An In-Depth View of the Economic Impact of COVID-19

Eric Tang, Hussein Hussein, Nicolaus Wong, and Andy Dai

Western University

October 31, 2020

Abstract

The COVID-19 Pandemic has changed virtually every facet of our lives. As the months continue to pass by, we are beginning to witness and understand the devastating long-term economic implications of this pandemic and the various shutdowns that it has necessitated. It has had far-reaching consequences beyond the spread of the disease itself and efforts to quarantine it. There have been supply shortages, lags in international logistics and manufacturing, and shortages of pharmaceuticals that affect even the most prepared of cities.

With the prevalence of the vast amount of data in society today, we have the ability to take a more in-depth and comprehensive look, at the economic impact and ramifications that stem from COVID-19. The aim of this study is to leverage the data that is available to the public, construct models, and determine insights about the economic impact in cities and counties since the pandemic hit. Our methods utilize SIR models to analyze disease transmission , as well as Linear Regression to pull inferences and interpret key drivers of economic impact.

Keywords
COVID-19, Coronavirus, Economic, Unemployment, Data Science, Differential Equations, Compartmental Models, Economic Impact

1 Introduction

At first glance, the coronavirus pandemic appears to be hitting the economy similar to the World War 2 era of battlefield tactics; from all sides and fronts. Spending on all fronts has seen a dramatic crash, with people avoiding public areas like never before. There are shortages of supplies and food in many areas, with technology companies forecasting even more delays to shipments of goods.

However, the effects of the pandemic may not be as simple and uniform as it appears to be. Cities and counties rely on more than 1 industry to function and flourish. With each industry having a different risk factor in response to the pandemic, different cities and industries are being uniquely and independently affected.

2 Methodology

2.1 SIR Model

The SIR model is used to analyze the dynamics of infectious disease transmission. The model divides the population into three categories. An individual can be categorized as susceptible ($S(t)$), infected ($I(t)$) or removed ($R(t)$, dead or cured), denoted by S, I and R respectively, along an independent variable; time [1].

The Susceptible Equation:

$$\frac{ds}{dt} = -\beta si \tag{1}$$

where β represents the infection rate which is the probability per day that an I-person can infect a S-person.

The Infected Equation:

$$\frac{di}{dt} = \beta si - \gamma i \tag{2}$$

The Recovered Equation:

$$\frac{dr}{dt} = \gamma i \tag{3}$$

where γ represents the recovery rate which is the probability per day that an I-person transitions into an R-person; becoming non-infectious permanently.

2.2 Linear Regression

Consider the problem of estimating the linear relationship between two variables y and X. Linear regression models the relationship as

$$y = X\beta + \epsilon, \tag{4}$$

where $y = (Y_1, ..., Y_n)^T$ is an n vector of responses, and $X = (x_1, ...x_n)^T$ is an $n \times p$ matrix with i.i.d. $x_1, ..., x_n$

3 Analysis

3.1 Risk Metric: Exposure to High-Risk Industries

The first step of the analysis is to identify high-risk industries that have been most severely affected by COVID-19. Based on literature review from [2], the high-risk industries are defined such as Accommodation and food services, Finance and insurance, Health care and social assistance, Manufacturing, Real estate and rental and leasing, Retail trade, and Transportation and warehousing. Using United States Census Bureau's publicly available dataset, All Sectors: County Business Patterns by Legal Form of Organization and Employment Size Class for U.S., States, and Selected Geographies (CB1800CBP), a risk metric is constructed based on the employment exposure of high-risk industries for every single county in United States.

The list of the counties is then sorted based on population and the risk metric. An example of top ten counties with the most risk exposure can be seen below:

It is hypothesized that higher value of risk metric should have a positive relationship with the proportion of population that are effected. We demonstrate the effectiveness of this metric by building a linear regression model where our only feature is the exposure to high-risk industries, and our target variable is the proportion of infected for each county on the sorted data.

