol. 60, no. 1, 2015, pp. 250–267., https://doi.org/10.1111/ajps.12185.
11. Blustein, David L., et al. “Unemployment in the Time of COVID-19: A Research Agenda.” *Journal of Vocational Behavior*, vol. 119, 2020, p. 103436., https://doi.org/10.1016/j.jvb.2020.103436.
12. Holmager, Therese Lf, et al. “Geography of Covid-19 in Denmark.” *Scandinavian Journal of Public Health*, vol. 49, no. 1, 2020, pp. 88–95., https://doi.org/10.1177/1403494820975607.
13. Mecenas, Paulo, et al. “Effects of Temperature and Humidity on the Spread of Covid-19: A Systematic Review.” *PLOS ONE*, vol. 15, no. 9, 2020, https://doi.org/10.1371/journal.pone.0238339.
14. Chen, Jian, and Belton M. Fleisher. “Regional Income Inequality and Economic Growth in China.” Journal of Comparative Economics, vol. 22, no. 2, 1996, pp. 141–164., <https://doi.org/10.1006/jcec.1996.0015>.
15. Blundell, Richard, et al. “Covid‐19 and Inequalities\*.” Fiscal Studies, vol. 41, no. 2, 2020, pp. 291–319., https://doi.org/10.1111/1475-5890.12232.
16. Xie, Yu, and Xiang Zhou. “Income Inequality in Today’s China.” Proceedings of the National Academy of Sciences, vol. 111, no. 19, 2014, pp. 6928–6933., https://doi.org/10.1073/pnas.1403158111.
17. Savvides, Andreas, and Thanasis Stengos. “Income Inequality and Economic Development: Evidence from the Threshold Regression Model.” Economics Letters, North-Holland, 18 Sept. 2000, https://www.sciencedirect.com/science/article/pii/S0165176500002937?casa\_token=4ADP9XnwUjgAAAAA%3AfEF6oQ71lQFUds7wkYPXROPXdqrsFdzhf4myDoj31fokyhWEZl4u\_Wl\_EVL35Ld2U18IGed7.
18. De Maio, Fernando G. “Income Inequality Measures.” Journal of Epidemiology and Community Health, BMJ Group, Oct. 2007, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2652960/.
19. Li, Yan, and Maria N. DaCosta. “Transportation and Income Inequality in China: 1978–2007.” Transportation Research Part A: Policy and Practice, Pergamon, 11 Sept. 2013, https://www.sciencedirect.com/science/article/pii/S0965856413001523?casa\_token=efJWewtEwGkAAAAA%3A\_Rv3oFf8rIM8E17FoZaWJ\_TrK-h5DCnQw8X2\_TJsZqYQvzGxpKwz9WwMsHZHSbenaWC5i92X.
20. Qiu, Yun, et al. “Impacts of Social and Economic Factors on the Transmission of Coronavirus Disease 2019 (Covid-19) in China - Journal of Population Economics.” SpringerLink, Springer Berlin Heidelberg, 9 May 2020, https://link.springer.com/article/10.1007/s00148-020-00778-2.
21. Shen, Jianfu, et al. “Covid-19 and Regional Income Inequality in China.” Frontiers in Public Health, Frontiers Media S.A., 11 May 2021, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8144473/.
22. Smith, Mark Kimura. “Relation between GDP per Capita and Income Inequality.” https://smartech.gatech.edu/bitstream/handle/1853/64475/GDP%20vs%20Income%20Inequality%20final%20draft.pdf?sequence=2.
23. Sicular, Terry, et al. “The Urban-Rural Income Gap and Income Inequality in China.” SpringerLink, Palgrave Macmillan UK, 1 Jan. 1970, https://link.springer.com/chapter/10.1057/9780230584259\_2.
24. Knight, John & Song, Lina, 1999. "[**The Rural-Urban Divide: Economic Disparities and Interactions in China**](https://ideas.repec.org/b/oxp/obooks/9780198293309.html)," [OUP Catalogue](https://ideas.repec.org/s/oxp/obooks.html), Oxford University Press, number 9780198293309.
25. 关于统计上划分城乡的规定, http://www.stats.gov.cn/tjsj/pcsj/rkpc/5rp/html/append7.htm.
26. Jung, Jaehun. “A Long Way to the Recovery: Covid-19 Will Not Disappear.” Journal of Korean Medical Science, vol. 36, no. 32, 2021, https://doi.org/10.3346/jkms.2021.36.e231.
27. li, binqin. 中国的不平等问题：现状、原因 ... -. https://www.researchgate.net/profile/Bingqin-Li/publication/286931679\_zhongguodebupingdengwentixianzhuangyuanyinjijianyi\_Problems\_of\_inequality\_status\_quo\_causes\_and\_policy\_implications/links/5670fbc208ae2b1f87ad0582/zhongguodebupingdengwentixianzhuangyuanyinjijianyi-Problems-of-inequality-status-quo-causes-and-policy-implications.pdf.
28. Li, W., Cai, Z. & Cao, S. What has caused regional income inequality in China? Effects of 10 socioeconomic factors on per capita income. *Environ Dev Sustain* **23,**13403–13417 (2021). https://doi.org/10.1007/s10668-020-01218-7
29. Donthu, Naveen, and Anders Gustafsson. “Effects of COVID-19 on Business and Research.” *Journal of Business Research*, Elsevier, 9 June 2020, https://www.sciencedirect.com/science/article/pii/S0148296320303830?casa\_token=IX\_5yi8dmjQAAAAA%3A3XqotBXh98wVpK02zG2QGXix8G3JzdlQMntQosbMSc7ZRRWh9g2oGZMJrkFba14fcEOfVA\_Q.