COVID-19 United States County Level Prediction and Policy Analysis

Wenqi Chen, Xingyu Liu, Jiahui Tang, Xiaohan Yang

December 13, 2020

Abstract

With the severe public health problem caused by COVID-19, our work proposed an effective approach to evaluate the influence of different policies on the future epidemic spread. On the one hand, we constructed county-level prediction models with Ridge Regression, SEIR and RNN based on the geographical, social-economics as well as the behavioral features, among which RNN reached MAPE of 0.077. On the other hand, we constructed state-level causal inference model for policy analysis and provided solid analysis and recommendations for policy design. On top of that, we combined the prediction and policy analysis together to demonstrate how the number of future cases will change and evaluate the influence of policy implementation on the cases.

1 Introduction

COVID-19 is at present one of the most important issues around world, as well as in the US. In order to mitigate the spread of COVID-19, various social interventions are taken place by different states in the US. Among those policies, the quarantine strategies and occupancy limit for business lies at the center of the trade-off between economic development and epidemic control.

Therefore, a vital problem is to evaluate the effects of different strategies and predict the future infectious cases based on these strategies, offering a valid recommendation for the government. What's more, with the solid predictions and causal analysis, it can also encourage public to be actively involved and support the strategies.

2 Problem Statement

In general, our project question is to focus on investigating the impact of different policies under different circumstances as well as predicting the future confirmed COVID-19 cases in different counties. In particular, we are interested in with the prediction of the future cases, how extending or lifting the policies, especially the different levels of limiting policies, will change the future number of COVID cases in different counties

More specifically, we intend to investigate or consider the following questions through our model.

1. **How will the number of cases change in the future in different counties?**

   Different counties have different patterns, due to the influential features of different counties. The final goal of our project is to reasonably predict the future change of cases with the factors. And to address this problem, we will take the following questions into consideration: How do the cases of surrounding counties influence the target county? How does the profile of the different counties affect the spread of COVID? What may be the important factors? How will people's behavior affect COVID spread. In this way, we will not only treat it as a time-series problem, but also customize the unique characteristics of the counties.

2. **How do the different levels of policies affect the spread of COVID?**

   Apart from the direct influence of policies on epidemic control, from figure 2 and the implementation of different policies, we found that the policies also can have a significant influence on people's behavior, which will indirectly affect COVID spread. Then the next step is to explore the detailed relationship between the social distancing factors and the policies and then take the total effect of policies into consideration.

3 Data Description

To address this question, we decide to conduct our analysis by deriving data from four different data categories, including Covid facts, mobility information, county profile, and relevant policies data. After obtaining and preprocessing the data, We then combine the datasets from different sources and try to explore the relationship between each other.

Table 1: Data Source and Description

Dataset	Source	Properties	Date Range
Covid Facts	USAFacts	daily numbers of confirmed cases	2020/1/1 - 2020/10/31
mobility info	aggregated GPS data from Maryland Transportation Institute	daily information about stayed-at-home population and total trips per county	2019/1/1 - 2020/10/31
social-economics data	US census, American Hospital Association, CDC Social Determinants of Health, Bureau of Economic Analysis, and United States Department of Agriculture Economic Research Service	county-level social-economic facts including hospital beds number, income, education, different age percentage, employment and so on	updated in 2018
policy data	KFF and the Berkeley group	state-level policies relating to Covid	updated by Oct.,2020
geographic info	aggregated location data from Prof. Benjammin Skinner at Florida University	list of adjacent counties	updated by 2019

4 Exploratory Data Analysis

4.1 COVID data

Figure 1 shows the geographical relationship between the new cases in different areas, indicating the cases mainly aroused from west and east coasts and then the newly confirmed cases mainly came from the surrounding states.

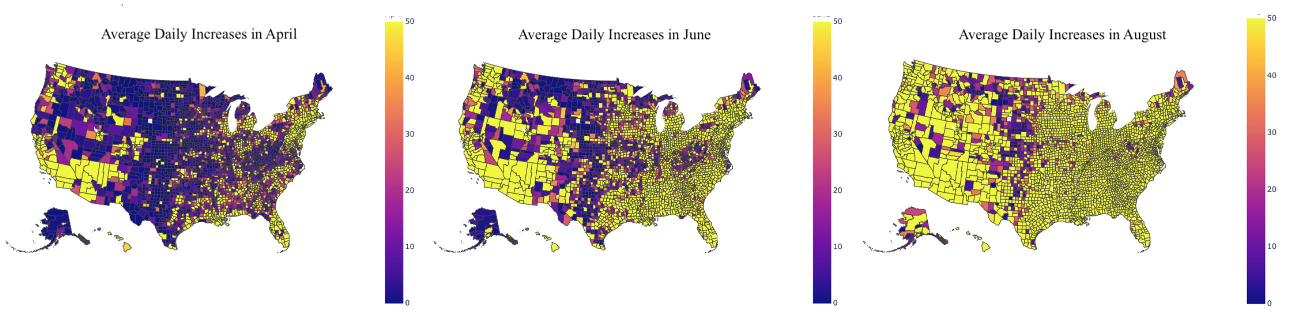

Figure 1: Heatmaps of monthly averaged cases(snapshots of April, June, and August)

Figure 11 shows the population distribution of different counties in US. Comparing 11 with 1, it shows the population is strongly correlated to the newly confirmed cases in different counties. Moreover, when we inspect data, we found that there are 418 counties that have negative increases, which is probably due to the data collection error (see Figure 12 for example). As our goal is to investigate the policy effect on large, we decide to drop those counties for simplicity.

4.2 Mobility data

From Fig 2, we can tell the difference of people's mobility between the pandemic period and the normal time. Moreover, we also notice that the SAH (Stay-at-Home) ratio during March to June is the highest. The drastic increase at the end of March is probably related to close-down policies, while the drop at the end of May could be explained by Black-live-matters protest after George Floyd's death in May 26th. Moreover, this graph also shows the differences among each states. In this dataset, we find certain counties have some missing values for one or two days. Thus, we impute the time series values for each county by k-nearest neighbor.

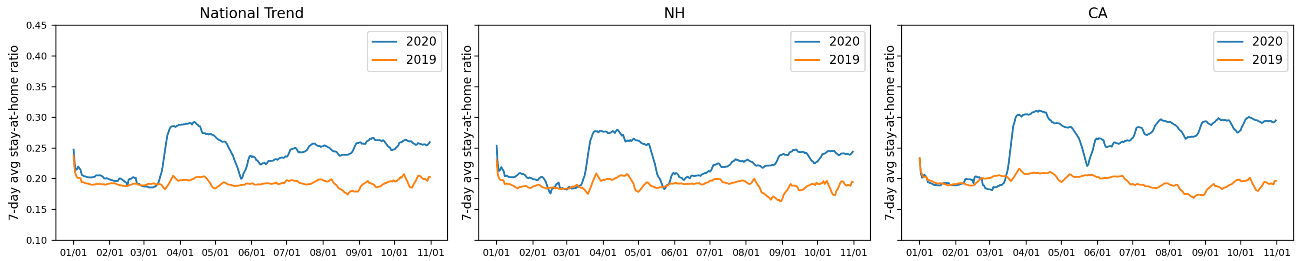

Figure 2: Figures of states' average stay-at-home ratio in 2019 and 2020 (see the full graph in Appendix A)

4.3 Social-economics data

The social-economics data are updated in 2018, and they would generally remain unchanged during the COVID period. Among all the social-economics features, We hope to find the most important ones and use them as static features to help train the model.

