Cover Letter

This study looked at how influenza can be prevented by accurately predicting an outbreak. By collecting health and climate data, it is possible to find correlations between health and climate, such as the concentration of carbon dioxide, temperature, wind speed, humidity, infection rate, and vaccine rate. From these results, we can determine the path coefficients. The most important aspect is that we can integrate (linear) algebra, such as mapping concepts, into statistical theory and operational research (Bayesian game theory), resulting in this author's butterfly effect philosophy (Lam, Oct 2018). With the use of a (functional) gradient descent, the butterfly effect philosophy can be applied to time series analysis in the example of temperature and infection rate to determine a best-fit line between them. As such, from a theoretical perspective, a well-calibrated gradient can be determined, as well as a best-fit plotted graph. By ascertaining accurate values for correlation, path, factor coefficients, and linear regression (standardised coefficients), a prototyped statistical model – and later a structural equation model or mathematical one relating to carbon dioxide emissions, temperature, wettest (humidity), wind speed, and influenza infection rate – can be developed. Eventually, the likelihood of future influenza outbreaks can be determined. A simulation of the outbreak would also be able to be performed. Thus, vaccination strategy and whether it is effective could also be evaluated. Indeed, the stochastic gradient descent can be used for further computational modelling and hence important values will be found. Finally, a well-refined simulation model of the relationship between weather and influenza outbreaks will be created. That said, adjustments to the elementary prototyped model may be required, and comprehensive decisions can be determined regarding whether vaccination would be necessary. While there is already a mathematical mutation model – based on frequency and fitness data at any given point in time (Marta & Michael, 2014) – which predicts the upcoming year's type of influenza, it lacks an explanation as to why a sudden

high rate of outbreak occurs in a particular year but not a normal yearly patterned rate. To be more specific, the model fails to take into consideration certain facts, such as human-weather interaction. One such example is why there is a higher chance of people catching influenza in winter, with cold and relatively dry climatic conditions (Jianhong et. al., 2014).

The most important discovery is that when matrix analysis (or linear algebra) is combined with statistics, they form a new field of data science. Taking this a step further, we have a principal component analysis mixed with matrix theory, which results in the big data analysis of statistics (i.e., the main focus of this study and the rationalisation of my butterfly effect philosophy). Additionally, this philosophy can also be positioned alongside other academic theories. These theories are in fact regression approximation concepts in the fields of statistics and mathematics as contributions. Hence, to some extent, this event is similar to Einstein's integration of mathematical tools into classical physics; this is how the theory of relativity was born from mathematical physics.

Indeed, I am the only author of this paper and, as far as I know, I have quoted all relevant material with corresponding suitable citations. Whenever and if it is possible, I discover any new and important un-cited material for this paper after publishing, I will send an email to make any necessary amendments.

Yours Faithfully,

Lam Kai Shun

Declara'on:

I declare that all of the contents presented in this paper are purely my work and that all of the references have been well quoted under corresponding citations (as I have known). There is no conflict of interest in this paper. There is also no funding sources for this paper. I thank my former Department's professor — Prof. Siu Man Keung and Dr Leung Kam Tim. I also thank the library of the University of Hong Kong for her kindly borrowing of the related books as this paper's references work. The library inspires me very much.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Forecast the prediction or predict the forecast — A precise view of investigating Butterfly Effects

Lam kai Shun

h9361977@connect.hku.hk

B.Sc., M.Sc., M.Sc. (HKU)

Abstract: Influenza infection is a large global issue. In Hong Kong, the common stages of an influenza outbreak are a decrease in temperature (during winter), which leads to the avian flux, then to the animal flux, and gradually infects humans—leading to an outbreak. However, in Australia, the story is quite different. Excessive carbon dioxide emissions cause extreme weather changes and climatic factors such as strong wind, temperature change, and wet weather. The result is an out- break of human influenza. These factors constitute a series of correlated sequences. The outcome gives rise to a Bayesian decision tree, which predicts the probability of taking vaccine given that the per- son is infected. This predictive vaccine efficacy is based on Bayes' probability, rather than the tradi- tional definition of efficacy or effectiveness. This is the definition of "forecasting a prediction".

Conversely, if Bayes' probability tree is understood, and the same categorised results are grouped by a random variable, the consequences of the domino effect will be formed. More specifically, this "predicts the forecasting". One of the aforementioned butterfly effect's applications (besides an influenza outbreak) is the prediction of an earthquake. It can forecast an earthquake together with its corresponding feasible predicting probability and issue necessary warnings in advance, according to daily phenomena. This act has the potential to save millions of lives.

Background

Humans have suffered from the influenza virus (the flu) for hundreds of years. The situation becomes even worse when sudden outbreaks occur, such as the 1889 Russian flu (1 million deaths), the early 20th Century Spanish flu (20–100 million), the 1950s Asian flu (1–1.5 million), the 1960s Hong Kong flu (0.75–1 million), and the 2009 flu pandemic (swine flu), which accounted for several hundred thousand deaths [1].Although the number of people infected has decreased since the introduction of the influenza vaccine, there is still a chance that future vaccines might be ineffective. Hence, a more accurate prediction mechanism is required in order to prepare for new strains of the influenza virus. In Hong Kong, our present public healthcare system, this will decrease pressure if there is another outbreak. Indeed, all influenza viruses have two types: H and N that ranges from (1–10) and (1–7). The various combinations of virus mutation are extremely large, and therefore it is difficult to predict future flu types. As such, finding the right vaccine is crucial.

This paper attempts to use Australia's human-weather data for influenza infection rates to see if there are any statistical correlations. The major aim is to find out whether there will be a sudden peak outbreak of an influenza virus. One of the defects in the current method of predicting common types of influenza diseases is the human-weather interaction factor. Thus, it is believed that when these factors are counted, it would be possible to develop an improved mixed vaccine to minimise the rate of infection. Moreover, one would also be able to determine the probability of being infected, based on deciding whether or not to take a vaccine. This would greatly reduce the risk of infection during peak outbreaks. This is similar to predicting earthquakes; by observing and detecting the phenomena of pre-earthquake events, one could determine the probability of a cat- astrophic earthquake.

1. https://en.wikipedia.org/wiki/Influenza#History

Literature Review

The author's previous studies on the butterfly effect originated from set theory, probability, and linear algebra. This conjecture is indeed quite predictive in nature. As such, it may be classified as a type of "predictive-ism" or the philosophy of prediction. It should be noted that forecasting is based on analysis, and prediction foresees an event before it actually happens. Clearly, prediction is more subjective and fatalistic in nature. In this study, the use of "predictive" is associated with "forecasting". This is because Bayes' theorem has some prediction elements, while the domino effect has forecasting (predicted) consequences. From a sequences of forecasted events, Bayes' prediction can be applied directly. More specifically, it is possible to forecast a predicted probability and predict the forecasting events by applying the proposed butterfly effects theorem into different situations.

The Historical Evolution of Forecasting

Throughout ancient history, it was common for people to try and predict the future. They did so for various reasons such as improving hunting and foraging , etc. This eventually extended into augury, such that one may divine the future. For example, a sign of an approaching storm is when seagulls stop flying. During the Roman Empire, the art of augury was applied to key social and military decisions, which were based on omens from the observed flight of birds. There were also many other divination practices such as haruspicy, where the entrails of sacrificed animals were inspected in order to predict the future.

Later, from the Classical to Middle Ages, sticks, beans, and other materials were drawn at random by a person from a set of collection—this is known as sortilege or cleromancy. Different cultures had various types of such drawings, from Judeo-Christian traditions (appears in the Bible) to the Chinese "I Ching" tradition (descended from bone divination). During "I Ching," coins were tossed and the subsequent texts were interpreted based on where the coins landed. In the process of

interpreting texts, both conscious and unconscious minds were used to solve problems. In such a case, some constraints on conventional thought was loosened, allowing for more innovative solutions and decisions to be achieved. This is known as the Age of Scrutinising Symbols.

Following this was the Renaissance period, where scientists began to use mathematics to study nature. Scientists such as Galileo Galilei combined logic, mathematics, and empirical observations to describe various aspects of nature. Galileo was able to demonstrate that the earth rotated around the sun, which was a ground-breaking milestone in astronomy. Subsequently, Isaac Newton and Gottried Leibniz developed calculus, which was a huge step forward in the area of forecasting. For these pioneers of science, the subject of calculus provided a "framework for modelling system where there is a change, and a way to deduce the predictions of such models" (K. Daniel)[2] .In fact, this thesis uses a statistical mathematical model, which will be explained in more detail in a later section.

