# The Impact of Measures on Covid-19 Transmission in Europe

Haoyuan Li, Jingyi Tang, William Xu

February 2021

## 1 Introduction

The Covid-19 virus is no doubt the most concerning health issue across the world. Up to 2021.2.17, there are 110 million people confirmed with Covid-19 infection across the world. It is important for policy-makers and public health departments to develop good measures to control the transmission of virus and recover the economy.

The fast response in China, Japan and South Korean shows that with proper public health governing, the spread of the virus can largely be suppressed. In the meantime, the public health response in the Europe has been less uniform and shows more variations. This creates a chance to study the effect of various policies in different conditions. In this report, we will study the important questions in this area with the daily COVID-19 and corresponding measure data in Europe together with additional climate information.

Here we pose the following quesitons based on the analysis of the datasets:

- **What are the most significant factors that determine the effectiveness of measures?**
- **Which kind of measures are the most efficient for different countries? Can we predict the trend of epidemic given certain set of measures?**

In Section 2 we show the procedures to pre-process and validate the data. In Section 3 we performed feature engineering and causal inference to investigate temporal structure of measures and new cases. In Section 4 we built a structured time series model to analyze the effects of different measures in different countries and make predictions.

Based on our findings, we are very positive about predicting the trend of the epidemics with temperature and measure data. Our analysis shows that both the type of the measures taken and the temporal structure of how measures are performed can have significant impact on the control of the transmission. For example, restrictions on workplace have the most significant impact to reduce the new cases in Norway and France. The impact of effective public health measures can dominate the impact of temperature change. We also have an interesting conclusion: despite of the diverse social, economical and cultural conditions of European countries, the features of overall good measures are similar: Starting the measures at an early stage and continuing the restrictions until the epidemic slows down.

# 2 Data Pre-processing and Visualization

## 2.1 Data Imputing, Normalizing and smoothing

For the Worldwide Covid-related statistics dataset(owid), we perform multiple imputation to fill the columns for which less than 3% of the total data is missed. Specifically, the mice package in R creates 10 copies of the dataset. In each copy, it imputes the missing data with values using the classification and regression tree (CART) method.

For time series data, the most recent data for some countries is missing. We fill those missing data with the nearest value. Some covariates like vaccine information are mostly missing. Since the date of the dataset stopped at Jan.2021, before the vaccination started in most countries, in this report we neglect the effects of vaccine and focus on how measures can change the transmission of the virus.

Based on our investigations, we found that the scale of covariates can vary a lot. We apply a Standard scalar to normalize them with mean zero and standard deviation 1 for distanced-based classifications.

Some of the data(stringency index and categorized measures) are non-continous. We apply a third-order Savitzky–Golay filter with a 7-day time window on them for time series analysis.

## 2.2 Test number

To obtain trustworthy conclusions, one needs reliable data. Among all factors, the most important one is the test number. One can only determine the actual infection rate with a large enough test number.

On the other hand, however, it is not straightforward to establish the criterion of enough sampling, because the simple statistical error is not the bottle neck of the accuracy of the result. The real issue is that there is no guarantee that the tested people faithfully represents the general infection rate.

Therefore, with the lacking of massive, accurate and timely virus screening of general population in most countries, one can only assume the validity of the data by self-consistence check.

According to an early press conference from WHO[2], 10% can be used as a standard to determine where there enough test or not for the data. However, as the virus continues infesting various countries, apparently, 10% is no longer

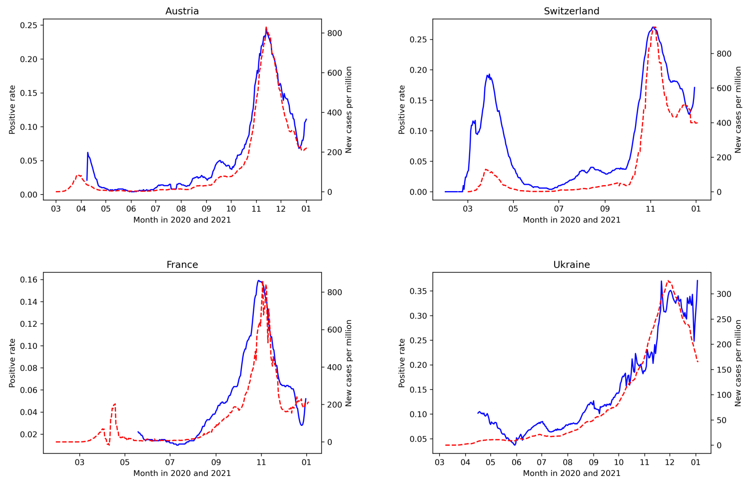


Figure 1: Daily positive rate data and the new cases data. The two curves show great similarity after 2020.6.2. Therefore, we assume that the testing number is sufficient after 2020.6.1.