Since we will predict COVID confirmed cases in October, we first average the COVID confirmed cases number in September per county, and calculated the number of confirmed cases per 1000. Next, we calculate the correlation between 'Confirmed Cases per 1000' and the social-economics features, as Figure 3 shows. The detailed pair plot can be found in Appendix Figure 14.

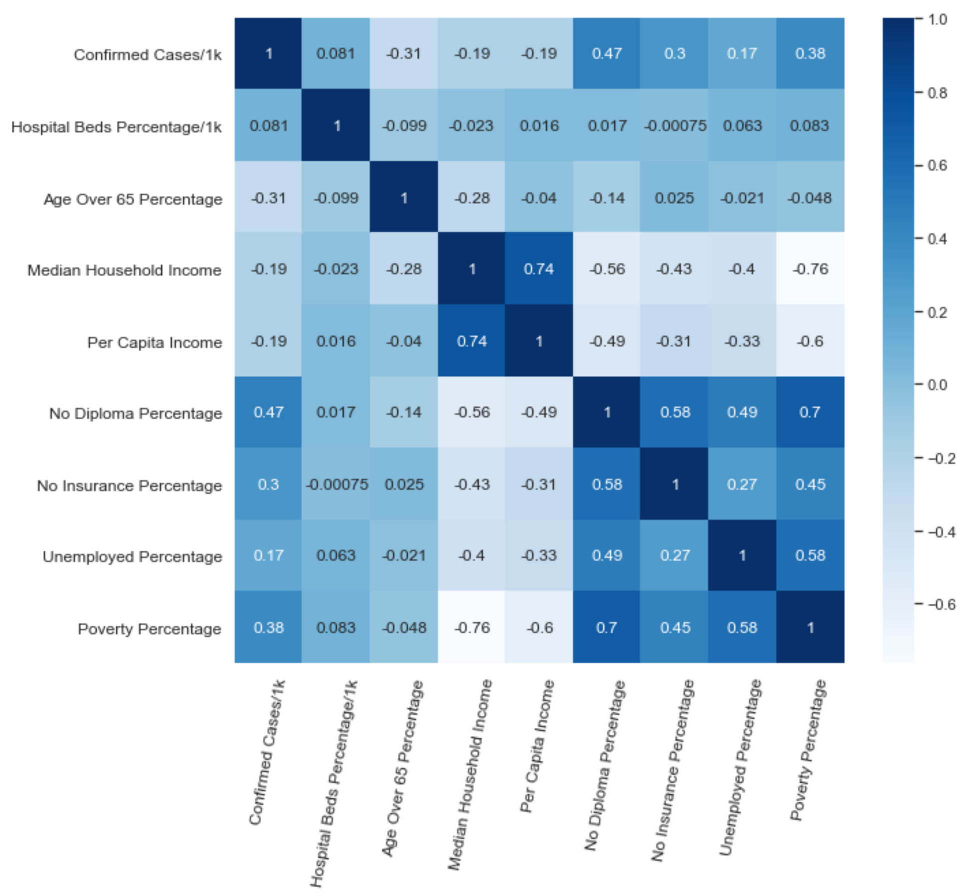

Figure 3: Correlation Coefficients between Confirmed Cases/1k and Social-economics features

From the heatmap, we could infer that:

- There appears to be a positive correlation between 'Confirmed Cases/1k' and 'No Diploma Percentage', 'No Insurance Percentage', 'Unemployed Percentage', 'Poverty Percentage'. That makes sense, because it may be harder for less-educated, illiterate and poverty to realize the severeness of COVID, or proactively take active protection like using masks.

- There appears to be a negative correlation between 'Confirmed Cases/1k' and 'Age Over 65 Percentage'. That's a little unexpected because old people are often more susceptible to infection. However, it is also possible that old people tend to stay at home and take more protective measures.
- Other features do not show strong correlation with confirmed cases, so in the following prediction modeling, we would pick the five features mentioned above.

5 Prediction Modeling Approach and Results

5.1 Model 1: Baseline Model

The baseline model is a Ridge Regression Model[6] focusing on daily Covid confirmed cases prediction on county level.

We use data between May and Oct as training data, and data after Oct is used as our testing data.

Training features includes the sliding window of number of previous 7 days' daily confirmed cases and previous 7 days' daily social distance features ('short_trip_ratio', 'avg_stay_at_home_ratio'). The response y is evaluated day's confirmed cases number. Thus, we have nearly $450k$ training points (i.e $3k$ counties * 150 in a rolling basis of 7 days period).

To get October testing results, we use two approaches:

Approach 1. Every time we have a previous 7 days ($T-7$ to $T-1$) period of true data to predict the target day's (T) response.

Approach 2. First, we use previous 7 days true data to predict the current day's (T) response. Then for the following predictions, we will re-utilize this prediction. For example, for the next day ($T+1$) use previous 6 days' data ($T-6$ to $T-1$), plus yesterday's prediction (T) to predict the target day's ($T+1$) confirmed case, and so on.

After predicting, we got each county's predicted daily confirmed cases in Oct and calculate the mean absolute percentage error (MAPE) of the two methods. The MAPE for approach 1 is 0.022, which is smaller than the approach 2 (0.335), indicating that approach 1 is obviously better than approach 2. However, assuming that we will always get the true previous 7-day data is not practical, so we choose to take the iterative approach for comparison and later refinement.

5.2 Model 2: SEIR

SEIR is a very classic compartmental model for infectious diseases, of which the structure is shown in Figure 4, so in this part, we will use SEIR to predict COVID cases. The SEIR model characterizes the dynamic interplay among the susceptible individuals (S), exposed(E), infectious individuals (I) and recovered/dead individuals (R) in a certain place. β is infection rate or the rate of spread, σ is the incubation rate or the rate of latent individuals becoming infectious ,γ is the recovery rate or mortality rate[13].

The basic hypothesis of the SEIR model is that all the individuals in the model will have the four roles as time goes on[5]. Thus, $N = S + E + I + R$ means the total number of population. The relationship between the number of four roles can be illustrated using the differential equations (1).

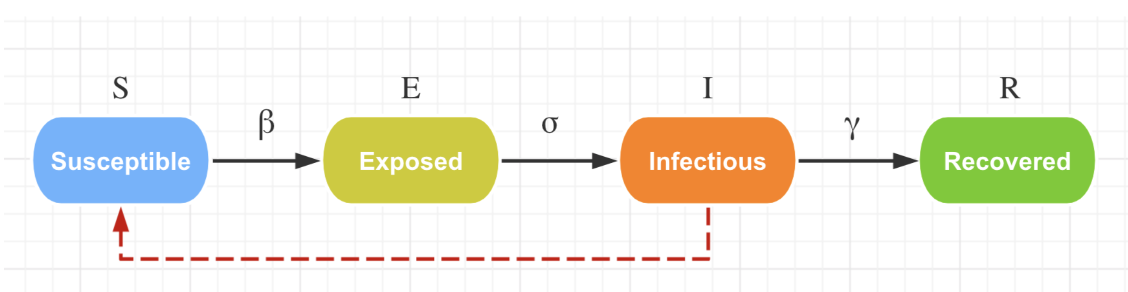

Figure 4: SEIR Structure

$$\begin{aligned}
\frac{dS_t}{dt} &= -\frac{\beta\,(I_t+E_t)\,S_t}{N} \\
\frac{dE_t}{dt} &= \frac{\beta\,(I_t+E_t)\,S_t}{N} - \sigma E_t \\
\frac{dI_t}{dt} &= \mu\sigma E_t - \gamma I_t \\
\frac{dR_t}{dt} &= \gamma I_t
\end{aligned} \tag{1}$$

Since in US, we don't have reliable recovered data, we only use the total confirmed cases data for training. Suppose in time t, the infectious cases number is I_t, and the recovered number is R_t, then the total confirmed cases is $I_t + R_t$, we use C_t to denote it. We fit a SEIR model for each county, which means the parameters $\theta = (\beta, \sigma, \gamma)$ for different counties are different.

