Continuous Intelligent Pandemic Monitoring

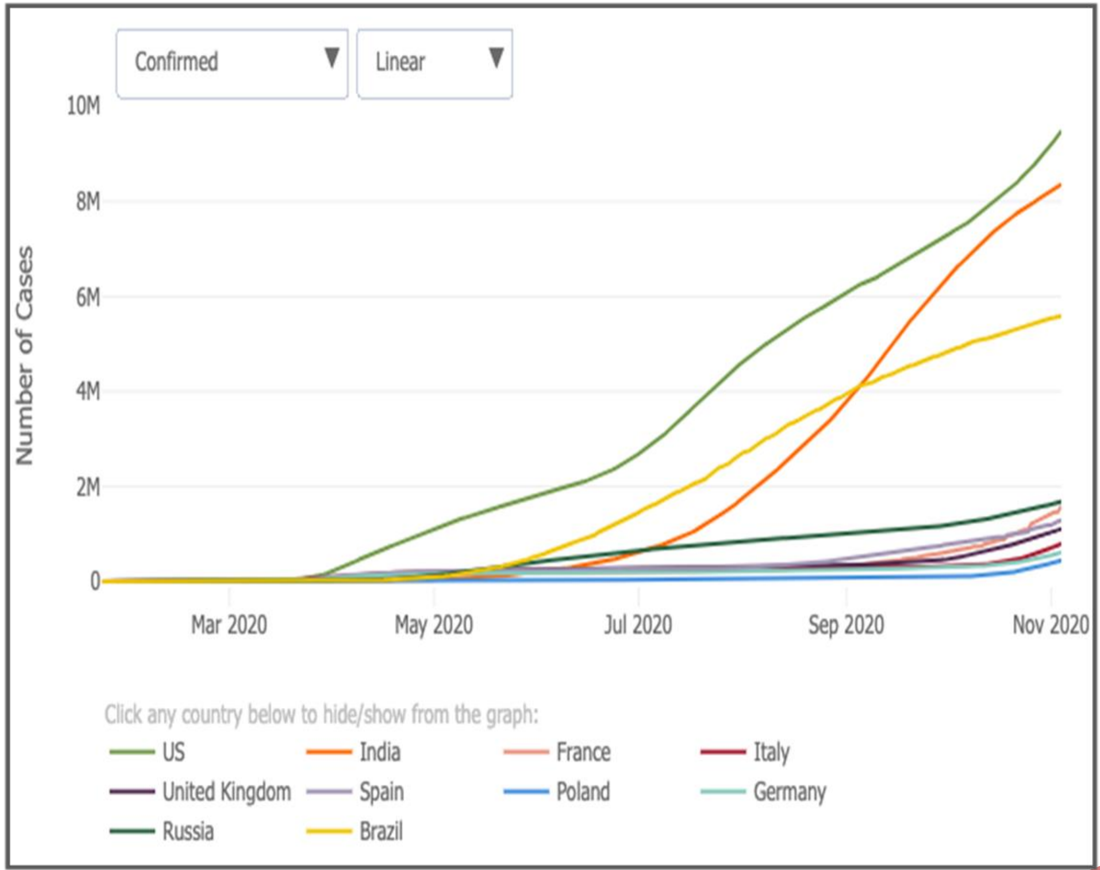
Huijue Kelly Duan, Hanxin Hu, Miklos Vasarhelyi

Accounting Information System

Rutgers, the State University of New Jersey

Introduction

- Global and US COVID situation
- Concerns over the accuracy of the numbers, the test
  - ❖Molecular test / PCR tests (Nasal/Throat swab or saliva)
    - FN (2% and 37%)
  - ❖Antigen test (Nasal/Throat swab): cheapest and fastest
    - FN (0% - 50%)
  - ❖Antibody tests (Blood test)
    - FN (0 – 30%)
  - ❖Large numbers of asymptomatic cases
    - Diamond Princess cruise ship passengers (46.5%)
    - Prison inmates (96%)
    - Poston homeless shelter (87.8%)



Source: Johns Hopkins Coronavirus Resource Center

Use measurement science (accounting), and assurance science (auditing) to examine the situation

Research objective

Continuous Intelligent Pandemic Monitoring (CIPM)

❖Utilizing various exogenous data to perform predictive analytics to validate the official disclosed epidemic numbers

❖Performing cross-sectional analytics to identify significant variables that could impact the disease severity

❖Assessing the disease severity level by utilizing Clustering approach

❖Providing guidance for policymakers based on simulations of different preventive actions

Model construction

Time Series Model Prediction

- **Model 1**
- Utilize different exogenous variables to predict confirmed cases, mortality, percentage of positive test
- Compare the predicted number to reported number to determine the reasonableness of public data