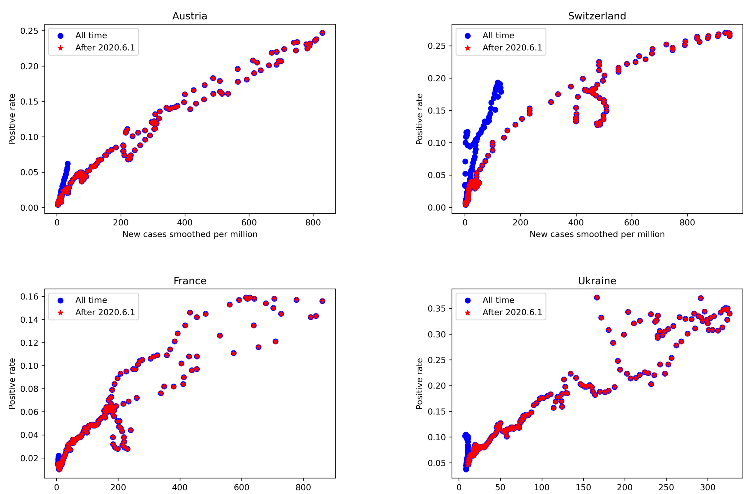


Figure 2: Positive rate and new case number per million people for different time range.

a valid standard for data validity. Therefore, we compare the *positive_rate* and the *new_cases_per_million* in the dataset.

The two curves are shown in figure.1 for several west European countries. As can be seen readily from this figure, after 2020.6.1, the two curves tightly follow each other. Therefore, we assume that, after 2020.6.1, the test number is large enough, and therefore the *new_cases_per_million* represents the actual infection rate well. (Notice that here, we are not claiming that the *new_cases_per_million* is a faithful value of the actual infection rate. Rather, here we just claim that, after 2020.6.1, the constitute of the tested people are stable, and that the *new_cases_per_million* is positively proportional to the actual infection rate of general population.) Similar conclusion can be drawn from figure.2. In this figure, the scatter plots of positive rate versus new case number per million people are shown for all the time and after 2020.6.1. Obvious branching are observed for these two time periods. This confirms the previous observation.

### 2.3 Stringency index and Categorization of Measures

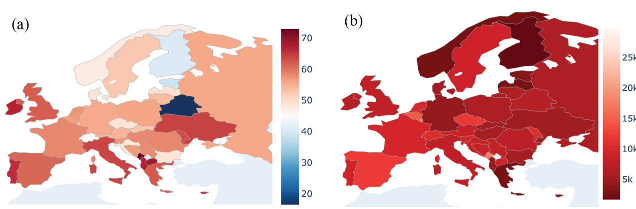


Figure 3: (a) Color map of the time averaged stringency index of European countries. (b) Color map the total cases number per million people of European countries.

The stringency index is an overall estimation of the strictness of the measure against the virus. The concrete definition of the stringency index can be find in ref.[1].

Beyond the overall index, we hope to analyze the effects all different kinds of measures. From European Centre for Disease Prevention and Control(ecdc) dataset we characterized all policies and manually by assigning a value to each of them. Details of this categorization can be find in the Appendix.

## 3 Correlation and Causality Analysis

In this section we use stringency index as an overall indicator on how strictly all kinds of measures are implemented and analyze its temporal structure. In the next section we will discuss the effects of different measures in detail.

### 3.1 Temporal Features of Measures

#### 3.1.1 Control the variables

European countries have different social, economical, cultural and natural conditions. Each of the conditions has the potential to influence the development and control of the convirus epidemic. Here we explore the different national conditions and investigate its correlation with total cases and deaths per million. We first conduct measure analyses on several "similar" country and then generalize the analysis to all European countries.