In order to find the optimal parameters of the SEIR model, defined by $\theta = (\beta, \sigma, \gamma)$, we use ODE solver in scipy to estimate the parameters which minimizes the following loss function.

$$L(\boldsymbol{\theta}; \mathbf{I}, \mathbf{R}) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{I}_t + \hat{R}_t - C_t\right)^2 \tag{2}$$

When fitting the SEIR model, we find in some county, even the training error can be very large, and will also yield large testing errors. We think this is due to the fact that the transmission rates may change with the epidemiological and status and may be impacted by the policies[12], however in SEIR model the rates are constant.

To solve this problem, we creatively propose a method of ensembling[1] the naive linear model and the SEIR model: Suppose we have fitted a naive linear model in 5.1, then for each county, we first fit the SEIR model, calculate the training error. If the training error is larger than the threshold(we use 1), then use the naive model for test prediction; Otherwise, use the fitted SEIR model.[2] With this method, the MAPE largely decreased from 0.557 to 0.139, and RMSE decreased from 288.201 to 159.546. This may be because it ensembles both the strengths of SEIR of modeling the spread trend and the strengths of Baseline model that consider the characteristics or other influential factors in each county. However, we can still improve it by introducing more complicated models to model both time trend and other factors. And thus we construct RNN.

5.3 Model 3 : RNN

In this part, we hope to better learn the temporal behavior of data and use non-linear functions to predict the dynamics[7]. Common approaches are autoregressive integrated moving average(ARIMA)[1] model, seasonal autoregessive integrated moving average (SARIMA)[8] model, and neural networks based models[2]. Considering the complexity of counties' feature and time series sequence nature of Covid-19 cases in the past year, we picked Recurrent Neural Network(RNN) as a starting point for our Neural Network model, because RNN can use their internal state to process variable length sequences of inputs[10], and can handle non-linear relationships. In our case, we set the timesteps of a sequence to be a 28 days timesteps window.

After some extensive literature review, we utilised a special RNN called Conditional RNN[9] as our primary architecture to predict daily confirmed cases' growth rate. This model is useful when we have time series data with other auxiliary static inputs that do not depend on time. We would like to initialize the RNN states with a learned representation of the condition (e.g. different counties metadata). This way we could model $P(x_{t+1}|x_{0:t}, cond)$

We used a Long Short-Term Memory(LSTM)[3] layer to process the sequence of vectors in our Conditional RNN. LSTM is known for overcoming long-term dependencies problems. With the changing landscape and uncertainty of Covid due to mutation, policies, new treatment, we would like to only use recent information to perform the task, where LSTM is a preferable choice here. A more detailed architecture of our conditional RNN is presented in Figure 5.

In particular, we set $log(increases)$ as our responding variable, and train the model based on static and temporal features. Specifically speaking, the static features are metadata on counties. They are unlikely to change in a 28

[1]Please note that this is may not be a typical of 'ensemble learning', since we will decide on which model to choose for test set prediction based on training errors

[2]Please note that in 5.1, we just fit one model for all the counties, while in 5.2, for different counties we will fit seperate models.

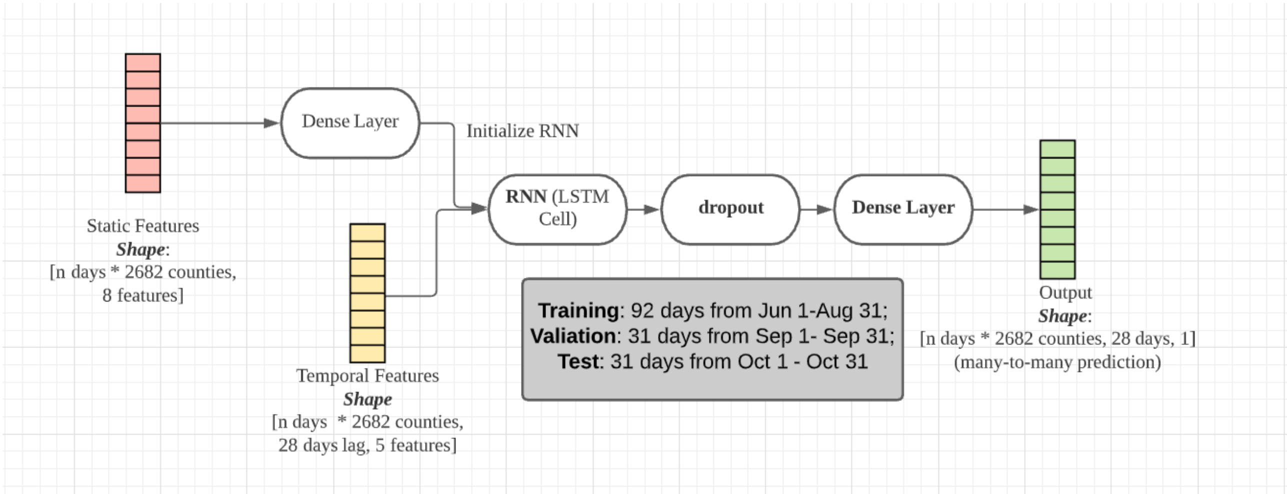

Figure 5: Conditional RNN Structure

days time window, such as counties' demographic, social-economics factor, healthcare, hospital beds predictors are introduced as extra information into models outside of the LSTM by means of additional layers. And other temporal features including previous daily increases of the county, the average increases of its adjunct counties, trips ratios, stay at home rate for the last 28 days are passed through LSTM as sequences to make predictions. The feature details are summarized in Table 2.

Moreover, to train the model, we split the train, validation and test sets by time periods as shown in Figure 5. In the training process, we use mean squared error as the loss function because this the response variable is a set of continuous numbers. As for the optimizer, we choose Adam since it considers both first and second moment. In terms of activation function, we use the default of LSTM, namely the hyperbolic tangent and the sigmoid. To prevent overfitting, we apply a dropout layer before the dense layer, and we also add the early-stopping callbacks.

Table 2: RNN Features Description

Features Categories	Feature Names
Static Features (8 features)	['neighbor_log_new_cases', 'over_65_percent', 'no_diploma_percent', 'no_insurance_percent', 'unemployed_percent', 'poverty_percent','over_65_percent.1', 'beds_per_1000']
Temporal Feature (5 features)	['log_new_cases','trip_ratio','short_trip_ratio', 'avg_stay_at_home_ratio','mean_neighbor_new_cases']

From Figure 7, we could tell that the LSTM model is valid as both its training and validation error decreases as the epoch increases. To further evaluate this model, we calculate the predicted accumulative confirmed cases from the prediction of $log(increases)$. In the heatmap at the right, we display the mean absolute percentage error (MAPE) of its prediction over counties. From this graph, we can tell that the RNN model performs well ($mape < 0.1$) in most counties that we have data. The MAPE heatmaps of other models could be found in the Appendix.