Clustering

- **Model 2**
- Utilize government's open data to identify demographic features that could have significant impact on the pandemic
- Incorporate migration data to the model
- Clustering approach
  - ➢Categorize counties into different clusters based on significant features
  - ➢Measure the centroid point distance to determine the disease severity of the counties
  - ➢Identify the counties that are highly susceptible to disease severity

Simulation

- **Model 3**
- Utilize identified significant factors and regional characteristics to simulate the Epidemic models (SEIQHRF)
- Provide guidance to policy makers
  - ✓Masks requirements
  - ✓Social distancing

Model 1. Time Series Model Predictions

1

Time series model: ARIMA

❖ With 30-day sliding window approach to assess the reasonableness of the number

- Confirmed case
- Mortality
- The percentage of positive cases

2

Dataset time period

❖ 9/5/2020 – 10/4/2020 training set

❖ 10/5/2020 – 10/11/2020 testing set - prediction window

3

Endogenous and Exogenous Data

❖ Endogenous Data
- NYC Open Data
- NJ COVID-19 Information Hub
- Johns Hopkins University data portal

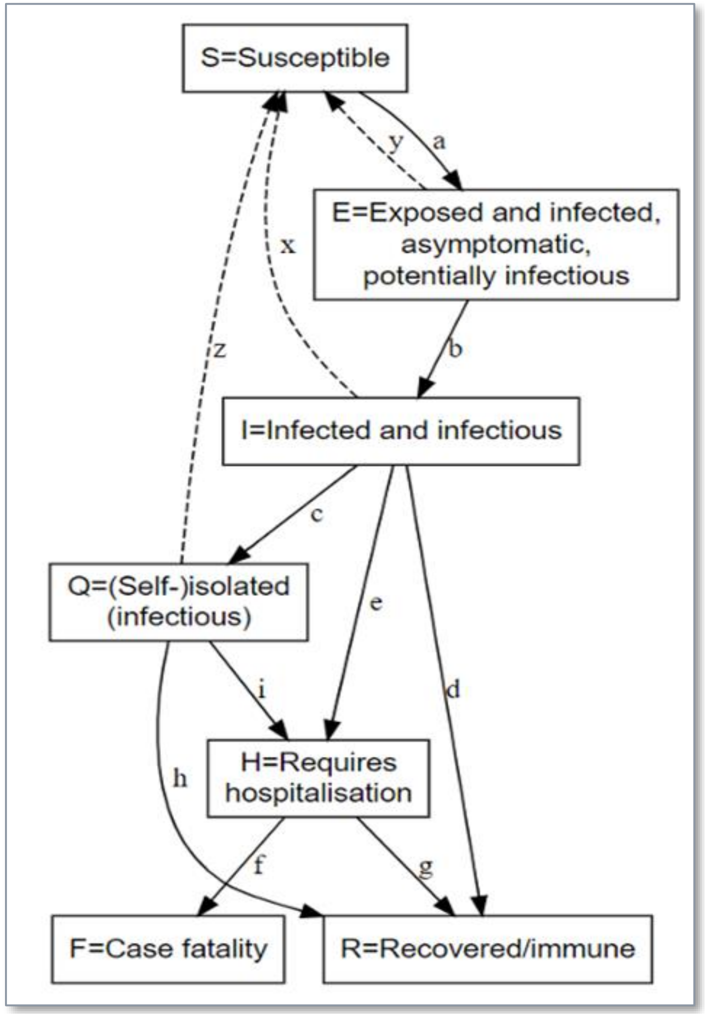
❖ Exogenous Data
- Google trends
- Apple Mobility report
- Subway Turnstile
- OpenTable
- Daily News Economic Sentiment Index

Model 2. Clustering Approach

- Perform cross-sectional analytics to identify potential high risky cities by using clustering analysis
- All counties in NY & NJ
    - Population
    - Population Density
    - Persons age 65 years and over, percent
    - Average household income
    - Persons in poverty, percent
    - Persons per household
    - Persons with a disability, under age 65 years, percent
    - Community Resilience Estimate
    - Mobility Data

Model 3. Epidemic Simulation

- Using SEIQHRF model to simulate the impact of preventive policies



Compartment	Functional definition
S	Susceptible individuals
E	Exposed **and** infected, not yet symptomatic but potentially infectious
I	Infected, symptomatic **and** infectious
Q	Infectious, but (self-)isolated
H	Requiring hospitalisation (would normally be hospitalised if capacity available)
R	Recovered, immune from further infection
F	Case fatality (death due to COVID-19, not other causes)