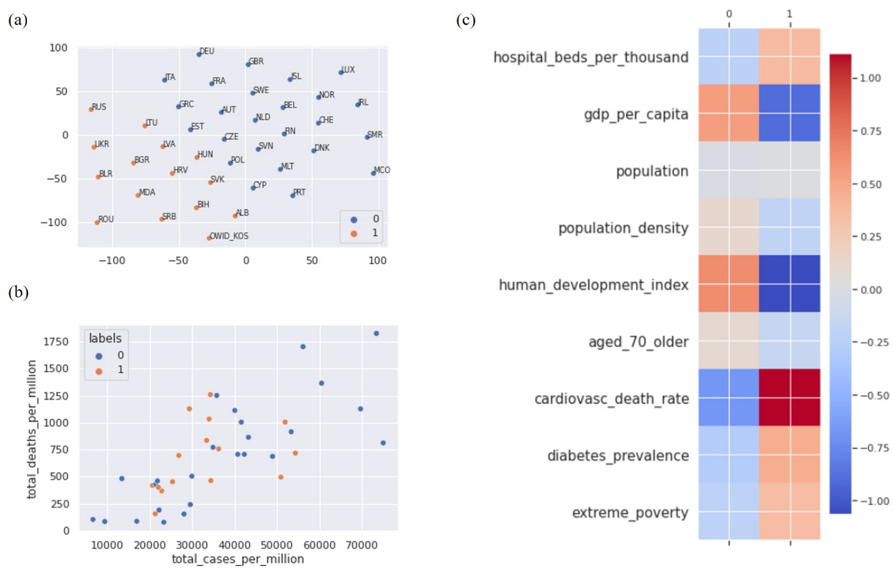


Figure 4: (a)TSNE embedding of the cluster of European countries classified by k-means classifier. (b)The distribution of total deaths per million and total cases per million of the two clusters. (c)Patterns of social and economical index for the two clusters.

Given the information in the owid dataset, we use 9 social and economical indexes(as shown in Fig. 4(c)) as features. We applied a k-mean clustering algorithm to group the countries into two clusters, labelled 0 and 1. The TSNE embeddings is shown in Fig. 4(a), verifying that the clusters are well formed. As indicated by Fig. 4(c)), cluster 0 is comparatively more developed countries with higher gdp per capita, human development countries, etc. We show distribution of total deaths per million and total cases per million of these two clusters in Fig. 4(b)). We find that these two groups don't show significant differences. Surprisingly some more developed countries even have larger deaths and new cases numbers compared with the average level.

To anaylze the effects of the temporal structures of measures, we would like to find sample countries with similar social, cultural and natural conditions to control the external variables. Also, there are multiple waves of coronavirus infection. For simplification we focus on the last wave, which generally start from Sep. 2020 and last till now. Before Sep. 2020, the daily new cases in almost all countries in Europe went to zero and therefore we can approximately regard this new wave as independent. We use the accumulated new cases per million population from Sep. 1st 2020 to Jan. 4th 2021 as an indicator of how

COVID19 is controlled in each country.

#### 3.1.2 Feature Engineering

We start with observing four similar countries in the Northern Europe: Norway, Finland, Sweden and Denmark. The smoothed new cases per million and stringency index is shown in Fig. 5. The stringency index shows distinct temporal patterns and severity of the epidemic also differs. Denmark and Sweden, as shown in Fig. 5(a) and (b), have much more daily new cases compared with Finland and Norway(Fig. 5(c) and (d)).

Through detailed analysis of the evolution of stringency index, we think it is probable that two kinds of structures may have significant impacts. First is the timeliness of the measures. For Sweden, when the measures were taken to deal with the new wave, as shown by the black arrow in Fig. 5(b), the daily new cases per million had increased to over 400 hundred. The delay of government response makes these measures less efficient in suppressing the new cases. As a contrast Finland and Norway took prompt actions when new cases per million was below 50. The second feature is the consistency of the measures. For Denmark, although it increased stringency index early, the restrictions were relaxed before daily new cases reached its maximum (shown by the arrow in Fig. 5(a)).

Inspired by these observations, we designed two features from the temporal structure of the measures. **Measure Start**: The daily new cases per million at the date when the stringency index begin to rise. **Measure Decline**: the aggregated decline amplitude from the measure start date to the date when new cases reach its maximum. The definitions are shown by the arrows in Fig. 5(a)(b).