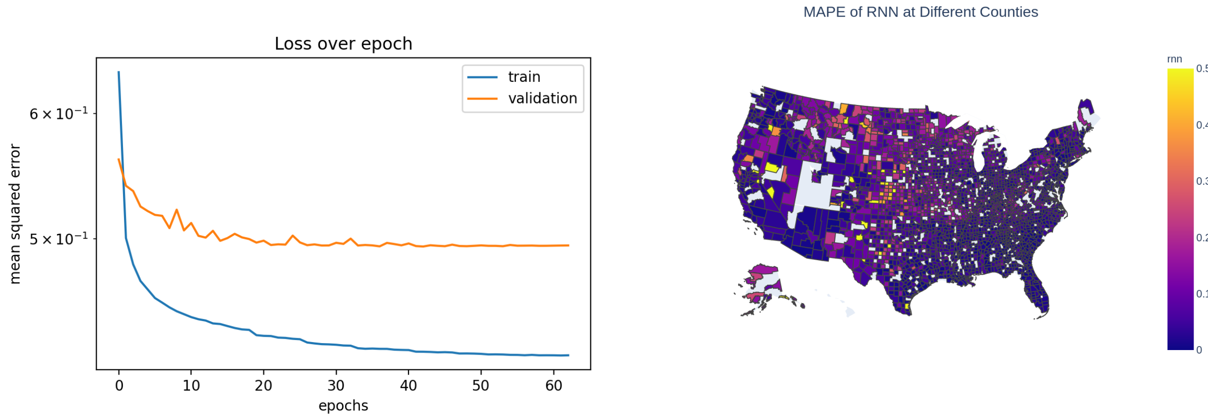

Figure 6: Results of RNN. The left is a diagram of the model's loss over epochs, while the right is a heatmap of average MAPE on the test set of each county.

5.4 Comparison of Models and Future Improvements

From the table 3, We can see the RNN has best prediction results in terms of MAPE (0.077), and we will use RNN as our main prediction model in the later analysis. Although RNN has slightly higher RMSE than baseline, the difference is tiny. This may be due to RNN has much better prediction performance in the short term, but performs a little worse in the long term when the number of confirmed cases is large. To solve this problem, we will in the future combine SEIR with RNN and implement Encoder-decoder to perform more stable prediction in the long term.

Table 3: Comparison of Results of Different Models

Metrics	Baseline	SEIR	Ens-SEIR	RNN
MAPE	0.335	0.557	0.139	0.077
RMSE	126.856	288.201	159.546	128.674

Besides, SEIR tend to have worst performance because it is mainly powerful in the short term but we conduct prediction in a month. But Ens-SEIR combine the strengths of both Baseline and SEIR and have a much better performance.

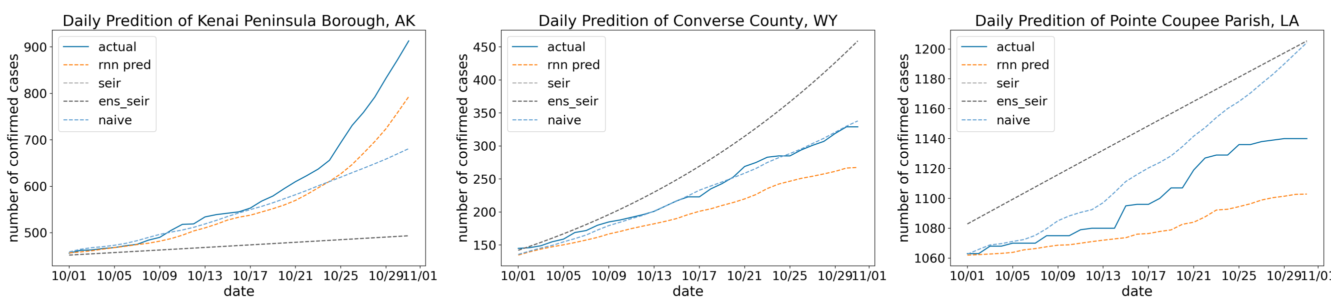

Figure 7: Comparison of Predicted Accumulative Confirmed Cases of Different Models[3]

Actually, different models still have their pros and cons and there are also future improvements we will do based on them, as is summarized in Table 4

6 Policy Analysis

In order to analyze the effects of different policies on COVID spread, we use causal inference to investigate the causal influence of policy on COVID growth.[4] Because the policy is state-level but the prediction is county-level, we

[3] Please note that the ensemble SEIR method choose naive/original SEIR model for test prediction based on training error, so the line of ens_seir on Figure 7 would always overlap with either naive or seir.

Table 4: Comparison of Models and Future Improvements

Models	Pros	Cons	Future Improvements
Naive	Linear, Interpretable, Easy to add more features to analyze;	Model is too easy, hard to capture complexity in data	Add non-linearity, like polynomial etc.
SEIR	Easy to data process; Match with epidemiology classic model's intuition; Different models for different counties	Hard to add features and interpret influence of policy etc.; Constant transmission rates over time	Combine with RNN, Or add more complexity into SEIR based on epidemiology
Conditional RNN	Conditional RNN captures static information to LSTM; Computational power better comparing to pure LSTM	Model still needs longer time to train; Hard to interpret	Possibly use SEIR to initialize hidden layer; Use Encoder-decoder conditional RNN method to expand to a wider range of time lag prediction

independently construct causal inference model instead of embedding the policy analysis into predictive model. At this time, we will focus on the direct and indirect effects of policies(through the effects on people's behavior). In this way, we can combine the policies and people's behavior(mobility data) together.

6.1 Causal Inference: Modeling

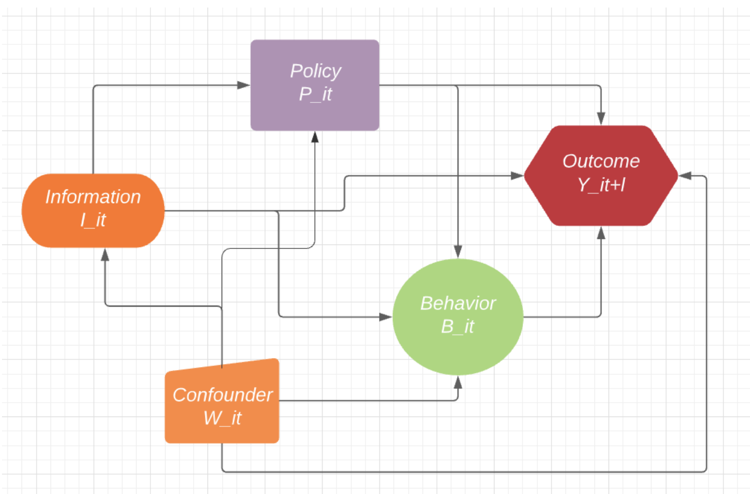

Figure 8: Causal Structure of Policy Analysis

The Figure 8 depicts how the policies, behavior and information interact together. And the causal ordering is determined in this sequence:

