Jingwen Ni

MACS 30200

**The Study of Covid-19's Impacts on Unemployment in China on Provinces' Level In 2020 Using Regression and Time Series Model**

***Abstract:***

Unemployment is a social problem that affects social stability and economic development. The existence of Covid-19 leads to unstructured changes in the unemployment rate worldwide. In the UK, Covid-19 promotes the unemployment rate nationwide. However, China's political and social condition is different from those in the UK, while China has a much larger land area. Covid-19 caused the economy and geography to be locked down in many provinces in China in 2020. Because the crisis led to different policies in provinces, the change of unemployment rate of provinces should vary. My result is partially consistent with the literature in the UK, but I use different methods to measure and analyze the Covid-19 impact on the unemployment rate. This paper will run a regression model and time series model to analyze how Covid-19 affected the unemployment rate in China at the province level in 2020. Data of control variables and dependent variables are scraped from the Chinese government. The data of Covid-19 was downloaded from Harvard University. After train-test the regression model and visualizing the result, the study finds that Covid-19 will promote unemployment in the nationwide model. However, Covid-19 promoted generally unemployment in 22 provinces at the province level, and 9 provinces had a decreasing unemployment rate in 2020. Overall, Covid-19 impacts will be divided geographically into two parts: Southwest and Northeast.

The study is a supplement to Covid-19 impacts on China because there are lacking study of the provinces level in China. One of the limitations is that data from the Chinese government could not be persuasive because the Chinese government sometimes posts fake data to keep social stability and avoid citizens' anxiety. The covid dataset was recorded by province instead of the region like the community. Moreover, the regression model is not accurate owing to the relatively small selection of control variables. Covid-19 will generally promote China's unemployment rate at the province level, but the impacts vary in different provinces.

*Keywords: Unemployment, Covid19, Time Series, Regression, China*

***Introduction***

Covid-19 later renamed Coronavirus, gripped China since the beginning of 2020 (Qiu et al., 2020). Wuhan in Hubei province was the first city to find covid19 cases in China. Covid-19 spread rapidly worldwide and brought an unignorable crisis in both economics and medical systems. By the end of November 2021, there were more than 63 million reported cases and 1.4 million deaths worldwide (Brodeur et al., 2021). Because Covid-19 is a respiratory infectious disease, countries published several policies about social distance and quarantine to reduce Covid-19 infections. The coronavirus disease disrupted these economic and labor market developments by exacerbating unemployment rates and causing a global economic recession. This contagious disease changed the world in economics and politics throughout history (Ceylan et al., 2020). Covid19 is the most substantial global crisis since the World War II, which negatively impact all countries over the world Boccaletti et al. ([2020](https://link.springer.com/article/10.1007/s00500-021-05871-6#ref-CR8)). The economic crisis caused by Covid19, unlike other economic crisis, whose shocks would last for many years. The degree of financial crash was more severe than the crisis in 2008(Castillo, Melin 2020). From European Commission's Spring 2020 Monetary Conjecture’s report, total national output will contract profoundly add up to EU nations: 7.5% for the EU, 4% for Poland ,9% in Italy, France and Spain (Economica [2020](https://link.springer.com/article/10.1007/s00500-021-05871-6#ref-CR18))

In theorical aspect, in economics, unemployment was a significant factor that would lead to poverty and income disparity. Unemployment was a significant cause of poverty, and the growing poverty became a significant factor that worsened inequality. This inequality has an increasing influence on migrant households (Xue & Zhong, 2003). In addition to the economic perspective, unemployment would also cause psychological and health problems. The unemployed group could have more significant symptoms like depression and anxiety than the employed group. Moreover, unemployment was also an important indicator as central to policy debate and aggregate resource utilization (Gali et al., 2022). Furthermore, the unemployed group was more likely to visit a physician than the employed group (Linn et al., 2011). In previous study, scholars used British Household Panel Survey to estimate dynamic panel data models of unemployment incidence for men and they found strong evidence that an individuals’ previous unemployment experience has implications for their future labor market experience. The impacts of unemployment on families not only including poverty and health but also the education aspect for their children. Unemployment could harm children’s development and their future employment (MCclelland, 2000). However, unfortunately, the coronavirus disease disrupted labor market developments by exacerbating unemployment rates and causing a global economic recession (Ceylan et al., 2020). Most companies were not recruiting new staff. A significant reason for the massive layoffs was the disruptions in global supply chains which saw many of these sectors, mainly export and manufacturing organizations, suffering order cancellations and incomplete orders.