With the definitions above we define measure flag as follows:

$$measure\_flag = \begin{cases} "good" & if\ measure\_start < 200\ and\ measure\_decline > -3 \\ "bad" & otherwise \end{cases} \tag{1}$$

It means we consider measures a country takes as "good" if it is both timely and consistent. We therefore divide European countries into two classes: the countries with good measures and countries with bad measures. Fig. 6 shows the distribution of total cases per million for each of the two classes. The countries with good measures show significant better control of the epidemic, despite of their diverse social, cultural and economical conditions. These result implies that first, how measures are conducted can be dominant factor for the epidemic control and second, timeliness and consistency can be universal requirements for good measure implementation.

### 3.2 Causal Inference

In this section we will perform more detailed analysis on causal relation between measures and severity of epidemic by Granger Causality. It is well known that

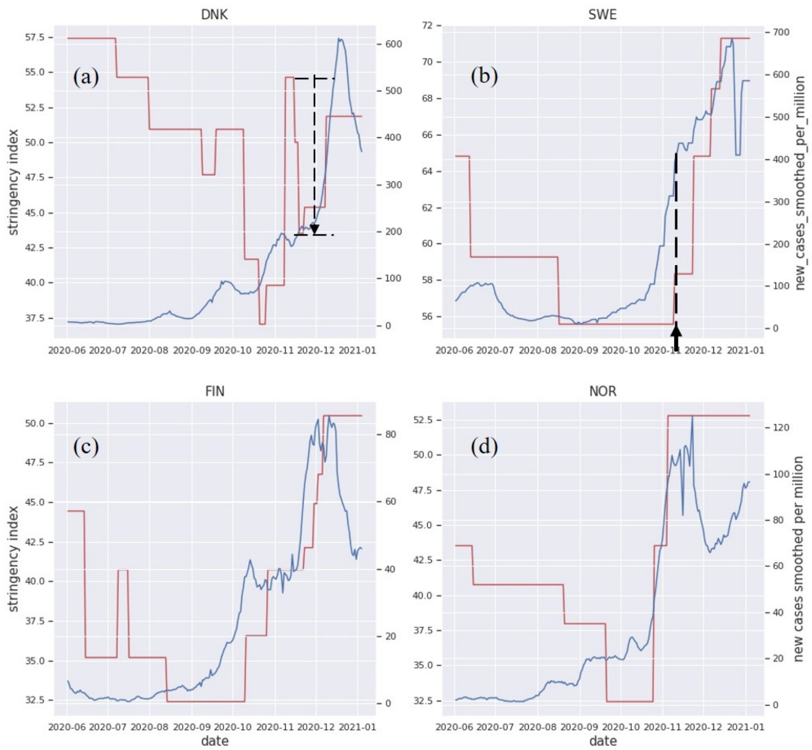


Figure 5: Smoothed new cases per million (blue line) and stringency index (red line) evolving with time for (a) Denmark, (b) Sweden, (c) Finland and (d) Norway respectively

increased severity of epidemic can cause an increase of stringency index, as governments make response when seeing increased number of new cases. Therefore we are more interested in the analyses with stringency index as an independent variable and total cases per million as depedent. Our previous analyses has indicated that different measure temporal structures lead to different cases increase. Here we will apply a formal statistical causality test and analyze the lag days after which measures take effects.

**Granger Causality Test:** *Let $y$ and $x$ be stationary time series. To test the null hypothesis that $x$ does not Granger-cause $y$, one determines if $y$ can be predicted by include in a univariate autoregression of $y$ and lagged values of $x$:*

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdot a_m y_{t-m} + b_1 x_{t-1} + \cdot + b_q x_{t-q} + error_t \tag{2}$$

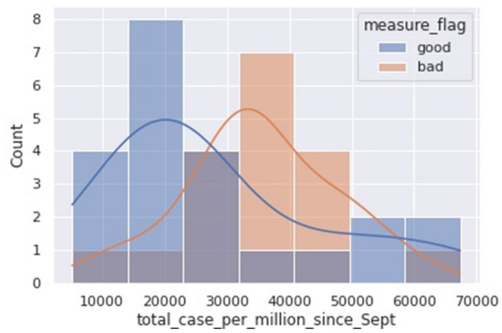


Figure 6: Distribution of total cases per million in European countries from Sept. 1st 2020 to Jan. 4th 2021. The blue line shows fitted distribution for countries with "good" measures and red line shows fitted distribution for coutries with "bad" measures.