Time Series Model - NYC

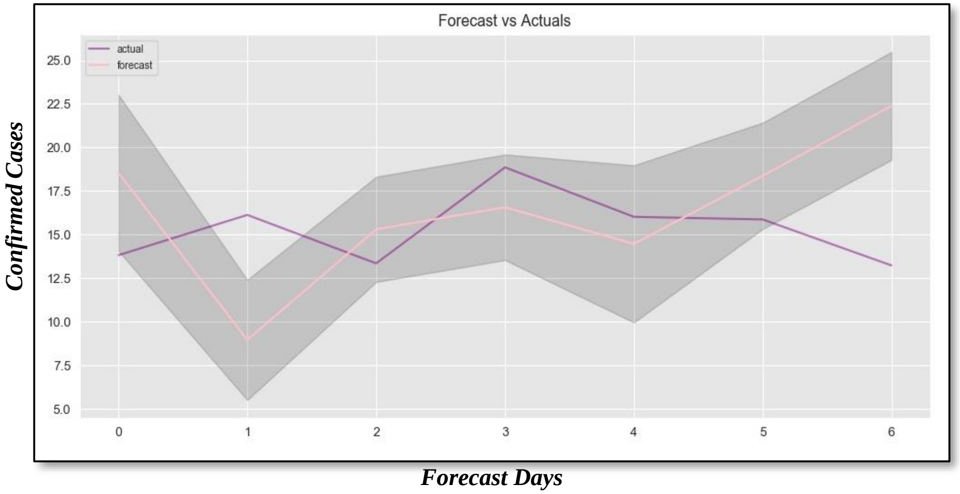


Figure 2-1: Predicted confirmed cases vs Actual confirmed cases

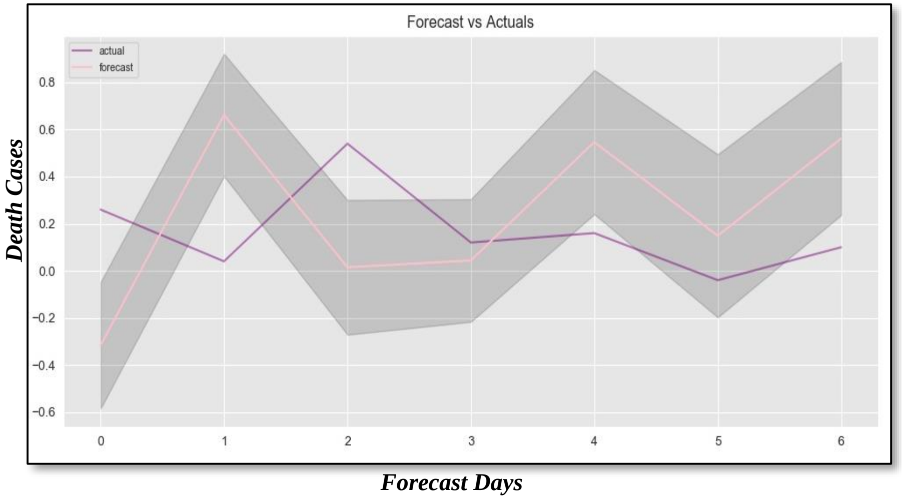


Figure 2-2: Predicted death cases vs Actual death cases

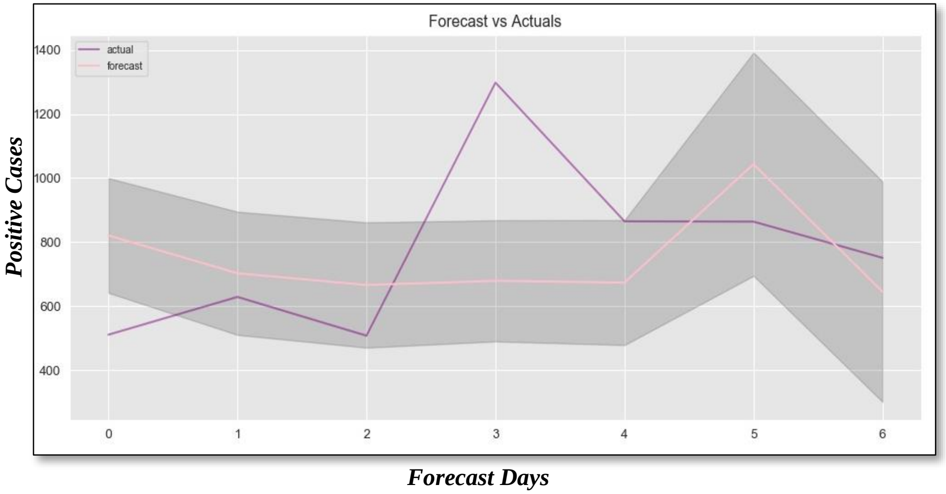


Figure 2-3: Predicted Positive Test Cases vs Actual Positive Test Cases

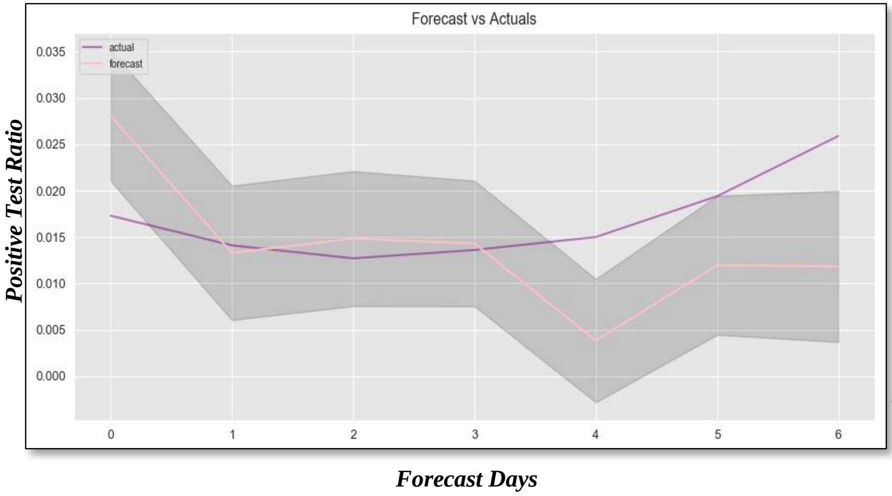