The cost of the Covid-19 was a burden on China's economy. The control measures to prevent the covid19 led to a 2.7% loss in China's annual gross domestic product. From BBC News, from January 23, Wuhan was locked down for 72 days, which means no one could exit and enter this city. Citizens in Wuhan cannot get out of their homes. They can only stay and wait for necessities delivery which the government arranged. Nationwide, China also took action. 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 Covid-19, exports and imports were limited, which led many businesses to get into trouble. The government implemented widespread lockdowns that effectively brought the entire domestic economy to a standstill. Not only were the majority of businesses closed, but all labor movements were halted.

However, the adverse effects of the pandemic on China's employment rates were diverse: some sectors, worker populations, and regions were more affected than others. For instance, migrant workers, predominantly female and senior workers (those older than 40), suffered the worst effects of the economic slowdown. In another case, unemployment could decrease because more remote jobs were provided during the Covid-19 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 (Blustein el, 2020). In addition, because of the remote option, people did not need to worry about the relocation problem to have more options for applying for jobs, decreasing the unemployment rate.

China is a socialist country, has witnessed fascinating economic growth over the past thirty years despite regional disruptions such as the Asian and 2008 financial crises, whose labor markets reformed and transformed into a market-driven market system. The country's high economic growth rates have been accompanied by massive shifts in employment patterns in all economic segments, from primary to secondary to even tertiary sectors. Some of the significant changes in the Chinese labor market were the growing flow of rural workers to urban cities, enforcing stricter labor regulations, and restructuring of the urban employment system. For instance, reforms in the 1980s and 1990s eased the Hukou system for registering households where the labor population was separated into a rural and urban workforce. New labor policies scrapped away the "iron rice bowl" arrangement where urban workers were given a permanent job in a state-owned enterprise (SOE) by allowing rural workers to move beyond their regions and seek employment in urban areas. China's decision to open its market to foreign investors resulted in the marketization of labor, fewer labor market frictions, and a more efficient labor distribution system. Hence, there were a lot of migrant workers affected by Covid19, the quarantine, and lockdown policies.

Fortunately, the Chinese government's quick response to the pandemic and commendable management of new infections greatly factored in the country's quick recovery in employment rates. The coronavirus epidemic was brought under control in urban and rural areas within a short period so that most employment-intensive sectors began resuming operations by April. In developing its economic recovery strategy, the Chinese government focused on containing the Covid-19 pandemic while protecting employment using phased resumption of economic operations. In the first stage, during the widespread lockdown, the government attempted to address the issue of unemployment through schemes for vocational training, job retention, and social assistance. However, marked improvements in employment rates were witnessed in the second stage, after new cases had dropped to single digits. The government directed all economic sectors to resume activities over and above work to protect workers' health. The first sectors to open were public services and transport, then manufacturing and construction, then retail and hospitality. This phased resumption of production saw a marked increase in employment rates post-pandemic. As markets started to recover, most migrant workers who were previously unemployed or underemployed had returned to pre-pandemic employment levels.

Unemployment is a historical problem and has 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 the mainland China was 1.39 billion, which was 18.3% of the world, while China only has a 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 et al., 2004).

There are 31 provinces in China, and there are various policies regarding Covid-19, and they depend on their situation. In addition, because of the geographic differences, the policy could also be affected. In the previous study, scholars analyzed the Covid-19's effects on unemployment in different races and gender groups in the United States. Furthermore, they found that the effects of Covid-19 on unemployment were significantly different on races and genders (Gezici & Ozay, 2020). In China, scholars studied the covid impacts on the national level in 2020 because it is evident that the unemployment rate increased in 2020 compared with previous years. In Brazil, researchers found that the burden of covid 19 is more significant in areas with high social deprivation. Until August 6, 2020, Bahia has 179,139 confirmed cases and 3,767 deaths. There was a spatial association between the epidemiological indicators and SDI observed. Twenty-two municipalities had the priority for incidence, which was 1.6 times higher than the state rate. Furthermore, 40 cities had a 1.2 times higher death rate than the state one, while they also had 4.1 times higher than the state rate (Souza et al., 2020). It motivated me to study the Covid-19 impacts on unemployment at the province level in China instead of national impacts. The unemployment rate of provinces has different trends, and some fluctuations were caused by seasonal, structural, etc., and could be affected by other factors. Covid-19 would not disappear in China and the world rapidly. Even though only one country in the world has the disease, the Covid-19 never ends. Accordingly, to face, solve, and predict the social problems from Covid-19, the study needs to analyze how many fluctuations were caused by Covid-19.