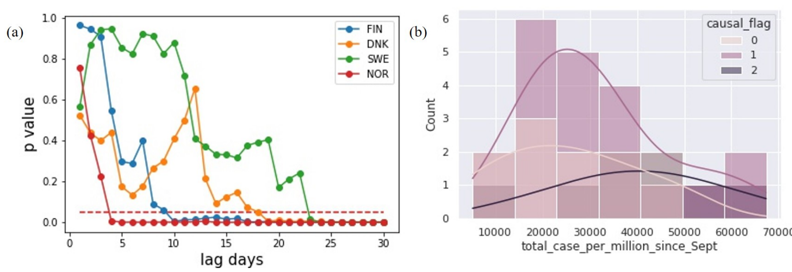


Figure 7: (a) Convergence plot for four European countries as a function as maximum lag days as an input parameter for Granger test. Red dahsed line shows 0.05 significance level. (b) Histogram of increase cases since September for European countries with different lag days.

We first apply a third-order Savitzky–Golay filter with a time window of 7 days on the stringency index data since it is not continuous contains high-order frequency component. Then we take differences on both stringency index and new cases smoothed per million to make them stationary. We tested if the two series is stationary by ADF null hypothesis (which assumes there is a unit root and series is non-stationary) and KPSS null hypothesis (which assumes there is no unit root and the series is stationary). We tested the approximate lag length with vector autoregression(VAR) model.

After that, we conducted F-test and and chi square test for Granger non-causality. If the p-value is below 0.5%, we reject the null hypothesis and conclude that x is a Granger cause to y. If p-value is above 0.5%, we conclude that x does not cause y.

As the first step we performed Granger causality test on the four European countries we mentioned before. As shown in Fig. 7(a), within 30 lag days the p values of all these four countries go below 0.5%, which verifies that stringency index can affect the new cases. We also noticed that the convergence rate is different for the four countries. For countries where epidemic is better controlled (Finland and Norway), the p-value converge faster.

We run Granger tests all interested European countries. The results show that within 30 lag days, stringency index is a cause to new cases for all countries. We define the lag day for each country as the minimum day for p-value to go below 0.5%. The observation above implies that it is possible to approximate the lag day as the time the measures need to take effects. We define the causal flag to be 0 if lag day is smaller that 3, to be 2 if lag day is between 4 and 10 and to be 3 if lag day is larger than 10. We plot histogram of total new cases for each of the three classes, as shown in 7(b). It shows that country with smaller lag days tends to have fewer new cases, which is consistent with our analysis.

In conclusion, we tested the causal relationship between the measures and the number of new cases with stringency index as an overall index. We also built up several indicators to predict if the measures a countries take is effective. The results show that the temporal structure of measures can have an dominant effect over other social, cultural and economical factors for the control of coronvirus.

# 4 Structured Time Series Prediction

## 4.1 Measures in Different Country

In the previous section, we analyze the causality between stringency index and the spread of the virus. Here, we decompose the stringency index into different concrete measures and try to investigate their effects.

As shown in figure.8, different countries have taken different measures. Different measures were taken at different times, which in turn corresponding to different infection conditions.

## 4.2 Model

Ideally, a controlled test is the most solid way to investigate the effect of different measures. However, in reality, such a controlled test is not always available. Therefore, here we resort to the structured time series model.

The general format of a structured time series (STS) model is the following:

$$f(t) = \sum_i f_i(t) + \epsilon \tag{3}$$

Here, $\epsilon \propto N(0, \sigma^2)$ is a random noise and each $f_i(t)$ represent component.

Specifically, for this model, we consider the 9 categories of measures together with the daily average temperature data for France and Norway between 2020.6.1 and 2021.1.4. i.e.