Figure 2-4: Predicted Positive Test Ratio vs Actual Positive Test Ratio

Time Series Model - NJ

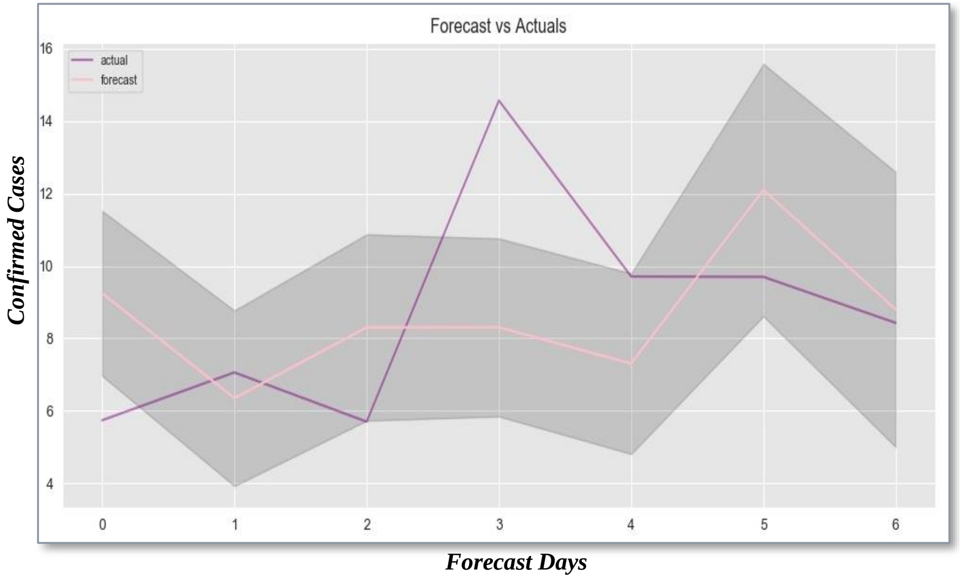


Figure 3-1: Predicted confirmed cases vs Actual confirmed cases

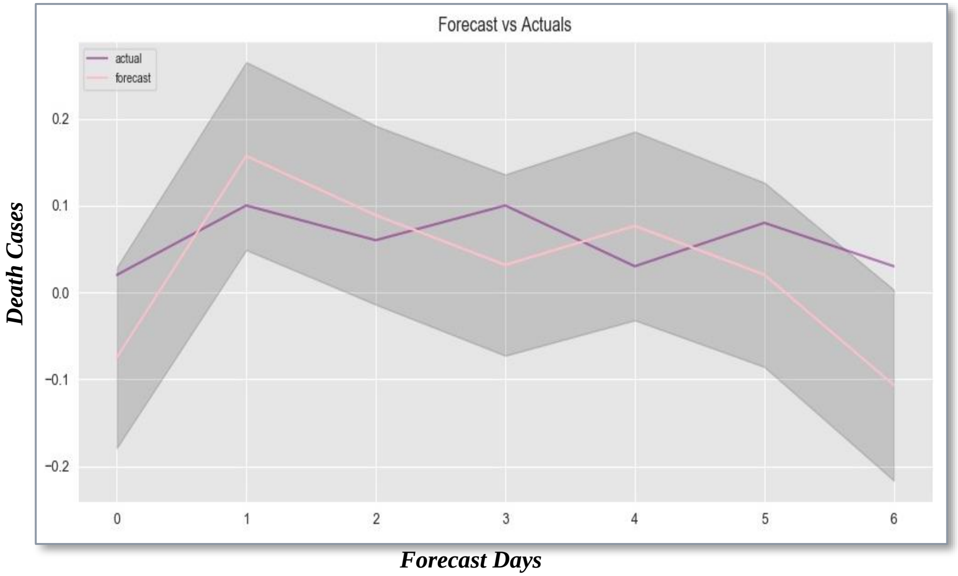


Figure 3-2: Predicted death cases vs Actual death cases

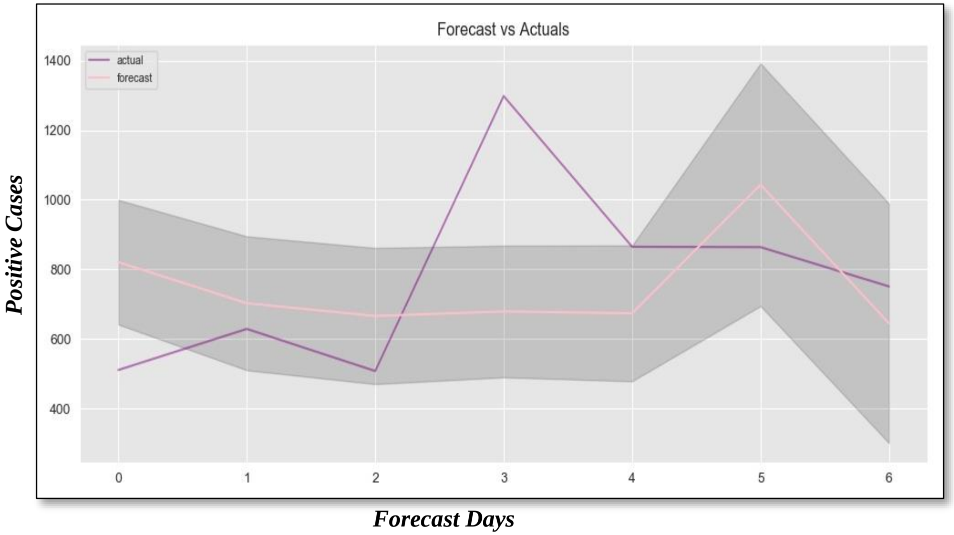