Statisticians appeared in the 18th century, when statistics was invented as another branch of mathematics. Although some statistical methods had already been used for around two millennia, probability only emerged in the 17th century. Modern statistics did not arise until the late 19th to early 20th century. The Pearson product-moment correlation coefficient was introduced by Karl Pearson. Other key concepts such as standard derivation, correlation, and event regression analysis were developed by John Galton. Finally, Ronald Fisher contributed the null hypothesis together with other important mathematical principles. Since then, these principles have been fundamental in data science, machine learning, and predictive analysis.

2. http://foresightr.com/2016/05/06/a-brief-history-of-forecasting/

After World War II, a combination of imagination and trend-watching was used for future

feasible developments. From the late 20th century, predictive analytics and data-based forecasting were created. Many computer-based models—from weather tracking to credit risk analysis—were invented. This subsequently led to predictive models being formulated. Finally, the introduction of more structured and quantified forecasts appeared in the early 21st century.

Models

There are three types of models that can be used to represent objects or systems with similar properties. This thesis will apply a statistical model using linear regression to represent human-weather relationships in Australia-based influenza cases.

In fact, the advantage of using a statistical models is that there is no need to have a thorough understanding of the governing mechanism (Wong, 2009). The dependent factor (influenza infection rate) relies heavily on other factors such as wind speed, average temperature, and wettest (humidity). Based on the data quality, a statistical model is the best tool to use. That said, it would be incorrect to apply data from a wet day to predict corresponding data on dry days. This paper's data-driven modelling approach uses multiple linear regression. More specifically, the outcomes of the statistical data can be viewed as probability distributions, rather than unique values only.

A deter- ministic model is one where the values of dependent variables of the system are completely deter- mined by the parameters of the model. The advantage is that it can be amended according to math- ematical analysis, such as in the case of using ordinary equations to model the geographic spread of an infectious disease within a human population.

• 3 https://fs.blog/2018/09/bayes-theorem/

Disadvantages of the Prediction

One disadvantage of Bayes' theorem is that its prediction is in terms of a value or only a probability. Therefore, it is possible for a predicted event to not happen. In simple terms, a high

probability of something being true is not the same as being true. For example ,

"A horse which has been often driven along a certain road resists the attempt to drive him in a different direction. Domestic animals expect food when they see the person who usually feeds them. We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken."

However, Bayesian reasoning states:

"After you've been steeped in Bayes' rule for a little while, it starts to produce some fundamental changes to your thinking. For example, you become much more aware that your beliefs are grayscale. They're not black and white and that you have levels of confidence in your beliefs about how the world works that are less than 100 percent but greater than zero percent and even more importantly as you go through the world and encounter new ideas and new evidence, that level of confidence fluctuates, as you encounter evidence for and against your beliefs."

That is the value of Bayes' theorem and the proposed butterfly effect theory.

A more generalised view of proposed Butterfly Effect Theory

For any event A1, A2, …, An, assumes each event branched with (A1, A2,…, An) and n levels, Then for those grouped final outcomes A1, A2, …, An, one may have A1,…,An ac-cording to their proper-

ties in the following ways:

$[A_1]$ ->…-> $[(A_1|A_1A_1\ldots A_1), (A_1|A_2A_1\ldots A_1), \ldots, (A_1|A_nA_n\ldots A_n)]$ -> RV_1 -> DE_1

$[A_2]$ ->… -> $[(A_2|A_1A_1\ldots A_1), (A_2|A_2A_1\ldots A_1), \ldots, (A_2|A_nA_n\ldots A_n)]$ -> RV_2 -> DE_2

. . .

. . .

. . .

$[A_i]$ ->…-> $[(A_i|A_1A_1\ldots A_1), (A_i|A_2A_1\ldots A_1), \ldots, (A_i|A_nA_n\ldots A_n)]$ -> RV_i -> DE_i

. . .

. . .

. . .

$[A_n]$ ->…-> $[(A_n|A_1A_1\ldots A_1), (A_n|A_2A_1\ldots A_1), \ldots, (A_n|A_nA_n\ldots A_n)]$ -> RV_n -> DE_n

More specifically, each outcome event must be grouped with similar properties together with a suitable random variable (RV). Each of these variables are followed by a series of domino events. Each event can be discovered through scientific research or statistical methods such as correlations with the main event under investigation. Thus, the generalised probability tree diagram, random variables, and the consequence domino effects become :

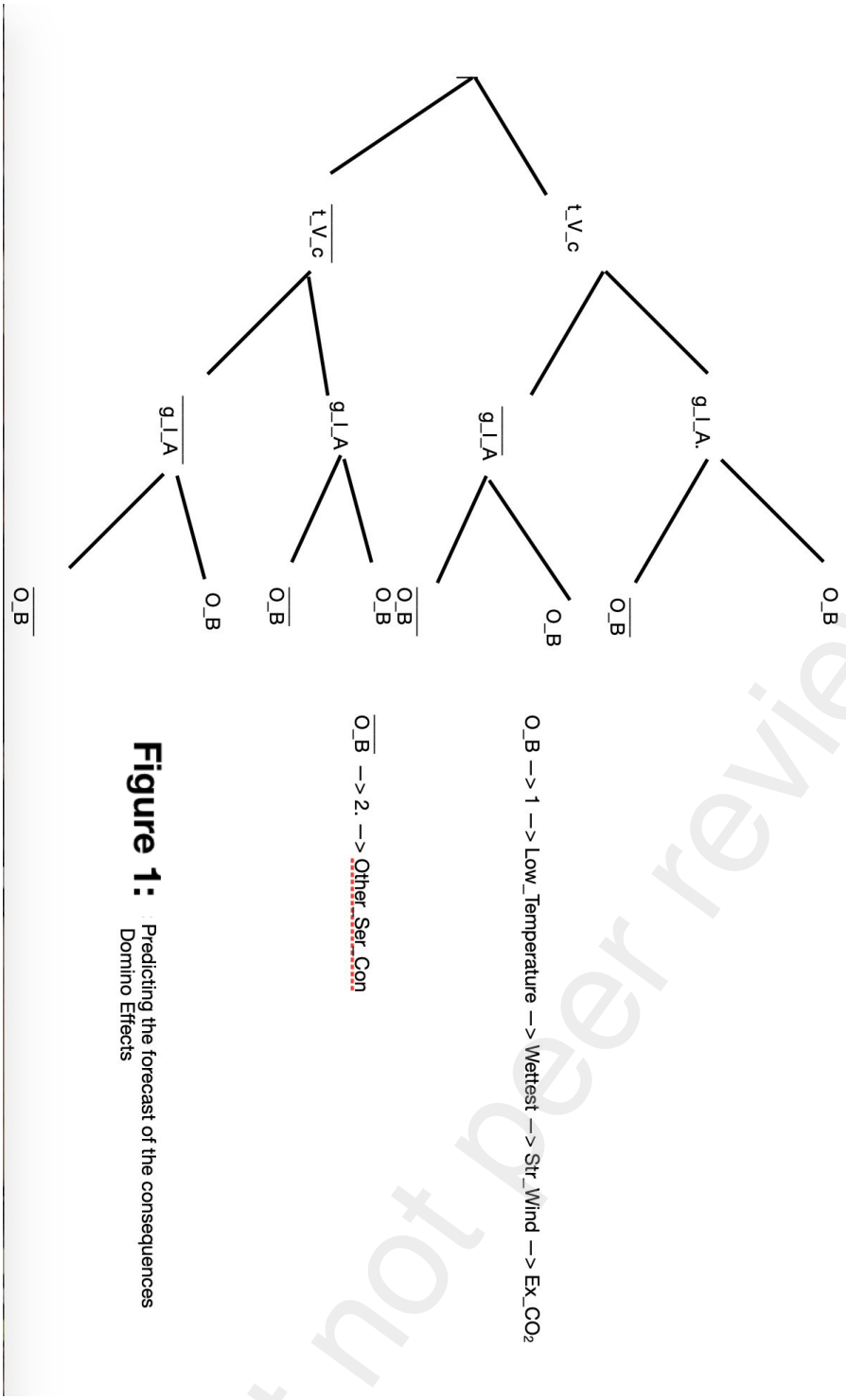


Figure 1: Predicting the forecast of the consequences Domino Effects

Key : Ex_CO2 — Excess carbon dioxide emitted; Str_Wind — Strong Wind;

O. B. — Out Break; G. I. A. — Get Influenza Affected;

t_V_c — Take Vaccine

2. only if Part of the proposed butterfly effect theory:

For each series of domino events (each event can be discovered through scientific research, or by showing a correlation with former events), one may then investigate backwards through a RV to the categorised events (in terms of vector and matrix) such as the following:

DE1 -> (RV1) -> [(A1|A1A1…A1), (A1|A2A1…A1), …, (A1|AnAn…An)] ->…-> [A1]

DE2 -> (RV2) -> [(A2|A1A1…A1), (A2|A2A1…A1), …, (A2|AnAn…An)] ->…-> [A2]

.