1. Information I_{it} and confounding variable W_{it} is determined at time t
2. policies P_{it} are set at time t at the certain state, given the information and confounder
3. behavior B_{it} is observed, given policies, information and confounders at time t
4. The outcome $Y_{i,t+l}$ of the policy is realized at time $t + l$ given the policies, information, confounders, behavior

At this time, P_{it} is the policy variable vector, including mainly three types of policies: Stay at Home Order, Business Limits(Restaurant Limits), Mandatory Quarantine for Travelers. These are policies limiting people's behavior from different levels. And Behavior variable vector B_{it} is Trip ratio and Stay at Home ratio, which are highly related to the policies we focus on. Outcome variable $Y_{i,t+l}$ is defined as the growth rate $l(l = 14)$ days later. Information variable vector I_{it} is the extra information, such as the new cases at time t, that can affect behavior and include the other effects of other policies we do not focus on. Confounding variable vector W_{it} is the confounding variables that may affect all the other variables and related to the characteristics of the states, such as population.[11]

Table 5: Part of the Results of Causal Inference

Policy	Direct Effect	Trip Ratio Effect	SAH Effect	Total Effect
Restaurant Limits-Closed Except for Takeout/Delivery	-0.007	0.214	0.011	-0.009
Stay at Home Order-Rolled Back to High Risk Groups	-0.002	0.347	0.011	-0.006
Mandatory Quarantine for Travelers-From Certain States (New)	-0.004	0.110	0.052	-0.005
Restaurant Limits-New Service Limits	-0.004	-0.050	0.014	-0.004
Stay at Home Order-New Stay at Home Order	-0.005	-0.288	0.001	-0.002
Mandatory Quarantine for Travelers-All Travelers	0.002	-0.194	0.018	0.004

Our quantitative model for the causal structure is given by the following potential outcomes model, where ϵ is the other information not included in the variables. :

$$Y_{i,t+l} = \alpha B_{it} + \pi P_{it} + \mu I_{it} + \delta_Y W_{it} + \epsilon_{it}^y \tag{3}$$

$$B_{it} = \beta P_{it} + \gamma I_{it} + \delta_B W_{it} + \epsilon_{it}^b \tag{4}$$

But for each behavior variable, equation(4) will be constructed accordingly as follows:

$$B_{it}^j = \beta^j P_{it} + \gamma^j I_{it} + \delta_B^j W_{it} + \epsilon_{it}^{bj} \tag{5}$$

With equation(3) and equation(4), it can be observed that the influence of policy on behavior is β and policy's direct effect on outcome is π. Based on these, we have the equation that depicts the total effect of policy on the outcome.

$$Y_{i,t+l} = (\alpha\beta + \pi)P_{it} + (\alpha\gamma + \mu)I_{it} + \delta W_{it} + \epsilon_{it} \tag{6}$$

where $\alpha\beta$ represents the indirect effect of policy on outcome, π represents the direct effect of policy, and $\alpha\beta + \pi$ is the total effect of the policy.

6.2 Results and Policy Conclusions

Based on the data from June 1st to September 30th, we also conduct t-test on the coefficients to determine the important policies. Here, we define the "important" as the policy has negative effect on the confirmed cases growth rate and either its effect on growth rate, or on one of the behavior variable is significant. And part of the results is shown in Table 5.

Note: All the coefficients here are significant under the level of 0.05. The complete table could be seen in appendix.

In the three types of policies, *Restaurant Limits-Closed Except for Takeout/Delivery, Stay at Home Order-Rolled Back to High Risk Groups, Mandatory Quarantine for Travelers-From Certain States (New)* are the most effective. We can also get the following conclusions and recommendations regarding the policies.

1. **Business-Level policies are more effective.**

   Compared with the Mandatory Quarantine order which is aimed at country-level and the Stay at Home order which is aimed at individual-level, the Restaurant Limit which is aimed at business-level tend to be more effective in controlling the growth of cases. All of the Restaurant Limits are important, which can be seen from the complete table. This may be due to the controllable nature of business. And it is recommended here that the business-level policies are implemented to control the spread of COVID.

2. **Government should be cautious at individual-level policies, like Stay at Home Order.**

   Compared with the strict Stay at Home Order, limit on the high-risk group is the more effective. This may be because stricter individual-level policies will offend people's rights of freedom and make people unwilling to obey the rules, even causing the opposite effect. As we can see, the effect of Stay at Home order(New) is mainly on limiting the trips, but has little effect on the stay at home ratio. it is recommended here that the individual-level policies are implemented carefully and focus on high-risk group.

3. **Country-level policies, like Mandatory Quarantine, should be more specific.**

   Similar with Stay at Home Order, the stricter orders of Mandatory Quarantine didn't lead to necessarily more effective result compared with the specific one, like limiting people from certain states. This is also because such policies will bring side-effect, like making people who feel offended about banning traveling to have a higher tendency of making short trips. Such behavior will undermine the total effect brought by the policy. Therefore, it is recommended here that the individual-level policies are implemented carefully and specifically.

4. **Business-level policies, like Restaurant Limits, should be more strict but not stop its main function.**

   Restaurant Limits-Closed Except for Takeout/Delivery is the most effective policy compared with partial limit ones. Although it is the most strict one, it does not prevent restaurants from functioning and supporting people's normal life. Since the business-level policies are most controllable, it is recommended here that the business-level policies are designed judiciously to be both strict and reasonable.

6.3 Policy Effect on Predictions

Based on the prediction results of RNN model and the causal inference of policy analysis, we predict the confirmed cases on test data (October 1st to October 31th) and also show how the prediction would change with counterfactual policies. Here counterfactual means if the current policy of the state is on, then we will show the real prediction with solid line and show the prediction of confirmed cases if policy is off with dotted line, as figure 9 and figure 10 show.

Figure 9 shows that in McCone, Luna and Morrow, their policies of Restaurant Limits-Closed Except for Takeout/Delivery, Stay at Home Order-Rolled Back to High Risk Groups, Mandatory Quarantine for Travelers-From Certain States (New) are all off. But once the policies are on, the counterfactual prediction of future confirmed cases will be less than the factual one. And the restaurant limits tend to have the largest effect.

Similarly, Figure 10 shows that in Chemung, Richmond and Warren, their policies of Restaurant Limits-Closed Except for Takeout/Delivery is off, but Stay at Home Order-Rolled Back to High Risk Groups, Mandatory Quarantine for Travelers-From Certain States (New) are both on. But once the later two policies are off, the counterfactual prediction of future confirmed cases will be greater than the factual one.

In all, Figure 9 and Figure 10 show that policies have strong effect on the spread of COVID. Besides, it has verified the effectiveness of our work that we can combine the prediction and policy analysis together to evaluate the influence of policies on the "future".