Figure 3-3: Predicted Positive Test Cases vs Actual Positive Test Cases

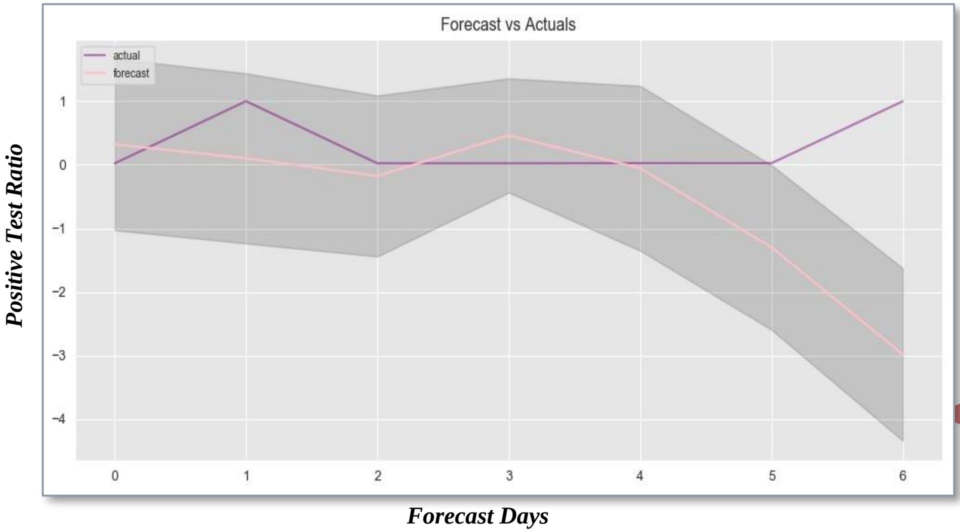


Figure 3-4: Predicted Positive Test Ratio vs Actual Positive Test Ratio

Clustering

- Method: according to Silhouette Score, the proper number of clusters is equal to 2; then K-means is conducted to identify peer city groups that may have similar vulnerability to the pandemic.