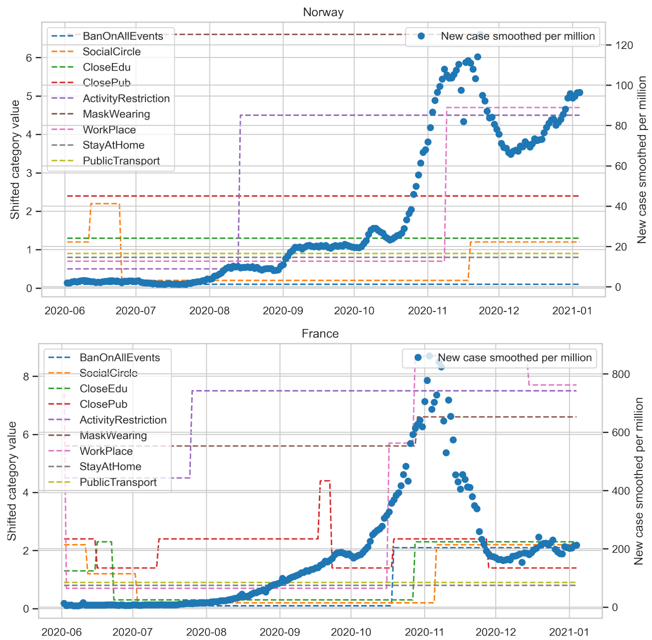


Figure 8: Measures and new case smoothed per million people for Norway and France after 2020.6.1. Notice that the category values of different categories are shifted to demonstrate the change better. Specifically, from "BanOnAllEvents" to "PublicTransport", the cagetory value are shifted by 0.1 to 0.9 respectively for each category.

$$f(t) = f_{temp}(t) + \sum_i f_{i^{th}\ cat}(t, t_{lag}) + \epsilon \tag{4}$$

where $f_{temp}(t)$ represent the effect of the average temperature and $f_{i^{th}\ cat}$ represents the effect of the $i^{th}$ category of measures shown in the Appendix. We lagged $f_{i^{th}\ cat}$ with $t_{lag}$ which is the optimum lag day calculated in Causal inference section

The effect of the temperature on the spread of the virus has been controversial. There are plenty of papers[4, 5, 6, 7] on both sides arguing whether temperature significantly determines the spread of the virus. On the one hand, apparently, the two major global outbreaks of the virus are around winter time where the temperature is significantly lower than June where the global positive rate reaches it's minimum after the first outbreak around 2020.2. Therefore, we decide to include the effect of the temperature. The data of the average temperature per day per country is obtained from ref.[3].

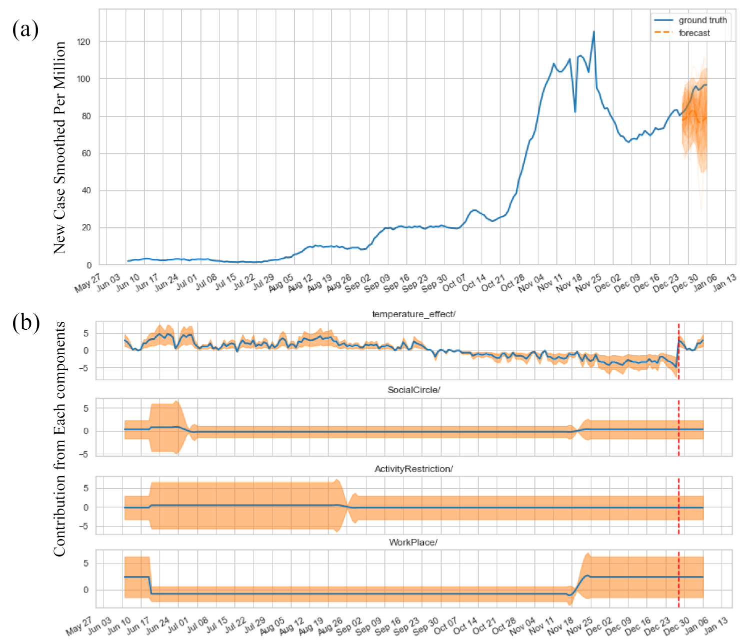


Figure 9: (a) Prediction of the STS model. (b) Contributions from each components. Here, we only show components with non-zero contributions.

### 4.3 STS Analysis and Predictions

With the model specified above, the spread of the virus was first investigated for Norway. As shown in figure.9(a). The this model reasonably predicts the the dynamics of the spread of the virus. The contributions of each components are shown in figure.9(b). According to this decomposition, policies on "SocialCir", "ActivitiyRestriction" and "WorkPlace" shows significant influence on the new cases number per million people and the influence of the temperature is less significant. For the prediction, most uncertainty comes from the evolution of the virus and smaller influence from the temperature uncertainty.

We also investigate the spread of the virus in France. France has more population than Norway and has taken much more different measures to control the virus. According to the STS analysis, almost all measures shows noticeable effects to stop the spread of the virus. The influence from the temperature is not significant.