.

.

DEi -> (RVi) -> [(Ai|A1A1…A1), (Ai|A2A1…Z1), …, (Ai|AnAn….An)] ->…-> [Ai]

.

.

.

DEn -> (RVn) -> [(An|A1A1…A1), (An|A2A1…A1), …, (An|AnAn…An)] ->…-> [An]

Thus, the corresponding forwarding consequence domino effects, random variables and probability tree diagram will be drawn in the coming page as shown below:

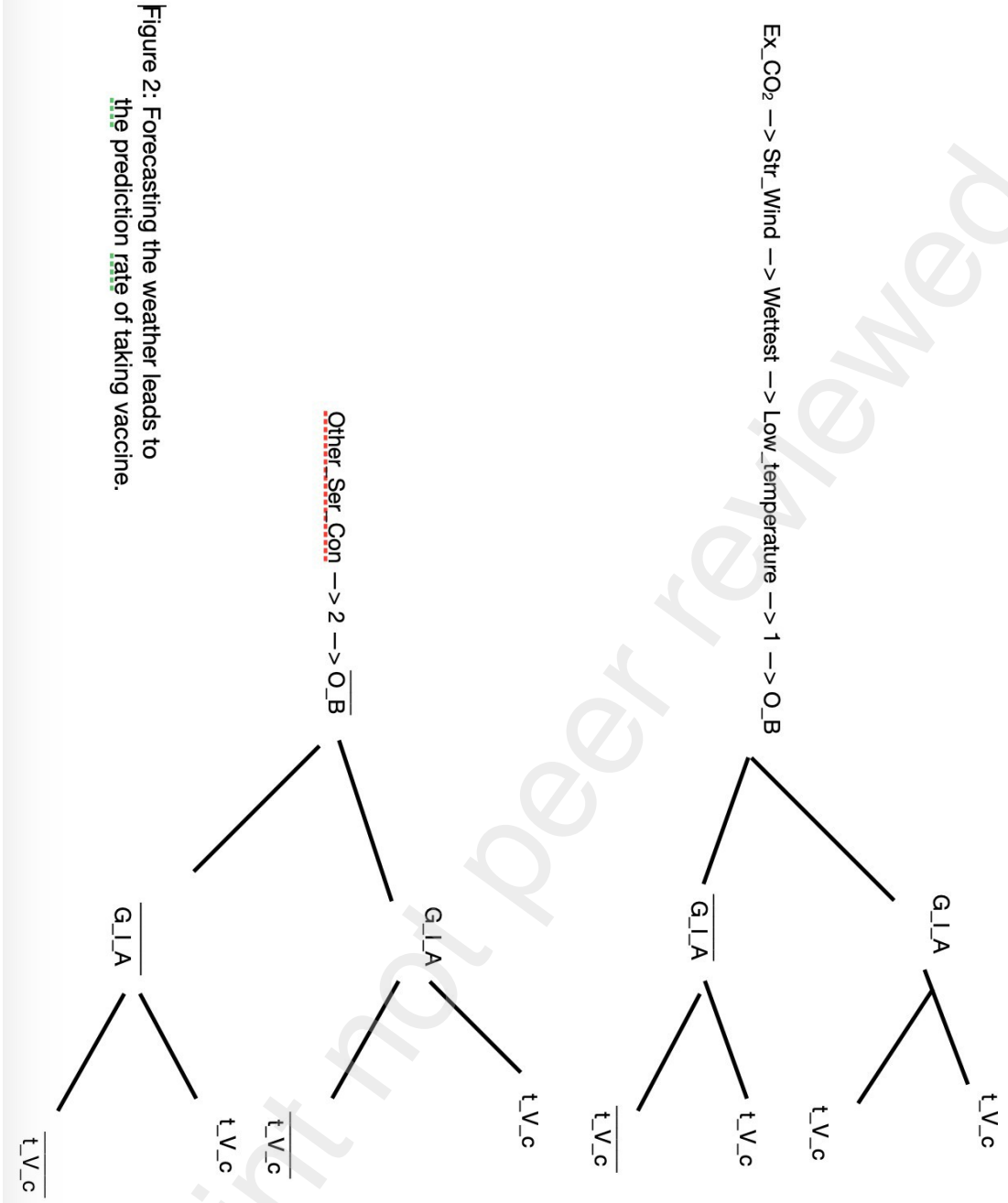


Figure 2: Forecasting the weather leads to the prediction rate of taking vaccine.

Key : Ex_CO2 — Excess carbon dioxide emitted; Str_Wind — Strong Wind;

O. B. — Out Break; G. I. A. — Get Influenza Affected;

t_V_c — Take Vaccine

Domino Events: Previous Incidents of an Influenza Virus Outbreak

Based on historical influx data, it was discovered that climate change (caused by excess CO2 emissions) may lead to bird flu, and then mutates into another type of animal flu, causing a recombination of genes or sudden mutation. This ultimately leads to another outbreak of the influenza virus. The best time for this to occur is during cool temperatures (around 15–20 degrees Celsius) with high levels of humidity (very wet). Hence, the domino events are the following:

Emission of CO2-> climate (cool + decrease in wettest)-> bird flu-> animal flu-> human flu.

However, there are some areas where bird flu and animal flu do not apply, such as Australia. In such a case, the domino events should be amended to:

Emission of CO2-> climate (cold + decrease in wettest + strong wind)-> human flu.

In addition, the opposite of the above can also be true.

For instance, the peak (or sudden extreme) outbreak of the human influenza virus that occurred in southern Australia in 2015 and 2017. Therefore, other than bird and animal flu, it is highly likely that climate factors (particularly strong winds) determines the rate of transmission for the human influenza virus. A pioneering study will use Australia's weather statistics in an attempt to develop a prototyped model between climate (regression analysis) and influenza rate. Later a more precise structural equation model will be used to fully describe the relationship between weather and the rate of influenza infection in Australia. Hence, the aim of this research is to show that human-weather interaction should be considered during the prediction of a peak influenza outbreak, rather than just adopting virus mutation frequency and fitness model (Michael et. Al., 2014).

Pioneering Research and the Major Results for Discussion

Pioneering research will first be performed using weather data from the Australian Bureau of Meteorology, which contains all the necessary materials. Looking at data from 2013–2018, this study considered nine climate-related factors: rate of carbon dioxide emission, number of infected, hottest day, coldest night, wettest day, warmest overall, coolest overall, wettest overall, and strong-

```
REGRESSION
  /VARIABLES= Wettest_1
  /DEPENDENT= Influenza
  /METHOD=ENTER
  /STATISTICS=COEFF CI R ANOVA BCOV
  /SAVE= PRED RESID.
```

Model Summary (Influenza)

R	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*
.39	.15	.14	.47

ANOVA (Influenza)

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Regression	2.22	1	2.22	10.24	.002
Residual	12.34	57	.22		
Total	14.56	58			

Coefficients (Influenza)

	Unstandardized Coefficients		*Standardized Coefficients*			*95% Confidence Interval for B*	
	B	*Std. Error*	*Beta*	*t*	*Sig.*	*Lower Bound*	*Upper Bound*
(Constant)	4.02	.14	.00	28.98	.000	3.74	4.30
Wettest_1	.00	.00	-.39	-3.20	.002	.00	.00

Coefficient Correlations (Influenza)

Model		Wettest_1
	Covariances	

```
REGRESSION
  /VARIABLES= Temperature
  /DEPENDENT= Influenza
  /METHOD=ENTER
  /STATISTICS=COEFF CI R ANOVA BCOV
  /SAVE= PRED RESID.
```

Model Summary (Influenza)

R	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*
.64	.40	.39	.39

ANOVA (Influenza)

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Regression	5.87	1	5.87	38.55	.000
Residual	8.69	57	.15		
Total	14.56	58			

Coefficients (Influenza)