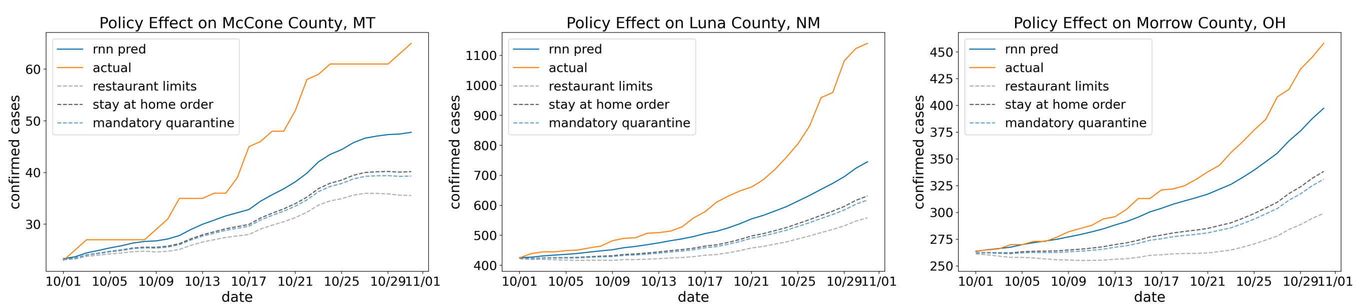

Figure 9: Figures of three counties where the real policies are off and dotted lines represent the counterfactual prediction if the policies are on, indicating the confirmed cases will be less than factual

7 Conclusion

Our work focused on the county-level prediction for COVID-19 in US and the state-level policy analysis. And we successfully combined them to demonstrate how the number confirmed cases will change and what the different levels of policies will affect it.

For the county-level prediction, we constructed and compared three models: Ordinary Linear Regression, SEIR and Conditional RNN. Different models have their own pros and cons but we chose Conditional RNN to be our main forecasting model as it has the best prediction performance. And we combined time-series data, county profile, behavioral data and geographical data to build it, successfully taking the characteristics of each county into consideration.

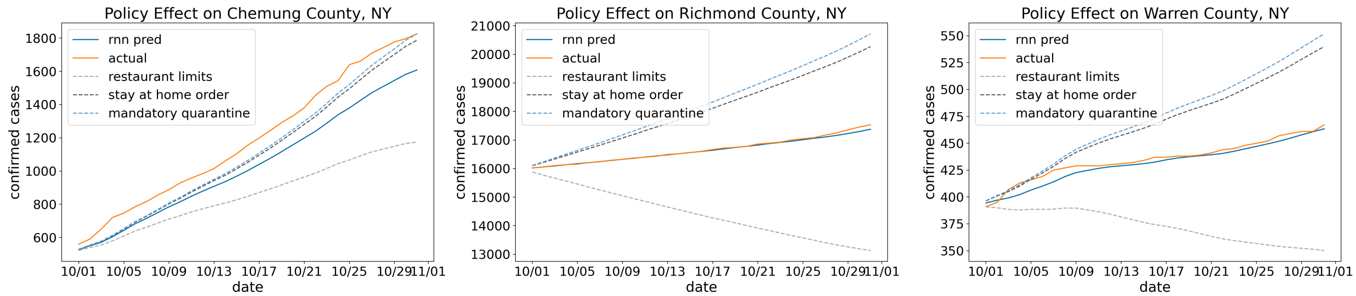

Figure 10: Figures of three counties where the restaurant limits is off, stay at home order and mandatory quarantine are on and dotted lines represent the counterfactual prediction if the policies are off/on, indicating the confirmed cases will be greater/less than factual

For the state-level policy analysis, we constructed causal inference model to analyze both the indirect influence and direct influence of policy. And we provided solid analysis and recommendations based on the results.

Finally, we integrated the two parts together to evaluate the influence of policy implementation on future cases, which will be especially meaningful for government and the whole society.

8 Discussion and Future Work

8.1 Strengths

Instead of analyzing the policy effect through observing the coefficients of policy features, we separate the prediction model and policy analysis model. This approach has two main strengths: 1) The prediction model could be focusing on accuracy instead of sacrificing accuracy for explainability. And for the policy analysis, we could use causal inference to deeply investigate both the magnitude of the policy effects and the relationship between policy and other features. In this way, we could maximize both the accuracy and the explainability. 2) The prediction model could be focusing on county-level and the policy analysis could be focusing on state-level. Actually, counties have their own unique features and states also have their confounding variables. If we combine them, it is hard to know the effects of the features separately.

8.2 Limitations and Future work

1. **Analyze the influence of other factors on COVID spread.** Although we successfully constructed the causal inference model to analyze the policies, our work did not detailedly analyze the other important characteristics of the counties, such as the geographical and social-economics features. Limited by this, our work failed to provide a complete analysis on all the important features in order to deeply understand the underlying reasons of different spread pattern in different counties. In the future, we would implement feature importance and relationship analysis, especially for RNN model.

2. **Analyze the uncertainty of our prediction model with confidence interval.** Now because of the computational limitations, we can not perform bootstrap for RNN to provide the confidence interval of the prediction, which is vital for the policy designers to understand how certain our analysis is. In the future, we will provide both the undertainty analysis for the prediction as well as the causal inference to set concrete confidence interval of policy influence in the future.

3. **Analyze the trade-off between economic development and epidemic control.** An essential issue of policy analysis is to investigate how much it will hurt the economic development and people's daily life and understand the trade-off between economics and prevention of COVID spread. But our work temporarily did not consider it. In the future, we will provide a thorough analysis on it to offer more practical suggestions with Machine Learning approaches.

9 Impact Statement

During this special time, our work can provide a valid scientific support for both the government to design effective policies and the related department to prepare for the future spread of COVID based on the prediction.

Above all, since our prediction of future confirmed cases is based on the characteristics of each county, the government and related department, like hospital and school, etc, can targetedly prepare for the epidemic spread and assist them to make plans and strategies. Besides, our causal inference model of policy analysis provide valid explanation for the effectiveness of the policies and in what way they are affecting the growth of the cases. In this way, government can have a reasonable and scientific reference on how to design the most powerful policies. Last but not least, our work will also have a broad positive impact on the whole society. Some people may not be willing to follow the instructions from the government and feel their freedom is restricted. Our work offer the scientific explanation of how serious the future would be with people's behavior and how effective the change of people's actions will be. In this way, we can make more people actively involved and support the strategies.

However, there are also several ethical concerns and risks that need to be taken seriously. These include:

1) **The risk of overusing the prediction models.** What need to be remembered is that model is just a reference tool, instead of a bible for us totally depend on. When using our model to make predictions for the future, it should be noticed that there may be some extra or uncertain features that we do not consider. Therefore, sometimes there may be important biases that need to be cared about. To mitigate these risks, we recommend paying attention to what is the difference between the real scenario and the scenario we consider in the model.

2) **The risk of highly relying on the policy causal influence.** Our analysis of policies also should be used carefully. Real-world cases are really complicated and there are other factors that may affect the influence of the policies. To mitigate these risks, we recommend using our analysis to understand what the policies can do but also take care about what they can not do.

3) **The emotional issues of over-generalizing the conclusions to individuals.** Our analysis is based on group-level, which is focusing on the effect of the actions of group of people. It is unfair and unethical to use it to evaluate individual's actions. To mitigate these risks, we recommend paying attention to the similarity between groups as well as the difference within groups.

References

[1] Domenico Benvenuto, Marta Giovanetti, Lazzaro Vassallo, Silvia Angeletti, and Massimo Ciccozzi. Application of the arima model on the covid-2019 epidemic dataset. *Data in brief*, page 105340, 2020.

[2] Vinay Kumar Reddy Chimmula and Lei Zhang. Time series forecasting of covid-19 transmission in canada using lstm networks. *Chaos, Solitons & Fractals*, page 109864, 2020.