To study the research question of how Covid-19 affected the unemployment rate in China at the province level in 2020, I utilized both regression and time series models to make predictions and analyze the results with the dataset scrapping from the National Bureau of Statistics of China. Before starting the work, the hypothesis I had is Covid-19 would promote the unemployment rate in China at the province level in 2020. The regression model can show the feature importance and the causal effects between covid19 and unemployment. From the regression model, one unit of confirmed increase will lead to a 0.066 unit increase in unemployment, proving the hypothesis that Covid-19 will promote the unemployment rate of provinces in China. Time series forecasting is a technique which can make prediction based on historical trends and assume future trends will be similar as the past trends. is a In this case, the time series can predict unemployment by avoiding using control variables' data which could be affected by Covid-19. I found the Covid-19 impacts varied based on geographical reasons and then led to different impacts on the unemployment rate of provinces.

The limitation of the dataset also cannot be ignored. In the regression model, the selection of control variables was not enough comprehensive. The dataset of control variables was scraped from the Chinese government, which could not be the actual value, to avoid citizens' anxiety, the Chinese government could hide the real data, which may cause the models to perform not ideally. Additionally, the time series makes predictions by trends over time, so it cannot consider the special cases in the model like policy impact. It is also likely that the investigator will have to make a few changes, such as including or excluding an item to increase the validity of data analyses. Another disadvantage of this methodology is the inability of the researcher to control or change the nature of the sample used by the original researcher. While the researcher can make significant changes to the data collected to align the original researcher's variables and measures with those of the study, such things as the nature of the sample used in the preliminary study cannot be changed or controlled. In addition to issues of validity is the challenge of reliability. Although the researcher can check for any inconsistency in values, thereby heightening reliability and representativeness, it is also likely that the original researcher came up with very different conclusions from those the study will arrive at. This divergence in conclusions could result from different agendas in both studies. For instance, the Chinese National Bureau of Statistics may have collected the data with a political or economic agenda. Therefore, these biases may be transferred into the study if the investigator is not keen.

Because Covid-19 will not disappear fast, this research can provide a guideline to the Chinese government and policymakers on which region should put more effort into reducing the unemployment rate. In addition, the model can make government predict future unemployment conditions and do better preparation. Although every country has different policies and conditions for the academic field, this work can still provide them with some ideas about analyzing Covid-19 regionally.

***Data***

The project's goal is to figure out the impact of Covid-19 on the employment rate of provinces in China. The hypothesis is that Covid-19 will hurt the employment of provinces in 2020, which means it will increase the unemployment rate. Many factors will impact employment besides Covid-19, so I choose them as control variables. Examples are the development level of provinces, the price level of production, the education level, and the change in the province's population (Zeng, 2020). To measure these concepts, I will use GDP to measure general development level, CPI to measure price level, the graduation number of universities and the college (Note: college is for students who failed to enter 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, the second industry increased income, and third industry increased income, total retail income to measure economic development. To measure the impact of Covid-19, I will use the number of Covid-19 confirmed, the number of deaths, and the number of recoveries in each province in 2020 as my dependent variable.

There are two parts of the research data collection: Covid-19 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 increase rate, and the first industry increased income, the second industry increased income. The third industry increased total retail income from the Chinese National Bureau of Statistics from 2008 to 2020. For this data source, I will run a data crawler to scrap the data from the National Bureau of Statistics because it is impossible to download and combine by hand, the data of each province are separate on 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 scraped the website, it blocked me several times because thousands of pages of data needed to scrape one by one. To solve this problem, I first added random breaks in the scraping code and then used three computers with different IPs to scrape the data. After collecting the data, I used the terminal to combine all CSV documents.