	Unstandardized Coefficients		*Standardized Coefficients*			*95% Confidence Interval for B*	
	B	*Std. Error*	*Beta*	*t*	*Sig.*	*Lower Bound*	*Upper Bound*
(Constant)	5.22	.26	.00	19.92	.000	4.69	5.74
Temperature	-.09	.01	-.64	-6.21	.000	-.12	-.06

Coefficient Correlations (Influenza)

Model		Temperature
	Covariances	

Model Summary (Influenza)

R	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*
.25	.06	.03	.49

ANOVA (Influenza)

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Regression	.92	2	.46	1.89	.161
Residual	13.64	56	.24		
Total	14.56	58			

Coefficients (Influenza)

	Unstandardized Coefficients		*Standardized Coefficients*			*95% Confidence Interval for B*	
	B	*Std. Error*	*Beta*	*t*	*Sig.*	*Lower Bound*	*Upper Bound*
(Constant)	-8.23	6.30	.00	-1.31	.197	-20.86	4.40
SW_log	6.56	3.44	.74	1.91	.061	-.33	13.44
Str_wind	-.02	.01	-.74	-1.92	.060	-.03	.00

Coefficient Correlations (Influenza)

Model		SW_log	Str_wind
	Covariances	Str_wind	11.81

est wind. A prototyped model describing the relationship between the weather and the infected will then be created (mainly using path coefficients). This can be done through linear regression by using software such as PSPP.

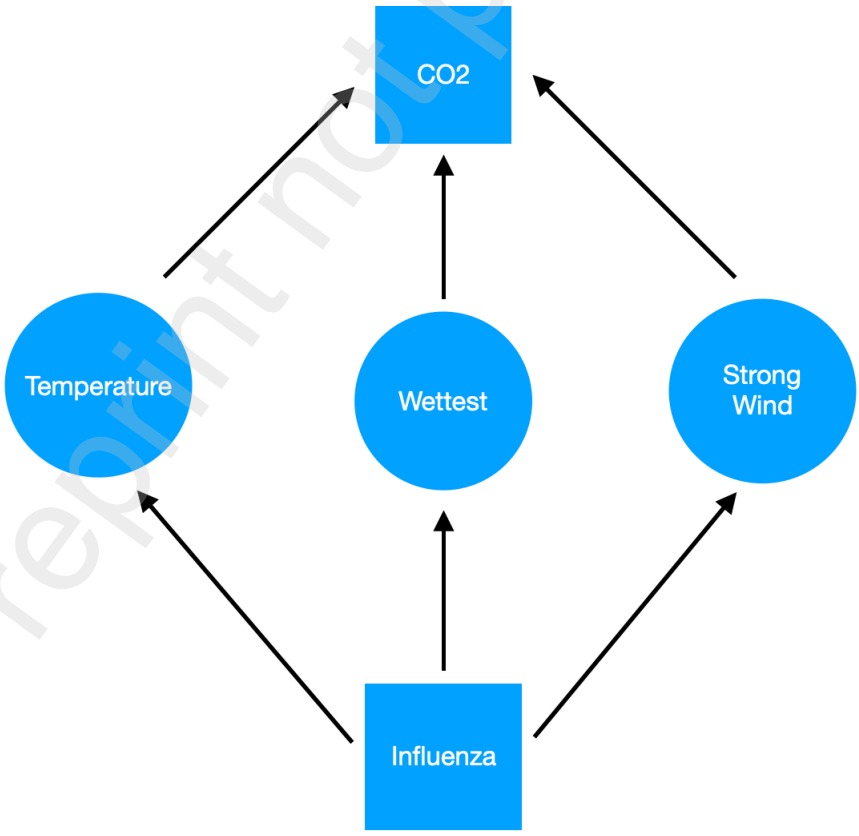
The following data was obtained:

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3400075

The predicted influenza infected rate over the coming 12 months are as follows:

3.19	**1548.81661891248**
3.20	1584.89319246111
3.20	1584.89319246111
3.41	2570.39578276886
3.60	3981.07170553497
3.92	8317.63771102671
4.25	17782.7941003892
4.10	12589.2541179417
3.92	8317.63771102671
3.59	3890.4514499428
3.55	3548.13389233575
3.30	1995.26231496888

Below is the suggested prototyped structural equation model.



The concentration of carbon dioxide in the atmosphere is affected by weather conditions such as temperature, wettest, and wind speed. These factors then relate to the number of people infected by influenza. The direction of the arrows can be both sided (upward and downward), which is reflected in the positive or negative values of the path coefficients (or, in this case, the linear regression parameters). The prototyped SEM (Structured Equation Modelling) constitutes this author's later parts of a forward predicted model—the first being a Bayesian probability tree diagram. It is obvious that the converse of the aforementioned arrow linkage is also true.

Other more precise models (or structural equation models) are described below. By using the statistical programming software R—using a linear regression approach—one is able to obtain two models: the first being the relation between the "number of people infected by influenza" and the "strongest wind speed, wettest (humidity), and average temperature"; the second being the connection between "the amount of carbon dioxide in the atmosphere" and the "strongest wind speed, wettest, and average temperature." The coefficients observed inside the paths were weighted and standardised. The aim being to overcome the multicollinearity among the variables other than the carbon dioxide concentration and the number of people with influenza.

The following computer plotted results as shown from R programming:

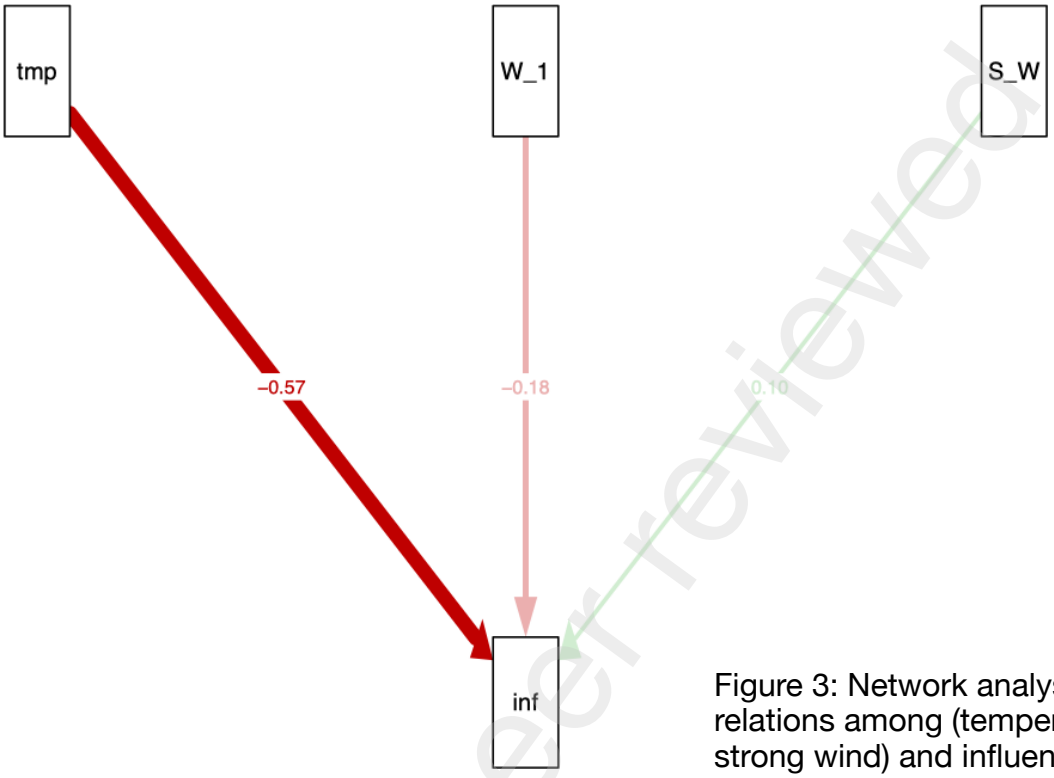


Figure 3: Network analysis for the causal relations among (temperature, wettest, strong wind) and influenza