[3] Tensorflow Documentation. Recurrent neural networks (rnn) with keras. Tensorflow Guide. URL`https://www.tensorflow.org/guide/keras/rnn#introduction`.

[4] Andrew Goodman-Bacon and Jan Marcus. Using difference-in-differences to identify causal effects of covid-19 policies. *DIW Berlin Discussion Paper*, May 2020.

[5] Shaobo He, Yuexi Peng, and Kehui Sun. Seir modeling of the covid-19 and its dynamics. *Nonlinear Dynamics*, 101(3):1667–1680, 2020.

[6] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[7] Gwenan M Knight, Nila J Dharan, Gregory J Fox, Natalie Stennis, Alice Zwerling, Renuka Khurana, and David W Dowdy. Bridging the gap between evidence and policy for infectious diseases: How models can aid public health decision-making. *International journal of infectious diseases*, 42:17–23, 2016.

[8] Zohair Malki, El-Sayed Atlam, Ashraf Ewis, Guesh Dagnew, Ahmad Reda Alzighaibi, Ghada ELmarhomy, Mostafa A Elhosseini, Aboul Ella Hassanien, and Ibrahim Gad. Arima models for predicting the end of covid-19 pandemic and the risk of second rebound. *Neural Computing and Applications*, pages 1–20, 2020.

[9] @philipperemy. Conditional rnn. Github, Aug 2020. URL`https://github.com/philipperemy/cond_rnn`.

[10] Ahmed Tealab. Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal*, 3(2):334–340, 2018.

[11] H. Kasahara V. Chernozhukov and P. Schrimpfr. Causal impact of masks, policies, behavior on early covid-19 pandemic in the u.s. *Jounal of Econometrics*, September 2020.

[12] Jin Wang. Mathematical models for covid-19: applications, limitations, and potentials. *Journal of Public Health and Emergency*, 4(0), 2020.

[13] Difan Zou, Lingxiao Wang, Pan Xu, Jinghui Chen, Weitong Zhang, and Quanquan Gu. Epidemic model guided machine learning for covid-19 forecasts in the united states. *medRxiv*, 2020.

Appendix

Supplementary EDA

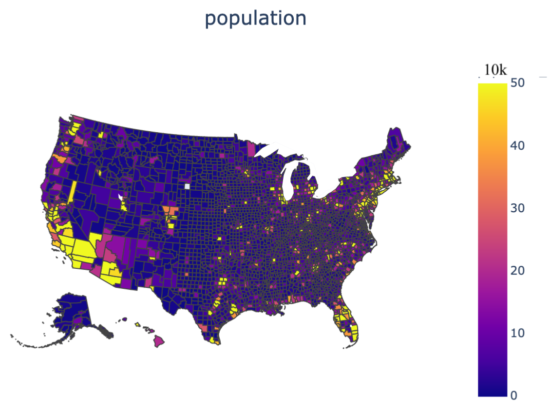

Figure 11: Heatmap of the population in different counties

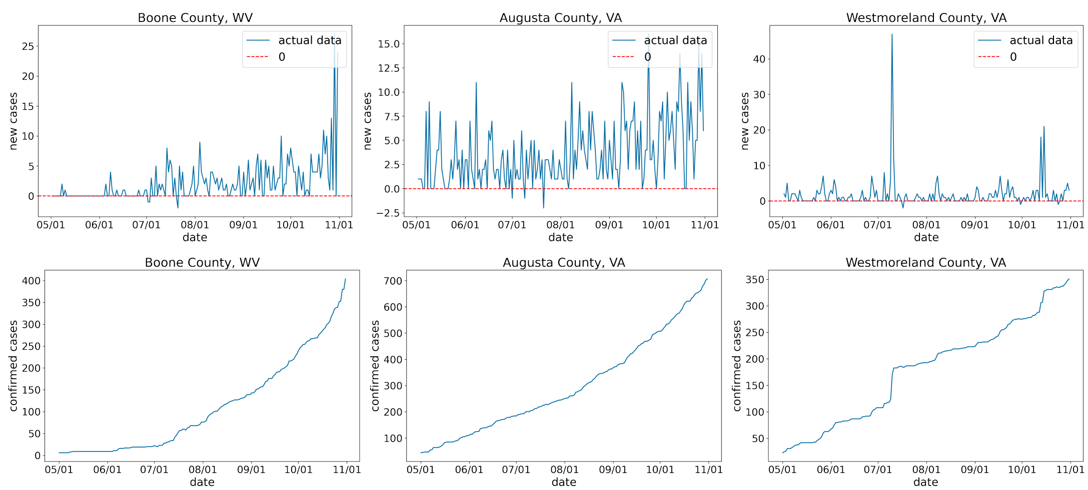

Figure 12: Figure of Confirmed Cases of Counties with Negative New Cases

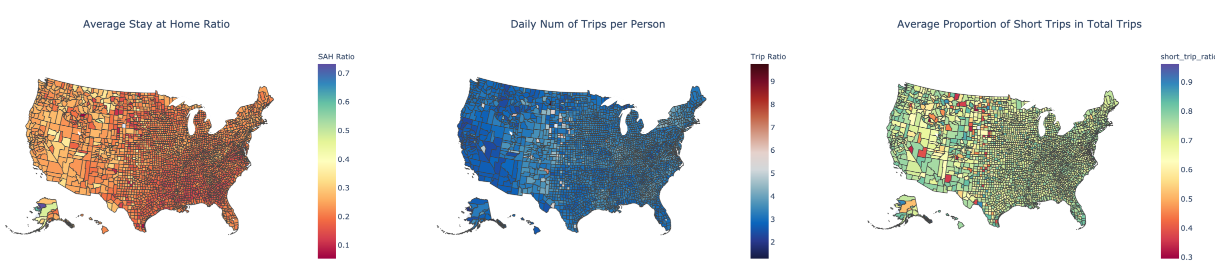

Figure 13: Figure of Counties' Mobility Pattern. From the left to the right are the daily average SAH ratio, trip ratio and short-trip ratio respectively. Here, trip ratio means the average trips taken by residents in a county, while the short trip ratio means the average proportion of trips shorter than 10 miles in total trips.