This residence population dataset is updated yearly with four columns: Year-end population, region for province name, time for data collection, and value for the total resident population. The mean value of the population is 4407.40 million. The minimum value is 292.33 million, while the maximum population is 12,489 million, and the median value is 3823.52. The CPI of province-level data, including four columns: CPI for the province, name of the province, time of collection, and the value of CPI, is updated for half a year. Hence, this dataset needs to apply the mean value of the first half and the second half of the 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 numbers of universities and colleges also includes four columns while the first column becomes the predicted graduation number, and the unit of data is ten thousand. The average number of graduates from university has a unit of a 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 graduates from college has a unit of a 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, and 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's 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's 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 increasing rate is 5.34, the maximum value is 11.47, the minimum is -1.01, and the median value is 5.54. The data on the unemployment rate in China by province from 2008-to 2020 is found in the Chinese National Bureau of Statistics either. The average unemployment rate is 3.34, the maximum value is 4.57, the minimum is 0, and the median value is 3.41.表格

描述已自动生成

*Figure1: Variables Description*

The dataset includes 31 provinces in China. For most provinces, the unemployment rate increased rapidly in 2020, while a few provinces in the southern part of China have a decreasing trend.

Second, the Covid-19 data is collected from Harvard, and the dataset is called China Covid-19 cases which includes the Covid-19 cases from 2020.01.15-2021.01.15. This dataset of Covid-19 was recorded daily, and data of confirmed, death, and recovery are separate in three worksheets, so I need to count the total number of confirming, death, and recovery number in each province. After collecting all the 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 increase rate, the first industry increased income, the second industry increased income, the third industry increased income, total retail sales, and row is the year. The dataset has no NA value, and it is recorded in detail, so it does not need 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, the natural population increase rate, total retail sales, the first industry increased income, the second industry increased income, the third industry increased income on province-level from Chinese National Bureau of Statistics are lacking authenticity because Chinese government may conceal accurate data to maintain social stability. The study will rely on secondary data for analysis: Covid-19 and control variables data. This secondary data will be sourced from various sources, including the Chinese National Bureau of Statistics. One of the advantages of this data collection method is that it relies on already collected data and is, therefore, less time-intensive and cheap. Rather than collecting primary data from actual participants, which would be expensive and take much time, the study will employ data that a relatively credible agency has already collected. There will be no need to spend time or resources on otherwise necessary research processes such as getting approval from relevant research oversight bodies. Statistical data is also advantageous because the patterns and correlations of interest are visible and apparent. The study involves a topic of interest to many agencies similar to the Chinese National Bureau of Statistics. Therefore, the researcher can create a setting for the investigation. Besides, the data the study will rely on has been taken from extensive samples, allowing for high generalizability. This strength is also tied to the ability of the researcher to use and re-use data sets to ascertain various variables: the researcher can check for any inconsistency in values, thereby heightening reliability and representativeness.

***Method***

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

I utilized the regression model to analyze the causal effects (Aronow & Samii, 2015) between Covid-19 and the unemployment rate by checking the p-value and coefficient of variables in OLS regression. The time series model is used to make a prediction of the unemployment rate in 2020 if Covid-19 never exists based on the trend line from 2008 to 2019. The time series analysis is a method of analyzing a sequence of data set gathered over a specified time frame. I chose the time series model because of its usefulness in cleaning data: time series analysis is advantageous in that it allows the researcher to remove outliers and so filter out the noise as well as apply different averages to obtain a broader perspective of the data. The time series model is also good at uncovering patterns in data and so improve understanding of a data set. I chose the dataset from the Chinese National Bureau of Statistics because it is the most comprehensive and relevant of datasets available: the broad scale of the dataset allows the researcher to compare Covid-19-related employment impacts across China’s provinces. 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 the time series model to make a prediction of the unemployment rate in 2020 without Covid-19 effects instead of the regression model is the data of control variables in 2020, which needs to be plotted into the regression model, have already been affected by Covid-19, for example, the GDP, CPI, and the total retail sales because of the lockdown policy.

After making the prediction, there exists a gap between the predicted unemployment rate and the true unemployment rate of each province. The gap can be treated as the impact of Covid-19 on the unemployment rate. Then the gap value becomes the dependent variable of the regression model, and Covid-19 variables will also be added as independent variables. To satisfy the assumption that the variables should be mutually independent, I calculate the correlations and draw the Heatmap (figure1) to make it more visualized. The correlation over 0.9 is removed or added together to analyze the causal effects between Covid-19 and the unemployment rate in China at the provincial level, and this project ran a regression model (Aronow & Samii, 2015). 日历