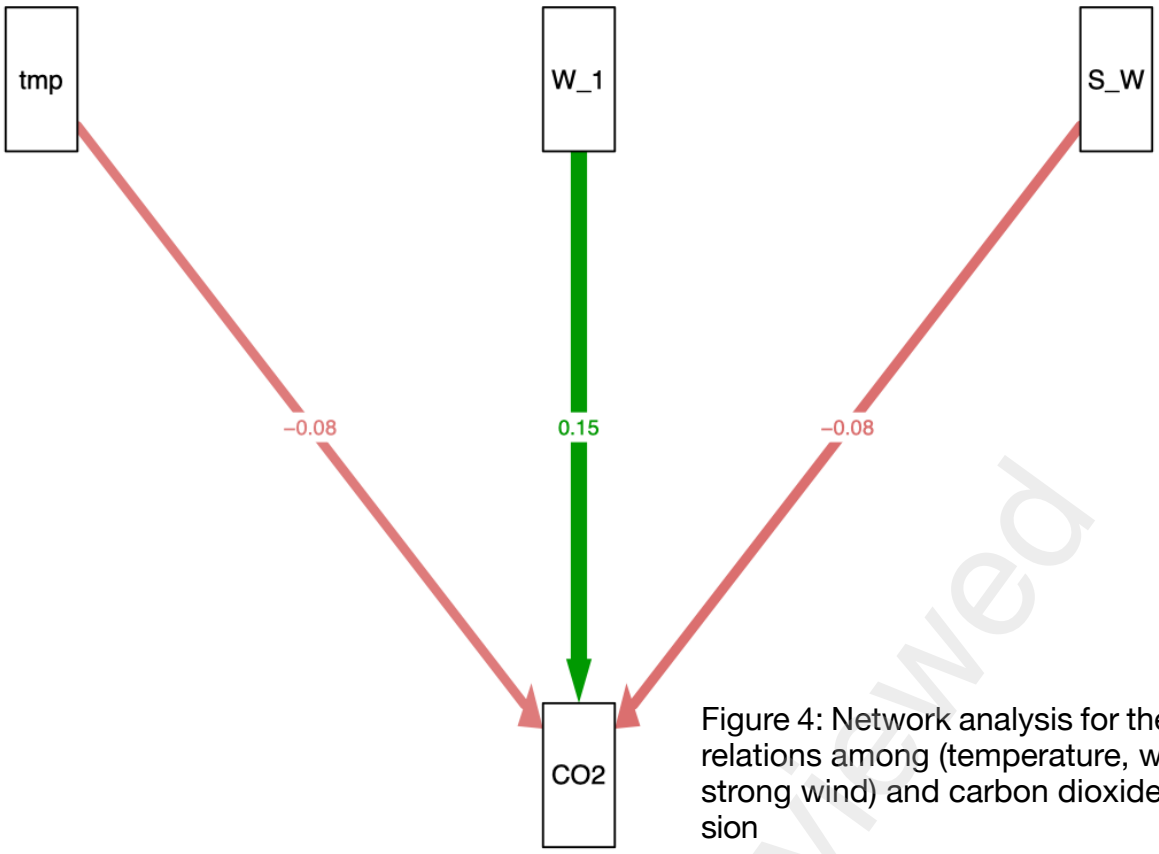


Figure 4: Network analysis for the causal relations among (temperature, wettest, strong wind) and carbon dioxide emission

This author notes that the negative values of the path—standardised and weighted linear regression parameter—means the increase in value "1" of the independent variable implies a decrease in the corresponding numerical value (as shown in the path) for the dependent variable. Therefore, an increase in one unit of carbon dioxide to the atmosphere will decrease 0.08 units of average temperature. This is because Australia is surrounded by ocean. The result is it can absorb some carbon dioxide. Other factors may also provide an abnormal modelling outcome in the concentration of CO2 and temperature. Similarity, a decrease in the average temperature by 0.57 °C will increase one unit of people infected by influenza. Simultaneously, a decrease in 0.18 unit of rainfall will increase one unit of people being infected. However, an increase in the strength of the wind will increase the number of infected. Hence, human-weather interaction in cold, dry, and strong wind environments means it is more likely for someone to get infected. This is because cold, dry weather can weaken a person's resistance to viruses, while strong winds will increase the speed of virus transmission.

Thus, the number of people infected with influenza increases and a sudden outbreak is more likely to occur under such conditions.

After the SEM is established, the simulation and predictions for the model will be performed using the gathered data. First, the aim of the simulation is to model Australia's human-weather interaction in relation to influenza. In other words, to develop a linear regression model for both prediction and simulation of human-weather interaction. Standard normal distribution was chosen for the random number simulator, since the infection rate and plotted weather data are in a time se- ries. To overcome randomness, a logarithmic function will be used to standardise the data. By cal- culating the mean and standard derivation of the infection rate, a linear regression model equation can be applied, which will use a 'sample' function in R for trialling further data. The Monte Carlo method was chosen, thereby determining the accuracy of the predicted values for future infection rates. These predicted values can be found through a linear regression model using PSPP software. Finally, with the linear model, it is even possible to calculate the accuracy of the prediction, to- gether with the accuracy of the model, through K-Fold Cross validation. The following are the R programming codes:

```
mysimulation <- matrix(nrow=24, ncol=1)
generat.path <- function(){months <- 12
set.seed(1)
simulation <- 24
i <- 0
for (i in seq(simulation)){
     changes <- rnorm(12, mean=3.615259, sd=0.49965)
     print(changes)
     sample.path <- -3.18*infected$temperature + 1.09*infected$Strong_Wind + 0.1*infected$Wettest + 6.16
     trials <- 200
     data1 <- sample(sample.path, size = trials, replace = TRUE, prob = NULL)
     sample.path1 <- cumsum(c(data1, changes))
     sample.path1
     print(i)
     sample.path1[1]
     sample.path1[2]
     sample.path1[months]
     mysimulation[i] <- sample.path1[months + i]
     print(mysimulation[i])
     i = i + 1
     print(i)
     }
final_test <- cbind(mysimulation, 1:24)
     print(final_test)
     plot(final_test)
     median.result <- median(final_test[,1])
     print(median.result)
     quan1.result <- quantile(final_test[,1], 0.95)
     print(quan1.result)
     quan2.result <- quantile(final_test[,1], 0.05)
     print(quan2.result)
}
```

The following R programming checks the accuracy of the prediction model:

```
set.seed(100)
trainingRowIndex <- sample(1:nrow(infected), 0.8*nrow(infected))
trainingData <- infected[trainingRowIndex, ]
testData  <- infected[-trainingRowIndex, ]
lmMod <- lm(influenza ~ temperature + Strong_Wind + Wettest, data=trainingData)
infectedPred <- predict(lmMod, testData)
summary (lmMod)
AIC (lmMod)
actuals_preds <- data.frame(cbind(actuals=testData$influenza, predicteds=infectedPred))
actuals_preds
correlation_accuracy <- cor(actuals_preds)
head(actuals_preds)
min_max_accuracy <- mean(apply(actuals_preds, 1, min) / apply(actuals_preds, 1, max))
mape <- mean(abs((actuals_preds$predicteds - actuals_preds$actuals))/actuals_preds$actuals)
library(DAAG)
install.packages("lattice")
library(caret)
fit1 <- lm(influenza ~ temperature + Wettest + Strong_Wind, data = infected)
cv.lm(infected, fit1, m=5)
library(tidyverse)
train.control <- trainControl(method = "cv", number = 5)
model <- train(influenza ~., data = infected, method = "lm", trControl = train.control)
print(model)
```

From the results, it seems that there will be an influenza outbreak in the coming July and August. This outcome is generally consistent with the data from the previous five years, where these two months (winter in Australia) will be cold, dry, and have strong winds. Hence, why there is a high infection rate. This author remarks that one can calculate the coming year's influenza infected rate

from first summing up the previous predicted number of flu cases being infected and then divided by the predicted total Australia population. Hence, one may get the predicted probability of being infected next year. In addition, one may also forecast the rate of people who will take vaccine in the coming year and so as the corresponding probability.

Next, the model and the accuracy of the other predicted data using R will be discussed. Looking at the first function ‘general.Path[ ),' the median value of the cumulative sum between the model equation and the normal distribution of the infection rate in 24 months is 119.3827. At the same time, the upper- and lower-95% quantiles are: 172.045 and 71.37293. The min-max accuracy of another predicted value is 0.916, while the mean-absolute-percentage-error is 0.0888. The mean value shows that the error is relatively low while the accuracy is high. The correlation accuracy is also high at 0.598. This means the actual values roughly have the same direction as the predicted ones. In other words, the actual value increase implies that the predicted value will also increase and vice-versa.