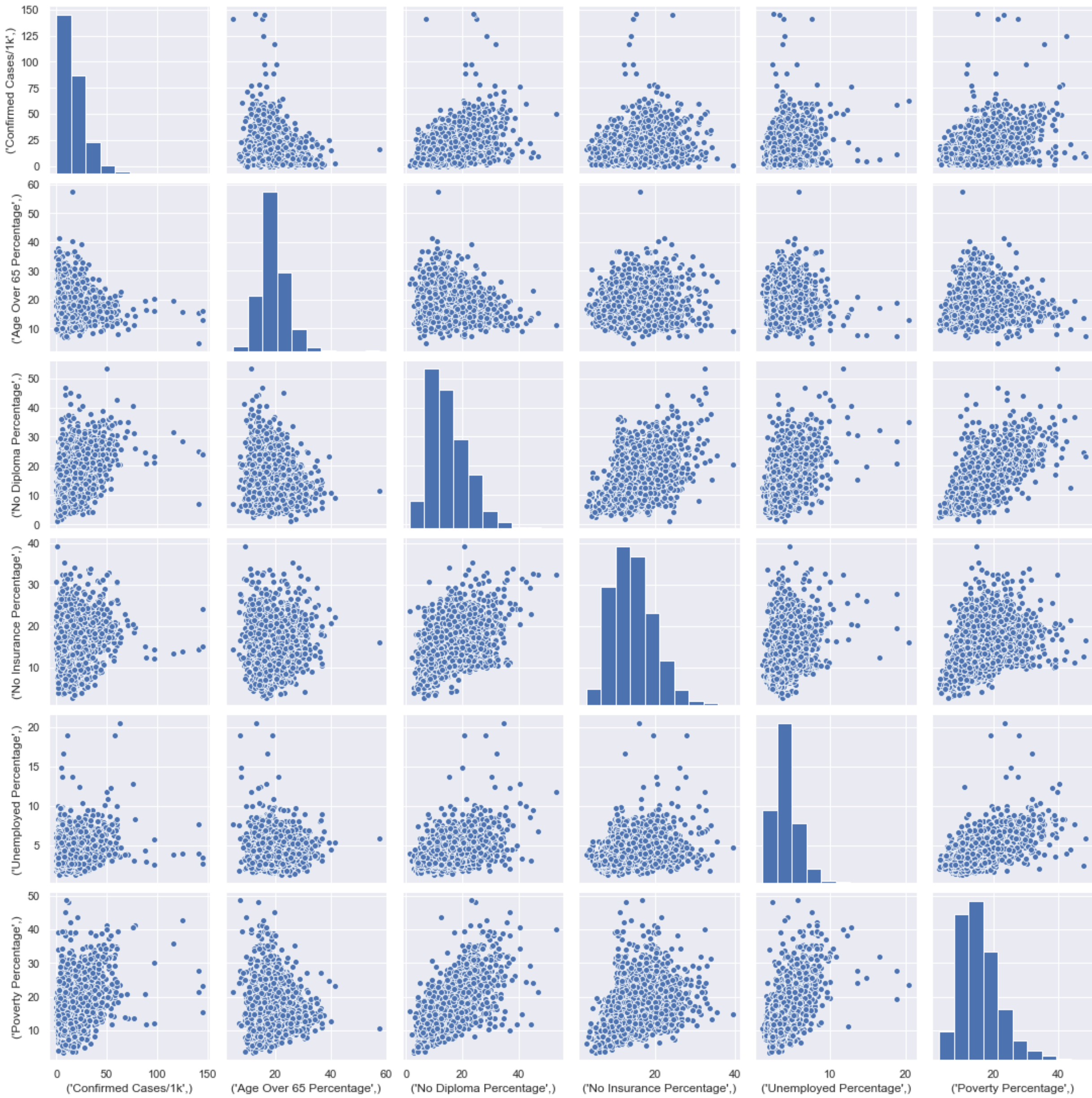

Figure 14: Pair Plot for Confirmed Cases/1k and Social-economics features

Supplementary Results

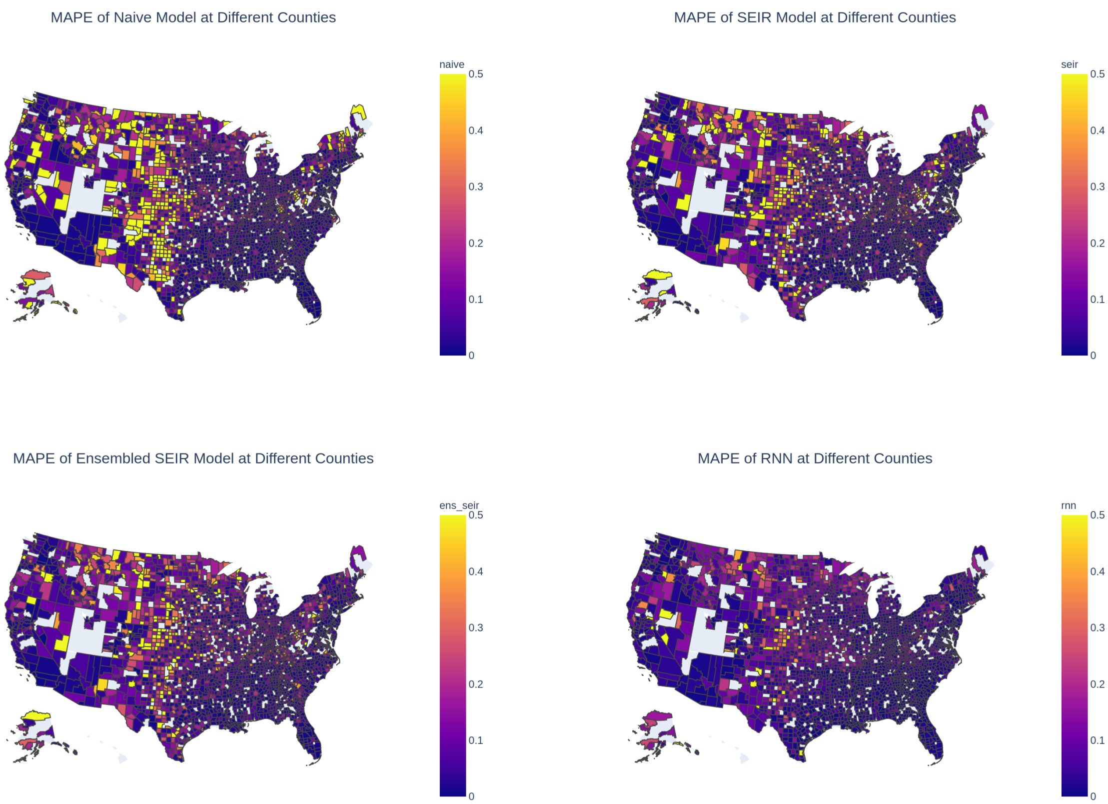

Figure 15: County-wise MAPE for naive model, SEIR, ensembled SEIR and RNN

Policy	Direct Effect	Trip Ratio Effect	SAH Effect	Total Effect
Restaurant Limits-Closed Except for Takeout/Delivery	-0.007	0.214	0.011	-0.009
Restaurant Limits-New Capacity Limits	-0.009	-0.250	0.036	-0.007
Stay at Home Order-Rolled Back to High Risk Groups	-0.002	0.347	0.011	-0.006
Mandatory Quarantine for Travelers-From Certain States (New)	-0.004	0.110	0.052	-0.005
Restaurant Limits-New Service Limits	-0.004	-0.050	0.014	-0.004
Mandatory Quarantine for Travelers-Rolled Back to International Travel	-0.003	-0.008	-0.003	-0.003
Mandatory Quarantine for Travelers-Rolled Back to Certain States	-0.001*	0.148	0.000*	-0.002
Stay at Home Order-New Stay at Home Order	-0.005	-0.288	0.001	-0.002
Stay at Home Order-Statewide	-0.003	-0.192	0.029	-0.001
Mandatory Quarantine for Travelers-All Travelers (New)	0.000*	0.000	0.000*	0.000
Mandatory Quarantine for Travelers-All Air Travelers	0.006	0.308	-0.018	0.003
Mandatory Quarantine for Travelers-All Travelers	0.002	-0.194	0.018	0.004
Mandatory Quarantine for Travelers-Lifted	0.004	-0.071	-0.001*	0.005

Note: * represents the coefficient is not significant under the level of 0.05.