## 5 Conclusion

Our analysis on European epidemic has the following findings:

First, active general public health measures can have significant impact to protect the general population and control the total infection rate. It can be

(a)

(b)

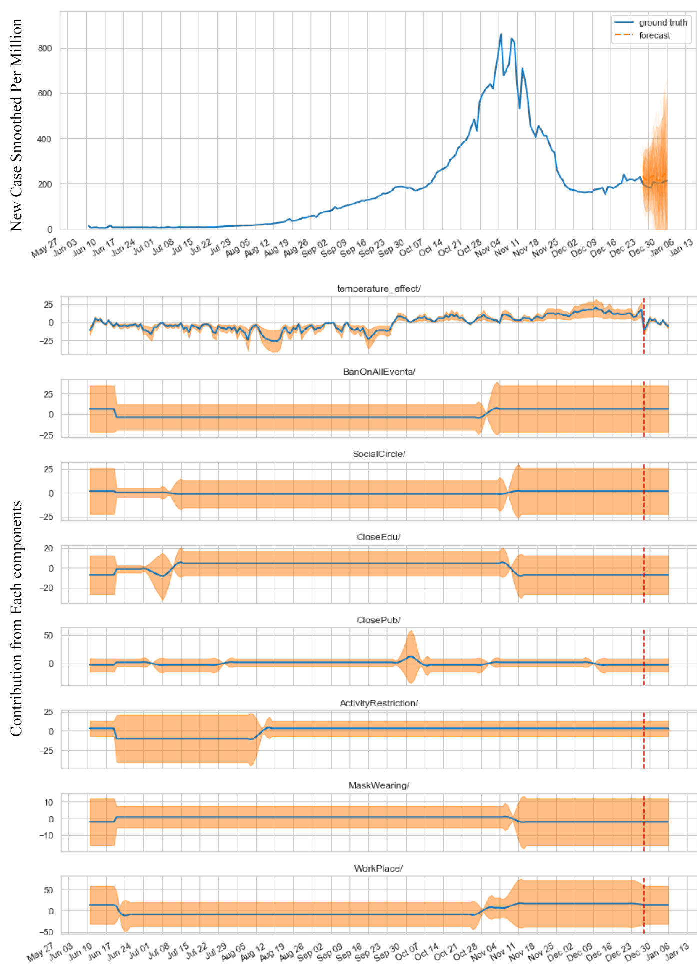


Figure 10: (a) Prediction of the STS model. (b) Contributions from each components. Here, we only show components with non-zero contributions.

dominant over the effects local temperature economical and social development level.

Second, the STS model are able to predict the infection rate in the near future with an acceptable uncertainty.

Third, Some measures are more efficient than others in certain countries. Therefore, with historical data and given measures, this prediction model may provide helpful information for the governments to get a better control of the spread of the virus.

# References


[1] Covid-19: Government stringency index. https://ourworldindata.org/grapher/covid-stringency-index?tab=table&stackMode=absolute&time=2020-01-22..latest&region=World. Accessed: 2021-02-21.

[2] If most of your coronavirus tests come back positive, you're not testing enough. https://www.npr.org/sections/coronavirus-live-updates/2020/03/30/824127807/if-most-of-your-coronavirus-tests-come-back-positive-youre-not-testing-enough. Accessed: 2021-02-21.

[3] Weather api and data for businesses and developers. https://www.worldweatheronline.com/developer/. Accessed: 2021-02-21.

[4] Alessio Notari. Temperature dependence of covid-19 transmission. *Science of The Total Environment*, 763:144390, 2021.

[5] Peng Shi, Yinqiao Dong, Huanchang Yan, Chenkai Zhao, Xiaoyang Li, Wei Liu, Miao He, Shixing Tang, and Shuhua Xi. Impact of temperature on the dynamics of the covid-19 outbreak in china. *Science of the total environment*, 728:138890, 2020.

[6] Aditya Stanam, Mamata Chaudhari, Durga Rayudu, et al. Effects of temperature on covid-19 transmission. *Medrxiv*, 2020.

[7] Jingui Xie and Yongjian Zhu. Association between ambient temperature and covid-19 infection in 122 cities from china. *Science of the Total Environment*, 724:138201, 2020.


# Appendix A: Categorization of Measures

All measures taken by European countries are summarized and categorized in table.5. Each measure are assigned with a value within this category manually. The value is determined based on the strictness of this measures according to the authors' impression. When no measures are taken in a specific category, the value is set to be 0 for that category.

# Appendix B: External Code and Data

External data and code are used to perform the analysis. Specially, the following code and website provide the majority of the external code and data we have reused for this project.