Finally, the values obtained in the mean absolute error (MAE) and root mean square error (RMSE) are 0.338 and 0.39. RMSE penalises those non-continued prediction values more heavily than MAE. However, RMSE is still used in many models since it is smoothly differentiable and is better for mathematical operations. At the same time, minimising the squared error will give us the mean value of a set of numbers. Similarly, the absolute error will give its median.

Conclusion

When a structural equation model is set alongside philosophy, a causal relation analysis

can be formed. Indeed, using concentration graphs, directed with a weight of the network analy- sis (see Fig. 3 and Fig. 4), which are both visualisations of the data, indicates a correlation be- tween variables or indirectly implies a causal relationship that su!ces to be verified through an experiment. Therefore, using Australia as a case study, it can be concluded that an increase in humidity implies an increase in carbon dioxide emissions or a strong causal of 0.15 (see Fig. 4). While there is also a weak causal outcome of 0.08 factor—increasing carbon dioxide emissions— it also means a slightly decrease in temperature (see Fig 4).

Finally, the results also show there is a strongly casual relation that a decrease in temperature will increase the number of humans infect- ed with the flu (-0.57) relative to the level of humidity (-0.18) and strong wind (0.1), as shown in Fig. 3. As a result, the possible causal relationships becomes:

humidity —> CO2 —> drop in temperature —> increase in human flu

Thus, the above outcome can be used as a comparison to the previous section, suggesting the following domino events:

emission of CO2 —> climate (cold + decrease in humidity + strong wind) —> human flu

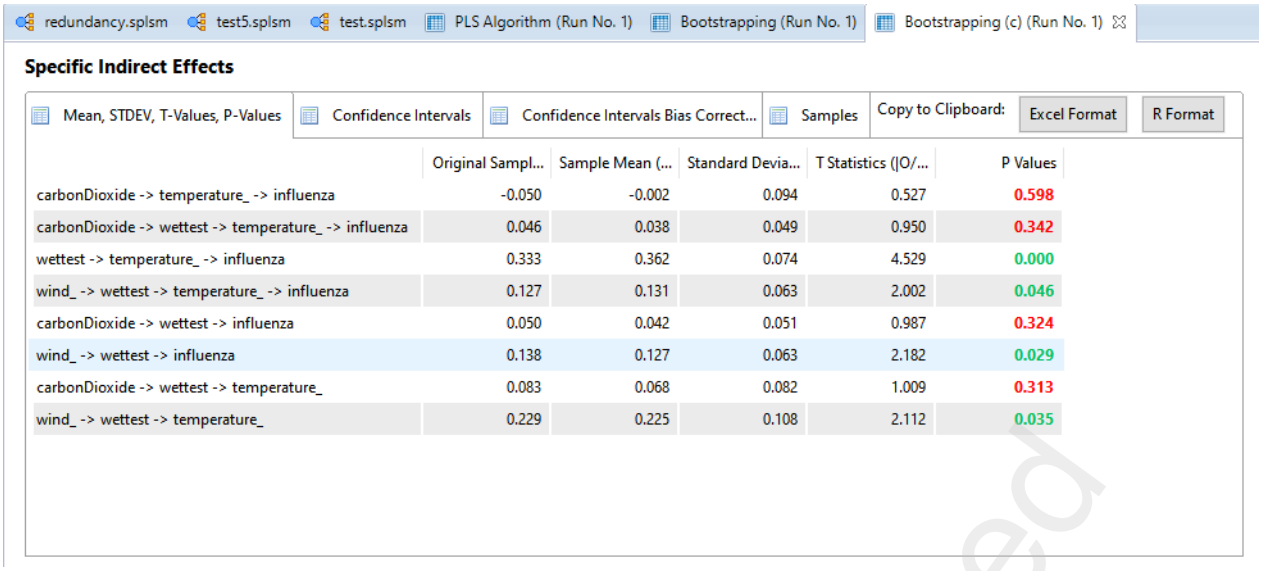
More specifically, based on causal analysis, humidity is shown to be a primary cause of human flu, instead of the proposed carbon dioxide emissions. The reason may be that Australia is surrounded by the ocean, and therefore its climate is heavily affected by humidity. On the contrary, suppose there were no greenhouse effects due to excessive carbon dioxide emissions, the network analysis results still show a rise in humidity levels. This is possibly because warmer temperatures cause more water to evaporate from the ocean, leading to higher humidity. Therefore, due to Australia's ocean surroundings, higher temperatures will automatically cause the country's humidity levels to rise. As heat energy is depleted due to evaporation, temperatures will decrease. Yet, the initial temperature rise encourages more vegetation to grow, thus increasing "surface roughness", creating more friction, and decreasing average wind speeds throughout the country. Therefore, it becomes possible to predict a wet and cool (low temperature) winter in Australia. Finally, this drop in overall temperature leads to increased flu cases in the human population. That said, the decrease in both temperature and windspeed implies an increase in carbon dioxide concentration. It is well known that climate change has been caused by excessive carbon dioxide emissions. This is the consequence of greenhouse effects, which contradicts the initial assumption that there is no such greenhouse effects. Hence, humidity (or in fact carbon dioxide and the greenhouse effect) might be the primary cause of human flu in Australia. The implementation of quantitative (social network) analysis in R merely explains the correlations between variables, but not their proposed cause and effects relationships. In fact, in the case of behavioural causal theory, the partial least squares (PLS) method can be applied for structural equation modelling to build and test it. To test this study's proposed causal relationship, the researcher will first design a series of corresponding experiments (in-

vestigating all complex factors associated with Australian winters) and test this relationship through a hypothesis between the causes (independent variables) and effects (dependent variables). A similar referenced experiment was performed by Foster in 2014 [4].During the experiment, Foster showed that flu transmission would happen in cases of high humidity with low temperatures. Moreover, low humidity implies 100% transmission rate. Foster concluded that the flu virus likes cold, dry weather. In the present article, one may apply the PLS analysis (that is the PLS algorithm, Bootstrapping together with mediation analysis etc) to test the data which is obtained from Australia government web-site and hence use the software SmartPLS to test those hypothesis (suggested casual relationships) as will be mentioned below.

4. https://sitn.hms.harvard.edu/flash/2014/the-reason-for-the-season-why-flu-strikes-in-winter

With those weather and health data obtained from Australia government web-site, this author import the data into software SmartPLS. Then one may get the deserved path model. Indeed the model is a reflective one or independent indicators are the cause of dependent latent variables. Furthermore, one can also perform mediation analysis and get the wanted specific indirect effects as

follow (which is obtained by Consistent PLS Bootstrapping) :



redundancy.splsm | test5.splsm | test.splsm | PLS Algorithm (Run No. 1) | Bootstrapping (Run No. 1) | Bootstrapping (c) (Run No. 1)

Specific Indirect Effects

Mean, STDEV, T-Values, P-Values | Confidence Intervals | Confidence Intervals Bias Correct... | Samples | Copy to Clipboard: | Excel Format | R Format

	Original Sampl...	Sample Mean (...	Standard Devia...	T Statistics (|O/...	P Values
carbonDioxide -> temperature_ -> influenza	-0.050	-0.002	0.094	0.527	0.598
carbonDioxide -> wettest -> temperature_ -> influenza	0.046	0.038	0.049	0.950	0.342
wettest -> temperature_ -> influenza	0.333	0.362	0.074	4.529	0.000
wind_ -> wettest -> temperature_ -> influenza	0.127	0.131	0.063	2.002	0.046
carbonDioxide -> wettest -> influenza	0.050	0.042	0.051	0.987	0.324
wind_ -> wettest -> influenza	0.138	0.127	0.063	2.182	0.029
carbonDioxide -> wettest -> temperature_	0.083	0.068	0.082	1.009	0.313
wind_ -> wettest -> temperature_	0.229	0.225	0.108	2.112	0.035

1. Wettest —> temperature —> influenza with P value equals to 0.00 or one rejects the null hypothesis that: The impact of wettest on the number of case of influenza infected is unaffected by temperature; or there is a causal relationship
2. Wind —> wettest —> temperature —> influenza with P value equals to 0.046 or one rejects the null hypothesis that: The impact of wind on the number of case of influenza infected is unaffected by wettest and temperature; or there is a causal relationship
3. Wind —> wettest —> temperature with P value equals to 0.035 or one rejects the null hypothesis that: The impact of wind on the temperature is unaffected by wettest; or there is a causal relationship
4. Wind —> wettest —> influenza with P value equals to 0.029 One rejects the null hypothesis that: The impact of wind on the number of case of influenza infected is unaffected by wettest.