描述已自动生成

*Figure 2: Heatmap of correlation*

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

At this time, we can check whether Covid-19 is statically significant through the p-value and feature significant through variables' coefficient in the OLS regression model. Finally, I chose the province with the largest gap in the unemployment rate and the one with the smallest gap to do data analysis. The limitation of the method is firstly on the control variables selection. The economic model for the unemployment rate includes many factors, but I only picked 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 a high R square over 0.85, but few of them are only around 0.25. Hence, the picked two provinces with the largest and smallest gap are not the real province with 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, the income of the household, etc., for 100,000 people in each province. The survey should run in community level so that the analysis could be in updated a micro level. Compared to the non-computational way, my approach can save more time and money because the data I used exists and the non-computational method need too many scholars to help with. Also, it is hard to find a non-biased sample that has a size of 100,000 per province. But the data in a non-computational way is more updated than my approach because the existing data was updated on 2020.8.16. Researchers can have a chance to communicate with participants and improve the research, which can help them make data interpretation comprehensively.

***Results***

Method 1: time series

For the time series method, I only choose the confirmed number of Covid-19 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 the y-axis is the provinces' names. From the graph (Figure 3), Hubei province has the most covid cases, which is 23,087,143, and Xizang province has the least, which is 351.

图表, 日程表

描述已自动生成

*Figure 3: bar chart of the number of covid confirmed*

After an overview of the Covid-19 condition of 31 provinces in China. I used 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 the province's unemployment. The whole sheet will have 31 separate graphs of 31 provinces. 图表, 折线图

描述已自动生成

*Figure 4: Time series prediction of unemployment rate in 2020*

The blue line was the true unemployment rate in 2020, and the yellow line was the predicted unemployment rate. To analyze the gap between the predicted and true unemployment rate, I am wondering if geographic reasons may cause it, so I combine the gap value and map (Figure 5). The darker color means a larger gap, and the lighter means a smaller gap.

地图

描述已自动生成

*Figure5: Time series prediction-true gap on unemployment rate in 2020*

The darkest area with a value of 2.512 is Beijing, and the lightest one with a value of -0.022 is Sichuan province. Beijing is the capital of China which has the largest population immigrants in China. In covid19, migrant workers were most affected by the pandemic because they accounted for the more significant percentage of the urban labor force. They were also concentrated in sectors that were hit the hardest by the disruptive lockdowns. Hence, it is reasonable that Beijing has the largest gap. Although Sichuan is not the province which has the least immigrants but it is a mountainous area that limited the immigration. Generally, the graph shows the larger gap concentrated in the Southern part of China, and the northern part has a smaller gap between the predicted and true unemployment rate. Hence, the covid will also affect by weather, temperature, etc., geographical reasons, and then lead to different impacts on the unemployment rate.

Method 2: Regression

I dropped Hubei province, has most covid19 cases, which is an outlier in the dataset and utilized OLS regression to see whether Covid-19 is statistically significant in the unemployment rate in China of different provinces. (Figure 6)

*Figure 6: OLS regression results*

|  |  |  |  |
| --- | --- | --- | --- |
| Term | Model 1 | Model 2 | Model 3 |
| (intercept) | 0.1866  [0.070] | 0.3218  [0.28] | 0.3018  [0.512] |
| Covid confirm | 4.48  [0.0069] \*\* | 1.182  [0.0179] \* | 2.006  [0.0412] \* |
| Covid recovery |  | -1.344  [0.0180] \* | -2.208  [0.0495] \* |
| Covid death |  |  | -1.07  [0.078] |
| Adjusted R-Square | 0.109 | 0.165 | 0.168 |
| Significant code | \*\*\*<0.001. \*\*<0.01. \*<0.05 | | |

From the figure 6, the adjusted R-square of the model 1 is 0.109, of model 2 is 0.165, of model 3 is 0.168. The p-value of confirmed in model1, model 2 and model 3 is less than 0.05 which shows the number of covid confirmed is significant in the gap value. The recovery in model 2 and model 3 is also statically significant. The death in model 3 has p value which is larger than 0.05 so it is not significant, the hypothesis of the research that Covid-19 promotes the unemployment rate in China can be supported. The result can make sense that the lock down policy is regarding to covid confirmed cases in each province. I divided the test into a nationwide part and a provincial part testing the hypothesis. I calculate the coefficient of variables nationwide by linear regression (Figure 7) and use the time series model coefficient to see the province level one by one.