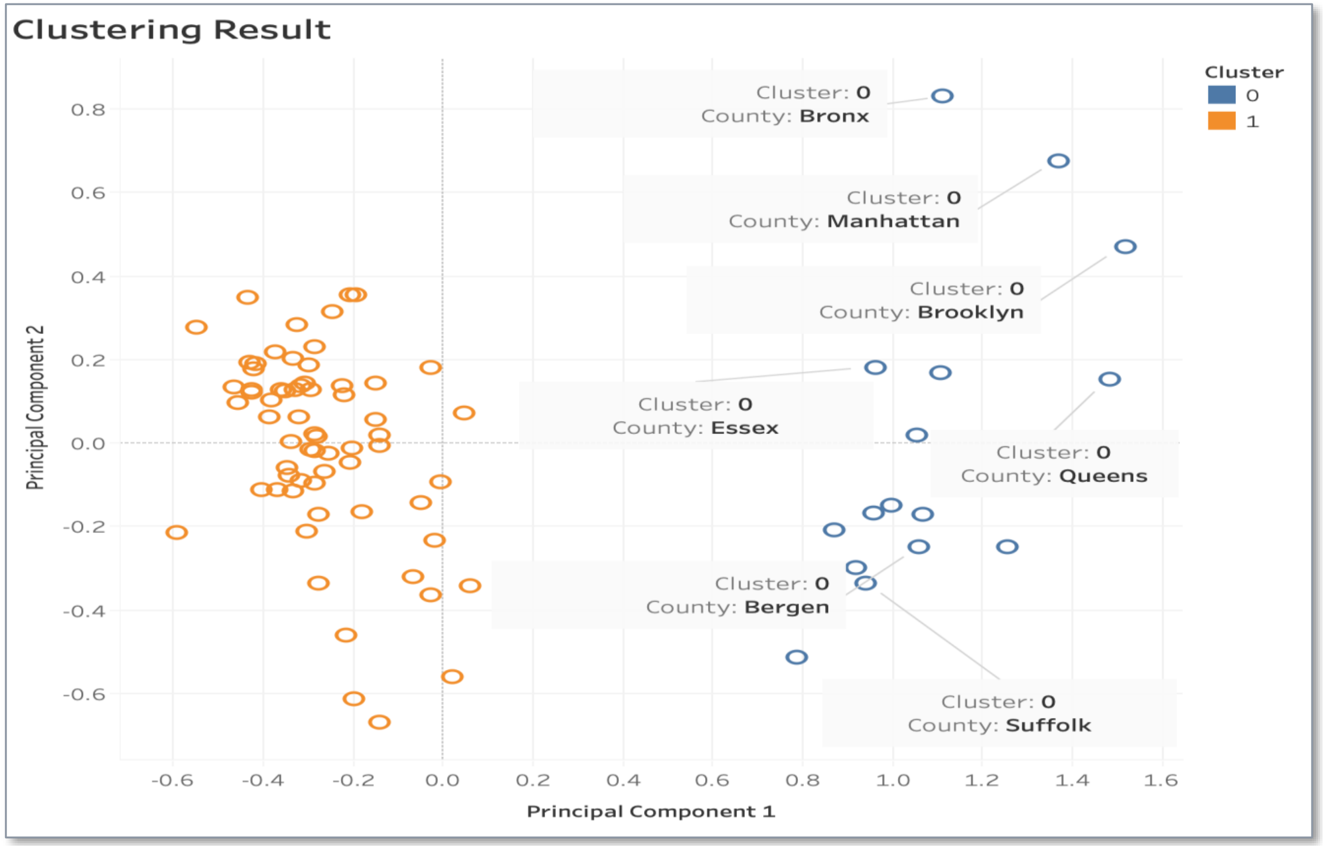


Figure 3: Clustering results

Clustering

- <u>Which city group is more susceptible to disease severity:</u> *we use the confirmed cases (as of 10/12) as the metric to measure disease severity*

Variable	Definition
Cluster_distance_0	The distance between the data point and the centroid point of "Cluster 0"
Cluster_distance_1	The distance between the data point and the centroid point of "Cluster 1"
Cluster ID	The Cluster ID: 0， 1

- <u>Regression results:</u>

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)          9.2448     1.2301   7.515 8.13e-11 ***
Cluster_distance_0  -3.9827     1.9240  -2.070   0.0418 *  
Cluster_distance_1   2.7552     1.5153   1.818   0.0729 .  
Cluster1            -0.5989     1.0039  -0.597   0.5526    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Clustering

- Method: according to Silhouette Score, the proper number of clusters is equal to 2; then K-means is conducted to identify peer city groups that may have similar vulnerability to the pandemic.