The above simulated results imply that wind and wettest have the cause-effect relation to the number of case of influenza infected. These outcomes are indeed generally consistent with the network analysis one obtained previously and agreed with this author's proposed explanations. More importantly, the above outcomes directly provide evidence for the proof of the causal relationships in my

proposed sequence of domino events of the Butterfly Effects philosophy.

This author also remarks that the specific indirect effects that obtained from PLS Algorithm is:

Indirect Effects

Total Indirect Effects | Specific Indirect Effects

	Specific Indirect Effects
carbonDioxide -> temperature_ -> influenza	-0.050
carbonDioxide -> wettest -> temperature_ -> influenza	0.046
wettest -> temperature_ -> influenza	0.333
wind_ -> wettest -> temperature_ -> influenza	0.127
carbonDioxide -> wettest -> influenza	0.050
wind_ -> wettest -> influenza	0.138
carbonDioxide -> wettest -> temperature_	0.083
wind_ -> wettest -> temperature_	0.229

1. Carbon dioxide —> temperature —> influenza with p value equals to (-0.05)

One fails to reject the null hypothesis that: The impact of carbon dioxide on the number of case of influenza infected is unaffected by temperature;

2. Carbon dioxide —> wettest —> temperature —> influenza with p value equals to 0.046

One rejects the null hypothesis that: The impact of carbon dioxide on the number of case of influenza infected is unaffected by wettest and temperature; or there is a causal relationship.

3. Carbon dioxide —> wettest —> influenza with p value equals to 0.083

One accepts the null hypothesis that: The impact of carbon dioxide on the number of case of influenza infected is unaffected by wettest.

The causal relationships among wettest, temperature, wind and influenza may form a multiple mediation. This can be found from their multiple specific indirect effect. Indeed, mediation analysis may be needed. Alternatively, meta-analysis is another way for detecting the required multi-level (two stages) model. From the result of ordinary PLS Algorithm, it implies there is another causal

relation that: wettest and temperature together are the mediator between carbon dioxide and the number of case of influenza infected. However, for the result obtained by Consistent PLS Bootstrapping, it shows there is no such causal relationship between carbon dioxide and influenza infected.

The main reason for the differences obtained between PLS and the Consistent PLS Boot- strapping is that:

The traditional PLS algorithm "tends to overestimate the loadings in absolute value, and to underestimate multiple correlations between the latent variables while the advantage is its well calibrated — I.e. it will produce the true parameter values for the models that one discusses when applied to the population" (Dijkstra & Schermelleh-Engel, 2015: 586). Indeed, PLS-SEM has advantages over (Covariance-Based)CB-SEM is that:

1. Sample size is small while high valued structural path coefficients are needed;
2. Applications have little available theory;
3. Predictive accuracy is paramount;
4. Correct model specification cannot be ensured;

But the disadvantages are:

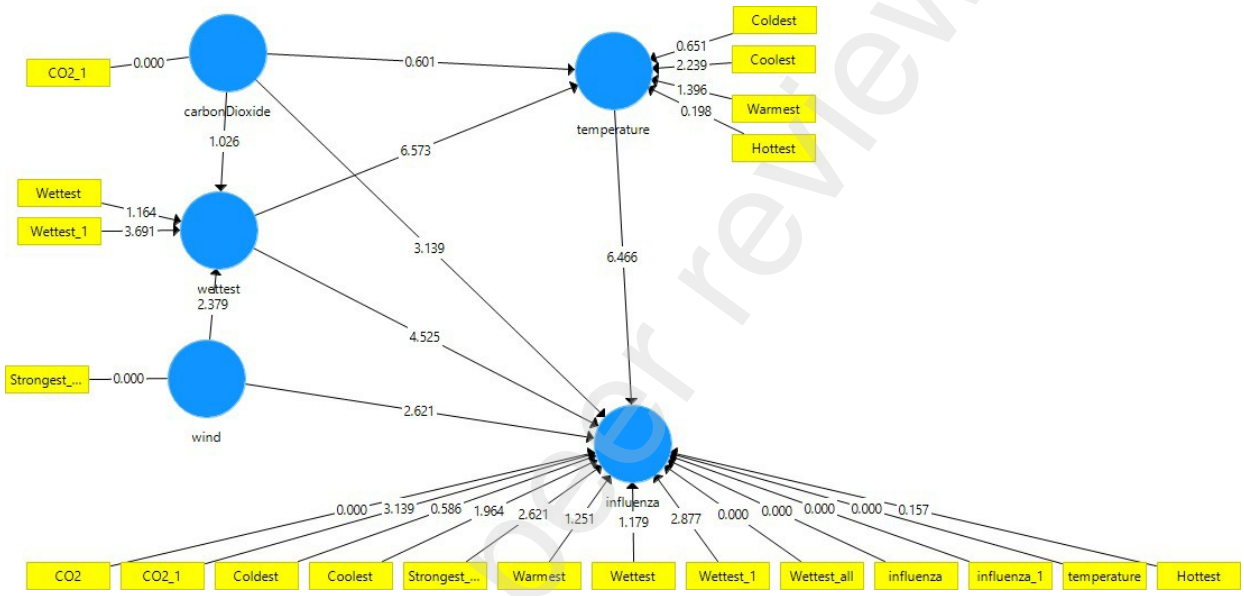
1. Problem of multicollinearity if not handled well;
2. Single headed arrow cannot model undirected correlation;
3. A lack of complete consistency in scores on latent variables may result based component estimation, loadings and path coefficients;
4. It may create large mean square errors in the estimation of path coefficient loading.

In the case of Bootstrapping, it has the drawback that when the samples do not have finite moments, small sample sizes, estimating extreme values from the distribution and estimating variance in survey sampling where the population size is N and a large sample n is taken. The reason for using it is that sometimes one can't rely on parametric assumptions and in some situations the bootstrap works

better than other nonparametric methods. It can be applied to a wide variety of problems including nonlinear regression, classification, confidence interval estimation, bias estimation, adjustment of p-values and time series analysis to name some applications.

Generally speaking, all of the above findings and results are consistent with this author's explanations previously — carbon dioxide is actually the real cause for the increase of cases in influenza infected if PLS model is accepted.

The formative measurement model finally obtained is shown as below:

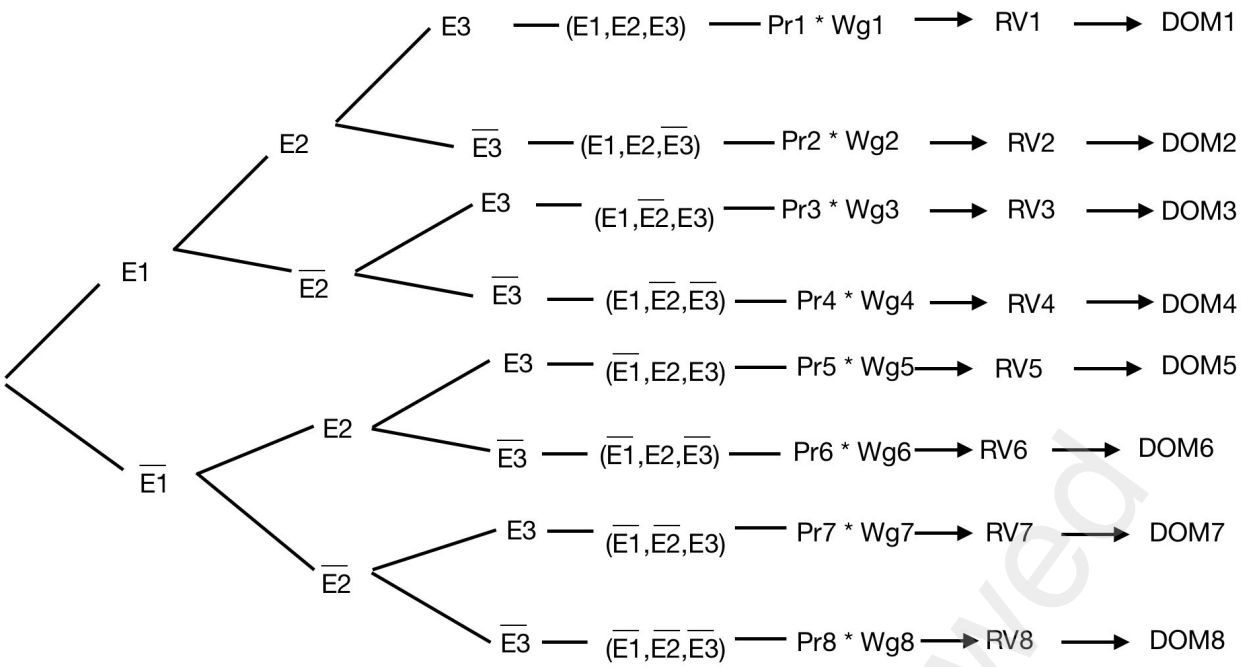


Indeed, both the (forwarding) Bayesian tree diagram and the (converse) tree diagram is a type of probability philosophy. While for the connecting immediate part — linear mapping is actually the application of linear algebra philosophy. Therefore, this author believes my proposed Butterfly Effects (as shown below in figure 6a, b, c) is indeed a type of predictive philosophy.