图表, 瀑布图

描述已自动生成

*Figure 7: Feature coefficient of the unemployment rate*

The recovery, graduations from university, and first industry income were confirmed to relate to the unemployment rate positively. And the CPI, GDP, graduations of college, population, and deaths have a negative relationship with unemployment. The coefficient of recovery is 7.025, of confirmed is 0.066, and of death is -7.098. One unit of recovery increase will lead to a 7.025 unit increase in the unemployment rate. One unit of confirmed increase will lead to a 0.066 unit increase in the unemployment rate. One unit of death increase will lead to a 7.098 unit decrease in unemployment. From the Covid-19 dataset, the covid confirmed is the sum of recovery and deaths. Hence, the number of covid confirmed can represent the condition of the Covid-19 cases. Because of the coefficient of confirmation is 0.66, which proves the hypothesis that covid will promote the unemployment rate nationwide in China.

On the provinces level, from the coefficient of time series model. There are 31 provinces that have different outcomes. The dependent variable is the unemployment rate, and covid cases are independent variables with other control variables. There are 22 provinces with a positive coefficient, which means 22 out of 31 provinces have an increasing unemployment rate and nine provinces have decreasing unemployment. At the provinces' level, the hypothesis was partially rejected, and it was true on the national level. The reason that the impacts were various geographically is temperature and humidity. Based on a previous study, there is a homogeneity regarding temperature and humidity's effect on the seasonal viability and transmissibility of Covid-19. The cold and dry conditions make Covid-19 more likely to spread than warm and wet weather (Mecenas et al., 2020). In the Southeast part of China, the temperature and humidity are higher than in the Northwest part. From the previous graph, it is obvious that the gap value between the predicted unemployment rate and true unemployment in the Northwest part is much larger than that in the Southeast part, which is consistent with the virus characteristics. 地图

描述已自动生成

*Figure 8: Gap value map of China*

Because of the difference of weather condition in northwest and southeast parts, the virus spread much faster and more broadly in the Northwest part of China. The China government started the lockdown policy in these provinces to keep social distance. Accordingly, the Covid-19 has better effects on the economy and society in the Northwest part. More companies in the Southeast were shut down or reduced employees to relieve economic pressure. Hence, the unemployment rate in the Northwest increases significantly, which is consistent with that the gap value between the predicted unemployment rate if Covid-19 never exists and the true unemployment rate with Covid-19 impacts is larger in northwest provinces.

***Discussion:***

The study sought to investigate the impacts of the Covid-19 pandemic on employment rates in China at the provincial level in 2020 using regression and a time series model. After train-testing the model and visualizing the result, the study discovered that the pandemic resulted in widespread unemployment at the national level. However, there were differences in unemployment rates across provinces, where 22 provinces reported increased unemployment while 9 provinces witnessed decreasing unemployment rates in 2020. The Covid-19 impacts were varied and could be split geographically into two parts: Southwest and Northeast. Geographical differences in Covid-19-related effects on employment rates were tied to regional variances in temperature and humidity: the colder and drier climate of the latter resulted in higher infections as well as larger unemployment impacts. Further studies could investigate other factors driving this divergence in unemployment rates across the country. For instance, it would be helpful if further findings indicated that the reason for the geographical division in Covid-19 impacts is influenced by the concentration of certain industries that were particularly affected by the countrywide lockdowns in that region or if socioeconomic factors like education, income, or even age result in one region experiencing greater unemployment rates than the other.

The literature review findings were not so different from those arrived at by the study: the findings suggested that differences in Covid-19-related unemployment rates were founded on regional characteristics. For instance, the literature review revealed that provinces with certain populations and characteristics were more affected than others in terms of availability of employment opportunities. The literature demonstrated that those regions populated by employee-intensive sectors sensitive to disruptions caused by government-imposed lockdowns and breakdown of global supply chains saw greater unemployment levels. Those cities with a large share of their gross domestic product (GDP) in these sensitive sectors, including retail, export, and hospitality industries, experienced the worst unemployment rates compared to cities with many more resilient sectors like finance and healthcare industries. The retail, export, and hospitality industries were greatly inconvenienced by the government's decision to restrict movement as well as the deterioration of global supply chains. They also happen to employ a large proportion of migrant workers, who are often employed informally and likely to be fired without the strictures of legal work contracts.