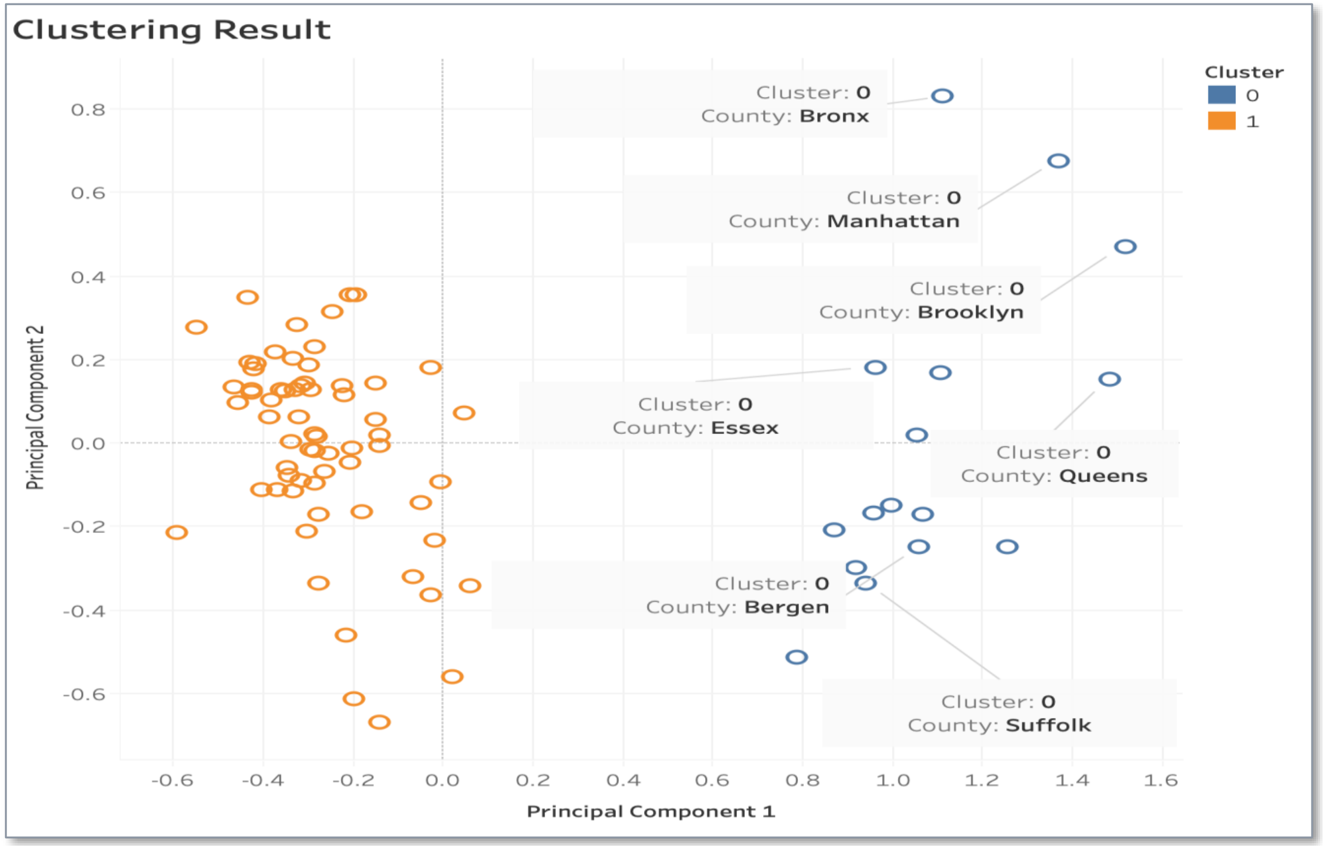


Figure 3: Clustering results

Simulations

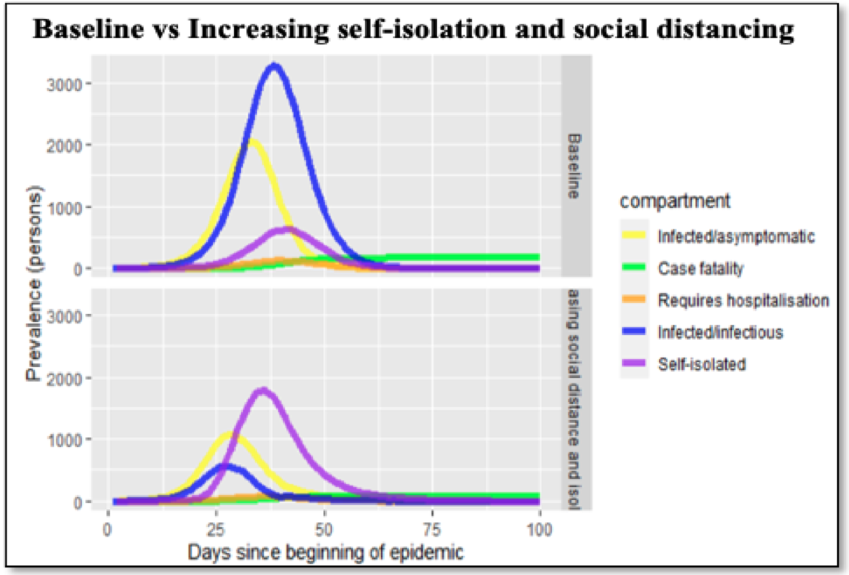


Figure 4.1

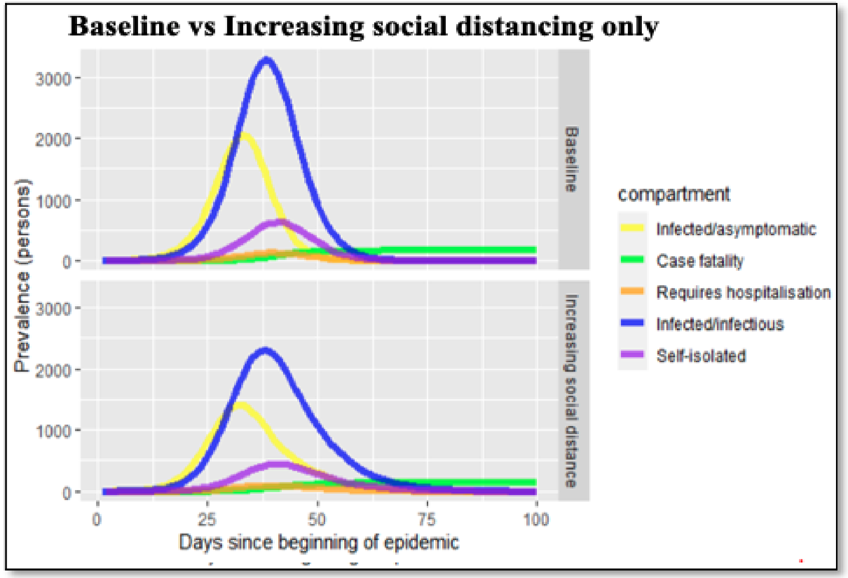


Figure 4.2

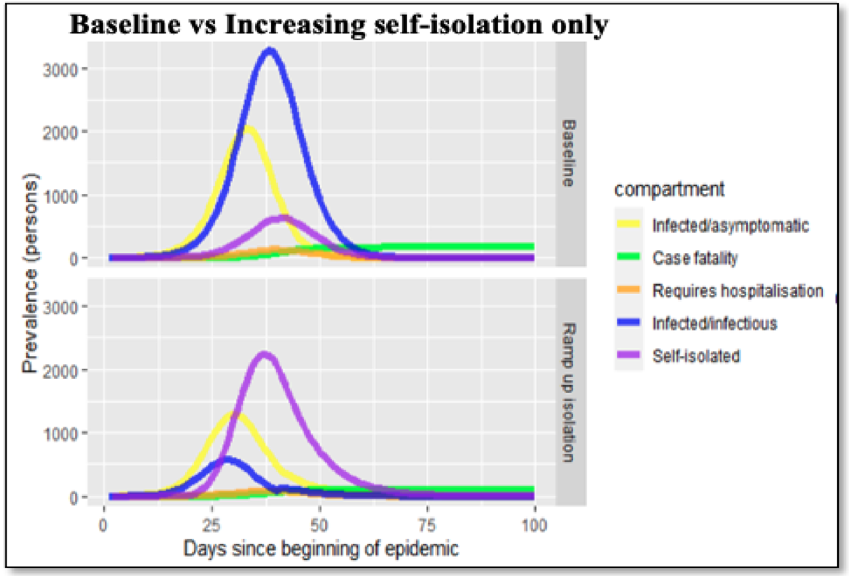


Figure 4.3

Figure 4: Use SEIQHRF model to simulate the impacts of different social interventions policies assuming the social intervention policies are implemented on Day 20. When enforcing self-isolation and social distancing, we can better control the transmission of COVID-19.

Contribution & Conclusion

- Examine the pandemic situation from the perspective of accountants and auditors
- CIPM can be used to continuously monitor the pandemic and generate alerts
- Support government decision-making
  - Validating the reasonableness of the current epidemic numbers by utilizing exogenous data
    - ❖ Unreasonable trend of recorded information could have significant negative impact on decision making
  - Assessing the disease severity level by utilizing Clustering method
    - ❖ Pinpoints the specific cities that are highly susceptible to disease severity
  - Providing guidance for policymakers based on simulations of different preventive actions
    - ❖ Provide simulation results based on current and forthcoming policies
    - ❖ Illustrate the timing impact of the policy implementation

Thank You!

Contact us at:

kelly.duan@rutgers.edu
hanxin.hu@rutgers.edu