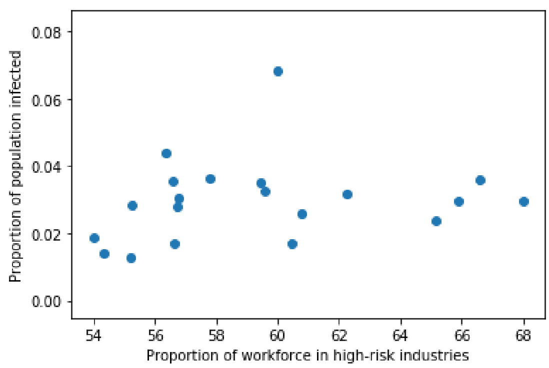

Figure 1: Relationship between exposure to high-risk industries and proportion of infected.

Figure 1 displays the relationship between our feature and target variable using the top 20 counties ranked based on population and risk metric. It can be observed that a significant linear relationship exist between the variables. This is reflected from the linear regression model as it achieved a R-squared of 0.866.

County	Risk Metric (%)	Infection Proportion(%)
Kings	68.00	2.96
Clark	66.58	3.60
San Bernardino	65.89	2.95
Wayne	65.14	2.39
Tarrant	62.26	3.18
Franklin	60.77	2.57
Cuyahoga	60.48	1.72
Miami-Dade	60.02	6.83
Bexar	59.58	3.27
Maricopa	59.46	3.52
Cook	57.80	3.62
Los Angeles	56.78	3.05
Philadelphia	56.72	2.81
San Diego	56.65	1.69
Suffolk	56.57	3.57
Broward	56.34	4.41
Hennepin	55.27	2.85
Allegheny	55.23	1.28
Alameda	54.35	1.42
Orange	54.00	1.88

Table 1: Exposure to high-risk industries per county.

3.2 Disease Transmission Forecasting

Given the top 20 counties ranked above, a SIR model is built for each county to predict the future infected cases. Figure 2 displays Hennepin county's predicted trends of COVID-19. Table 3 shows the prediction of future cases in one year, two years, and five years.

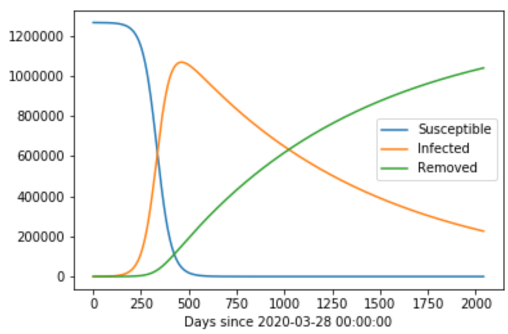

Figure 2: Hennepin SIR Results.

3.3 Economic Impact Inference

Given the top 20 counties, a linear regression is built for each county to model the relationship between its unemployment rate and the monthly changes in COVID-19 cases. To measure the strength of economic impact of COVID-19 to each county, we use the R^2 metric of each regression as an approximation.

In the previous two sections, we first predicted future infected cases using SIR models, then we build linear regression models to understand the impact of COVID-cases in each county's economy. In this section, we integrate the two models and use the predicted future cases to infer the predicted future unemployment rates in one, two, and five years.

4 Findings

In Risk Metric: Exposure to High-Risk Industries, we defined high-risk industry groups and noticed strong linear relationship between the population working in high-risk industries and infection cases for each county. The regression model was able to achieve a R-squared of 0.866, which implies that the exposure to each industry is an important factor to be considered when dealing with COVID-19 modeling. We believe the reason behind this relationship is because it is very challenging task for a county to quickly perform radical changes to its economic structure to deal with pandemics like COVID-19. In many of the high-risk industries defined previously, a common theme that shared by these industries is the high human density working environment - which promotes human interaction, and hence transmission of diseases.

In Disease Transmission Forecasting, we used SIR, a widely used compartmental model, to understand and predict the future dynamics of COVID-19 in each county. In particular, it can be observed that Alameda, Allegheny, and San Diego showed +40% potential increase from the existing levels in one year. Multiple other counties, such as Franklin, Los Angeles, and Hennepin showed +20% increase. While the models predict a continued increase in COVID-19 cases, most of the cases is estimated to fall back lower in two and five years.