The economic disruptions of the pandemic resulted in these sensitive sectors shutting down operations or reducing activities. As a result, many migrant workers in these sectors, particularly women, were affected, causing higher unemployment rates in said regions. On the other hand, those cities with a large share of their GDP in more resilient sectors to government lockdowns and disruptions in global supply chains reported lower unemployment rates. The financial, informational technology, and healthcare industries continued to operate even during the peak of the pandemic, and therefore a significantly lower number of workers lost their jobs. Overall, the literature review revealed that divergence in the impact of the Covid-19 pandemic on employment rates across China's provinces was influenced by the province's economic and workforce characteristics. For instance, those provinces with a large number of cities with migrant workers and sensitive employee-intensive sectors were more affected than provinces with a large number of cities with fewer migrant workers or resilient sectors.

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

The regression score and accuracy of time series prediction are not high enough. Time series is better than run regression because the regression model needs to plot the data of control variables in 2020 to predict. Although I have not added Covid-19 impact when I run the regression for the prediction, the data of control variables like GDP, CPI, etc., have already been affected by Covid-19, which could reduce the impact of Covid-19 on the unemployment rate and make a 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 a prediction based on the trends over time.

One of the biggest weaknesses in the study was the relatively few numbers of control variables which significantly limited the accuracy of the regression model. The economic model for the unemployment rate includes many factors, but I only picked 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 with over 0.9 correlation, some variables have a correlation over 0.7. For the time series model, the R squares for each province are different. Most provinces have a high R square over 0.85, but few of them are only around 0.25. Hence, the picked two provinces with the largest and smallest gap are not the real province with 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 the community level. The data could not be updated and comprehensive enough. The covid hospitalization and death rate are concentrated in people older than 70 years old while the population statistics did not show the observed regional difference in covid19 indicators could be led to differences in the age distribution (Holmager et al., 2021). The dataset of control variables was scraped from the Chinese government. Moreover, the primary disadvantage of relying on secondary data is closely related to the issue of validity. The original investigator's aims for conducting the study and collecting the dataset may be slightly different from these studies. It is also possible that the original researcher may have used different variables and measures from the one the researcher wishes to use. Since the data has already been gathered and is inherently founded on the original researcher's variables and measures, validity issues become more prominent when these are different from those of the present study. The data collected may be helpful in the study but not appropriate to guide the study's analyses. The researcher may be confronted with the additional burden of investigating if the measures used by the original researcher are sufficiently similar to those of the present study to allow free use of the same.

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

For future studies, we should put more detailed information. The dataset of covid19 cases should be recorded daily at the community level. The age, gender, and recovery time should also be considered in studying the Covid-19 impacts. The unemployment rate is still too broad if we want to focus on it. Different groups will also have different effects from covid19. What kind of group is more vulnerable? Do older people are more likely to lose their job? Can first-year graduate students not find a job? After we decompose the Covid-19 impacts in detail, it can help government improve the condition fast and efficiently.

***Conclusion:***

The research question was answered by the prediction of the unemployment rate by the time series model and the p-value of the OLS regression model. The p-value of Covid-19 deaths and recovery cases is less than 0.05, which means a statistically significant gap in the predicted and true unemployment rate that Covid-19 has a large impact on the unemployment rate in China at the provinces' level. In addition, to test the hypothesis that the covid 19 will promote unemployment. I used a linear regression model to test the nationwide level and time series to test for the provinces' level. On a nationwide level, the Covid-19 promoted the unemployment rate, and on the provinces level, Covid-19 promoted unemployment in 22 provinces and decreased the unemployment rate in 9 provinces.

The research investigated the impacts of the Covid-19 pandemic on employment rates in China at the provincial level in 2020 using regression and a time series model. This approach was advantageous in that it combined the advantages of both methods including heightened capacity to identify patterns, impute missing values, eliminate outliers, and control for confounding effects thereby allowing for a more comprehensive and thorough analysis. However, the study was limited by inadequate control variables, which may have weakened the accuracy and comprehensiveness of the model. Moreover, the actual reporting of Covid-19 cases in China was done at the provincial level rather than at the community level and so the study was unable to narrow down unemployment characteristics to a granular level. Another weakness was the study’s reliance on data collected by the National Bureau of Statistics of China: depending on statistics from the agency helped save time and resources but also meant that some weaknesses, inherent in the original study, might have been transferred to the research.