- Temperature code: covid19/weather_data_extraction

- Temperature source: Weather API and Data for Businesses and Developers
- Structured time series: tensorflow/probability

Table 1: Category and values of measures, Part 1

| Measure | Category | Value |
|---|---|---|
| BanOnAllEventsPartial | BanOnAllEvents | 1 |
| BanOnAllEvents | BanOnAllEvents | 2 |
| SocialCirclePartial | SocialCircle | 1 |
| SocialCircle | SocialCircle | 2 |
| ClosDaycarePartial | CloseEdu | 1 |
| ClosHighPartial | CloseEdu | 1 |
| ClosPrimPartial | CloseEdu | 1 |
| ClosSecPartial | CloseEdu | 1 |
| ClosDaycare | CloseEdu | 2 |
| ClosHigh | CloseEdu | 2 |
| ClosPrim | CloseEdu | 2 |
| ClosSec | CloseEdu | 2 |
| ClosPubAny | ClosePub | 2 |
| ClosPubAnyPartial | ClosePub | 1 |
| EntertainmentVenues | ClosePub | 2 |
| EntertainmentVenuesPartial | ClosePub | 1 |
| GymsSportsCentres | ClosePub | 2 |
| GymsSportsCentresPartial | ClosePub | 1 |
| HotelsOtherAccommodation | ClosePub | 2 |
| HotelsOtherAccommodationPartial | ClosePub | 1 |
| PlaceOfWorship | ClosePub | 2 |
| PlaceOfWorshipPartial | ClosePub | 1 |
| RestaurantsCafes | ClosePub | 2 |
| RestaurantsCafesPartial | ClosePub | 1 |
| IndoorOver50 | ActivityRestriction | 4 |
| IndoorOver100 | ActivityRestriction | 3 |
| IndoorOver500 | ActivityRestriction | 2 |
| IndoorOver1000 | ActivityRestriction | 1 |
| MassGather50 | ActivityRestriction | 4 |
| MassGather50Partial | ActivityRestriction | 3 |
| MassGatherAll | ActivityRestriction | 2 |
| MassGatherAllPartial | ActivityRestriction | 1 |
| OutdoorOver50 | ActivityRestriction | 4 |
| OutdoorOver100 | ActivityRestriction | 3 |
| OutdoorOver500 | ActivityRestriction | 2 |
| OutdoorOver1000 | ActivityRestriction | 1 |
| NonEssentialShops | ActivityRestriction | 2 |
| NonEssentialShopsPartial | ActivityRestriction | 1 |
| PrivateGatheringRestrictions | ActivityRestriction | 2 |
| PrivateGatheringRestrictionsPartial | ActivityRestriction | 1 |

Table 2: Category and values of measures, Part 2

| Measure | Category | Value |
|---|---|---|
| MasksMandatoryAllSpaces | MaskWearing | 8 |
| MasksMandatoryClosedSpaces | MaskWearing | 7 |
| MasksMandatoryAllSpacesPartial | MaskWearing | 6 |
| MasksMandatoryClosedSpacesPartial | MaskWearing | 5 |
| MasksVoluntaryAllSpaces | MaskWearing | 4 |
| MasksVoluntaryClosedSpaces | MaskWearing | 3 |
| MasksVoluntaryAllSpacesPartial | MaskWearing | 2 |
| MasksVoluntaryClosedSpacesPartial | MaskWearing | 1 |
| Teleworking | WorkPlace | 6 |
| TeleworkingPartial | WorkPlace | 5 |
| WorkplaceClosures | WorkPlace | 4 |
| WorkplaceClosuresPartial | WorkPlace | 3 |
| AdaptationOfWorkplace | WorkPlace | 2 |
| AdaptationOfWorkplacePartial | WorkPlace | 1 |
| StayHomeOrder | StayAtHome | 8 |
| StayHomeOrderPartial | StayAtHome | 7 |
| RegionalStayHomeOrder | StayAtHome | 6 |
| RegionalStayHomeOrderPartial | StayAtHome | 5 |
| StayHomeGen | StayAtHome | 4 |
| StayHomeGenPartial | StayAtHome | 3 |
| StayHomeRiskG | StayAtHome | 2 |
| StayHomeRiskGPartial | StayAtHome | 1 |
| ClosureOfPublicTransport | PublicTransport | 2 |
| ClosureOfPublicTransportPartial | PublicTransport | 1 |