In Economic Impact Inference, linear regression models are deployed to understand the relationship between unemployment rates and COVID-19 cases. The goals was to investigate whether or not COVID-19 has a significant impact on a county's economy. Several counties showed significant R^2, meaning that a high proportion of the variance for the dependent variable is explained by our independent variable. Notably, counties such as Cook, Los Angeles, and Hennepin showed R^2 over 0.6. Implying that the unemployment rate at each of these counties has been largely driven by the impact of COVID-19 since the inception of the disease.

The goal of the research is to identify the county that will feel the impact of the economic impact of COVID-19 the most. Hence, a combination of the results from all three sections must be considered. One county that stands out is the Hennepin: 26.29% expected growth in infection numbers and a significant R^2. We expect Hennepin to be affected by COVID-19 economically the most among the sample space of our research.

Conclusion

In this research we identified and proved the significance of the relationship between exposure to high-risk industries and COVID-19 infection cases. For each county, we aggregated the employment in Accommodation and food services, Finance and insurance, Health care and social assistance, Manufacturing, Real estate and rental and leasing, Retail trade, and Transportation and warehousing. The counties are then ranked nationwide to identify the counties that are more susceptible to COVID-19. High exposure to working at the given industries imply a difficult transition to adapt to pandemics. Different from other industries, the high-risk industries require frequent human interaction and are prone to transmission of diseases. The strength of the metric is then

By using the SIR model, we have been able to predict and forecast the three components of population: susceptible, infected, and removed. Observing the predicted trends through a cross-section persepective, we noticed that the COVID-19 cases are expected to continue increase over the next two years before decreasing.

Increase in number of cases is not a direct indicator of economic impact - different economies have different structures and are affected by many different internal and external factors. In order to understand which county will feel the most economic impact, the relationship between COVID-19 cases the health of economy must be studied. To achieve this, we used unemployment rates of each county as a approximation for the overall level of economy. Linear regression models are then built to predict unemployment rates using infection data.

Using a combination of three methodologies, we conclude that Hennepin will be impacted the most by COVID-19 among its peers.

County	1 Year Prediction	2 Year Prediction	5 Year Prediction
Kings	834292	188331	2131
Clark	1826728	1386110	605391
San	1817524	1434344	704761
Wayne	552351	129033	1641
Tarrant	1767847	1423121	742358
Franklin	1074017	809736	346440
Cuyahoga	987018	690134	229747
Miami-Dade	2314818	1892299	1031851
Bexar	1588545	1174503	474453
Maricopa	3392408	2391773	838069
Cook	3555650	2196454	517259
Los Angeles	7294924	4908251	1494895
Philadelphia	1154026	715479	165807
San Diego	2964686	2472180	1421688
Suffolk	612684	410786	118117
Broward	1604576	1254819	599963
Hennepin	985987	683962	226160
Allegheny	981519	701385	254055
Alameda	1483843	1230895	692041
Orange	2667123	2056877	937994

Table 2: Prediction results from SIR models.

County	R^2
Kings New York	0.1904
Clark Nevada	0.0032
San Bernardino California	0.4242
Wayne Michigan	0.0744
Tarrant Texas	0.0053
Franklin Ohio	0.2515
Cuyahoga Ohio	0.0684
Miami-Dade Florida	0.3644
Bexar Texas	0.0006
Maricopa Arizona	0.1118
Cook Illinois	0.6598
Los Angeles California	0.6289
Philadelphia Pennsylvania	0.4763
San Diego California	0.2077
Suffolk Massachusetts	0.0008
Broward Florida	0.0708
Hennepin Minnesota	0.6777
Allegheny Pennsylvania	0.0824
Alameda California	0.2261
Orange California	0.2587

Table 3: Prediction results from SIR models.

Acknowledgements

We would like to express deep gratitude to Raahim Salman at Western University for valuable suggestions in the development of this work.

References

[1] David Smith and Lang Moore. The sir model for spread of disease - the differential equation model, Dec 2004.

[2] Fabian Stephany, Niklas Stoehr, Philipp Darius, Leonie Neuhäuser, Ole Teutloff, and Fabian Braesemann. Which industries are most severely affected by the covid-19 pandemic? a data-mining approach to identify industry-specific risks in real-time, 03 2020.