Still, even with these shortcomings, the study was able to determine those provincial characteristics (climatic conditions including temperature and humidity levels, sectoral and industrial features, and workforce characteristics) were the underlying reason for differences in pandemic-related employment effects. It identified two provinces with the biggest discrepancy in unemployment rates. By focusing on these two regions, the study was able to categorize several critical factors behind the uneven unemployment rates, brought about by the Covid-19 pandemic, among Chinese provinces. After train-testing the time series model and visualizing the results, the study discovered a general increase in unemployment rates at the national level. However, the study also identified discrepancies in unemployment rates across provinces: 22 provinces in China recorded higher unemployment levels while 9 provinces recorded decreasing unemployment rates in 2020. While the pandemic caused widespread unemployment across the country, this effect was disproportionate in that some provinces were more affected than others. For instance, the adverse impact of the Covid-19 epidemic on employment rates could be geographically divided into two parts: West-Southern and East-Northern. One of the reasons for the difference in unemployment levels during the pandemic was tied to the regions’ disparate temperature and humidity levels.

Previous studies have shown a correlation between regions with lower temperature and humidity with higher transmission and spread of coronavirus: regions with higher temperature and humidity tend to have lower transmission and spread of coronavirus. In the Northwest region of China, the temperature and humidity are much lower while in the Southeast region of China, the temperature and humidity are much higher. This difference in temperature and humidity levels between the two regions explain the discrepancy in Covid-19 infections and by extension the resulting impacts. Consistent with virus characteristics, the Northwest region of China recorded higher Covid-19 infection rates than the Southeast region of China. As a result, the former experienced disproportionate economic and social effects than the latter, including higher unemployment rates. The study recommends that future studies concentrate on exploring in depth this finding, particularly how demographic characteristics such as age and gender tie with Covid-19 impacts.

The literature review indicated that migrant workers, small and middle-sized business owners as well as their employees, together with women were the most affected populations. These groups tend to concentrate in sectors that were greatly impacted by government-imposed lockdowns and disruptions in global supply chains. Future studies should focus on identifying even more specific regional and worker characteristics that contribute to disproportionate Covid-19 impacts on employment rates. Further investigation of these phenomena will help the Chinese government in developing more efficient ways of combating another epidemic while limiting subsequent adverse effects on employment rates.

***References***

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>.

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.

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.

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>

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>.

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/.

Donthu, Naveen, and Anders Gustafsson. "Effects of COVID-19 on Business and Research." *Journal of Business Research*, Elsevier, June 9 2020, https://www.sciencedirect.com/science/article/pii/S0148296320303830?casa\_token=IX\_5yi8dmjQAAAAA%3A3XqotBXh98wVpK02zG2QGXix8G3JzdlQMntQosbMSc7ZRRWh9g2oGZMJrkFba14fcEOfVA\_Q.

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

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

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.

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.

https://docs.google.com/document/d/14nEfaKG8eyPWPbRk3YAoa5N-NLJqtHCPriuMqF2QhoY/edit?usp=sharing

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.

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.

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.

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

Li, Yan, and Maria N. DaCosta. "Transportation and Income Inequality in China: 1978–2007." Transportation Research Part A: Policy and Practice, Pergamon, September 11 2013, https://www.sciencedirect.com/science/article/pii/S0965856413001523?casa\_token=efJWewtEwGkAAAAA%3A\_Rv3oFf8rIM8E17FoZaWJ\_TrK-h5DCnQw8X2\_TJsZqYQvzGxpKwz9WwMsHZHSbenaWC5i92X.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.

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)

[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.

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.

Savvides, Andreas, and Thanasis Stengos. "Income Inequality and Economic Development: Evidence from the Threshold Regression Model." Economics Letters, North-Holland, September 18 2000, https://www.sciencedirect.com/science/article/pii/S0165176500002937?casa\_token=4ADP9XnwUjgAAAAA%3AfEF6oQ71lQFUds7wkYPXROPXdqrsFdzhf4myDoj31fokyhWEZl4u\_Wl\_EVL35Ld2U18IGed7.

Shen, Jianfu, et al. "Covid-19 and Regional Income Inequality in China." Frontiers in Public Health, Frontiers Media S.A., May 11 2021, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8144473/.

Sicular, Terry, et al. "The Urban-Rural Income Gap and Income Inequality in China." SpringerLink, Palgrave Macmillan UK, January 1 1970, https://link.springer.com/chapter/10.1057/9780230584259\_2.

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.

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>.

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.

McClelland, Alison. “Effects of Unemployment on the Family.” *The Economic and Labour Relations Review*, vol. 11, no. 2, 2000, pp. 198–212., https://doi.org/10.1177/103530460001100204.