Below is the extended version of proposed Butterfly Effect (or a predictive) philosophy:

Figure 6a

It should be noted that E1, E2 and E3 are di"erent independent events where it is possible that the out-



come probability Pr1 = Pr2 = Pr3 = Pr4 = Pr5 = Pr6 = Pr7 = P8. However, the weight multiplied must be di"erent. This will create a one-to-one linear mapping to the corresponding domino e"ects individually. The situation likes human's deductive reasoning that an outcome is based on another argument and determines a cause-e"ect relationship.

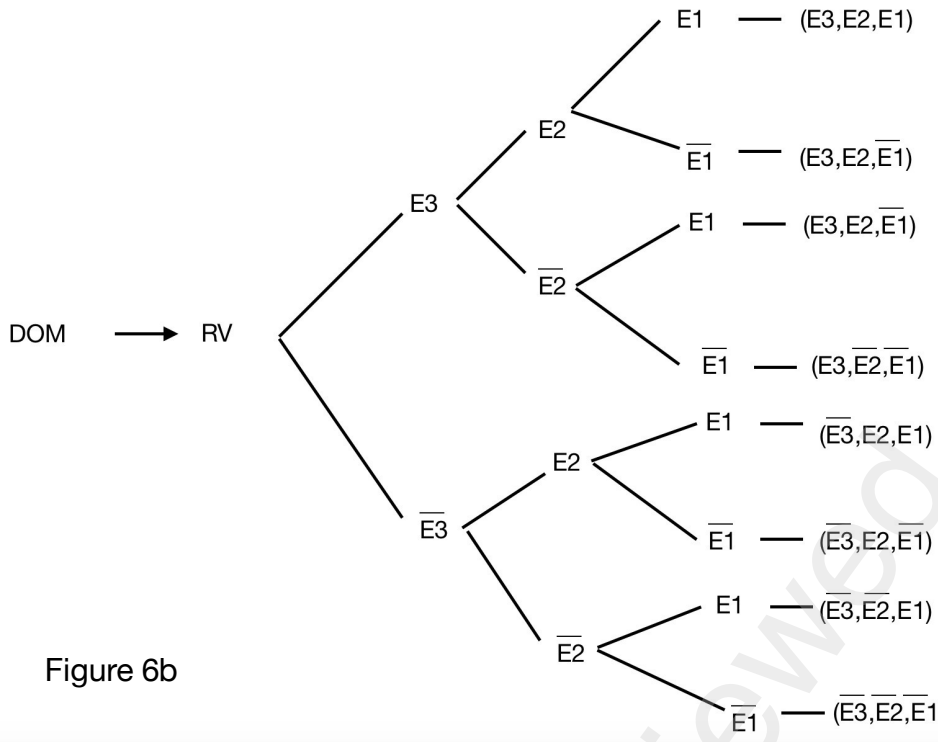


Figure 6b

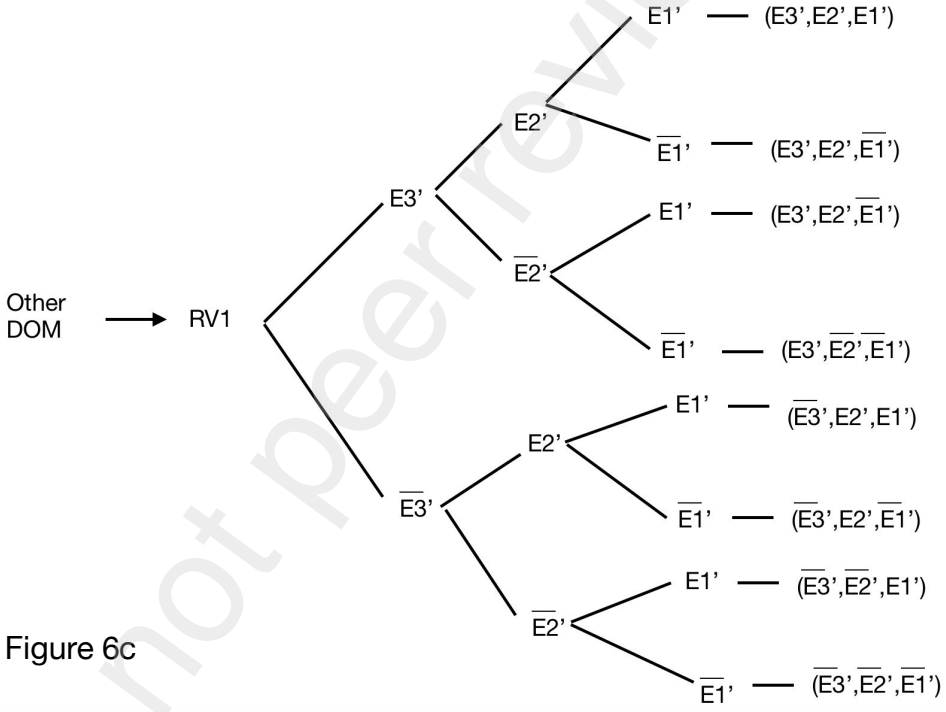


Figure 6c

It should be noted that the first domino event "Dom" corresponds to random variable RV and events E1, E2, E3. For the first main domino event, one may use those predicted values as calculated from a mathematical model of an investigation (data) to forecast the relevant probability for events (such as [E3, E2, E1] etc). While the other domino event "OtherDOM" corresponds to random variable RV1 with the relevant probability for other events (like [E3', E2', E1'] etc). The situation likes inductive reasoning that an inference is made with uncertainly; while the conclusion is likely but not guaranteed. Finally, th sequence of the domino events su!ce to the path analysis dependency calculation. This is used to find out the

order (or direction) individually among each of these events by applying Structural Equation Modelling.

In order to reduce the number of flu cases over winter, it is advised to re-design cities (such as those in Australia) so that it fits the wet and cool climate (with strong winds) during this period. It should be noted that this study is related to environmental epidemiology and the design of urban cities. In 2018, an article by Matt Hickman suggested the following ways to increase the temperature of cities during cold weather [5]:

1. Plant dense rows of trees like spruce, which act as effective wind blockers along popular walking trails and paths;
2. Other types of deciduous trees that can enable the bright winter sun to reach are also needed;
3. To achieve maximum sunlight exposure, the authority should orient buildings' adjacent outdoor space like patios and public plazas towards the North (in Australia);
4. Widely install push-button heaters at high-traffic bus stops etc;
5. Install barrier-free warming huts in public parks and along trails;
6. Improve cycling infrastructure so that one can increase wintertime bike commuting;
7. Design buildings with the ability to capture maximum sunlight, as it can remove most of the humidity during winter in a city.

[5] https://www.mnn.com/earth-matters/climate-weather/blogs/how-urban-design-makes-cold- weather-cities-less-bleak

Finally, whether vaccines can prevent an outbreak of human influenza is open to debate and

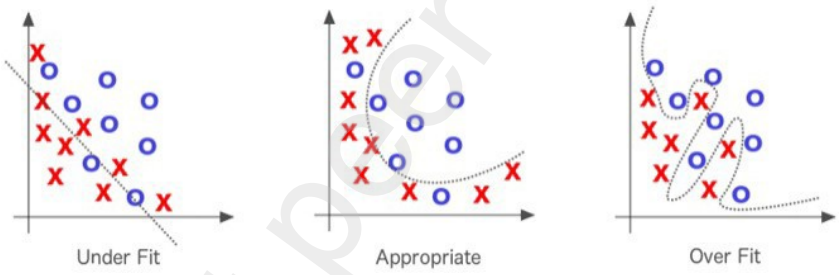
has no definite answer. This is because vaccines only provide around 60–70% protection. Their effective period is roughly six months; so if someone is vaccinated in the first half of the year, they may not be protected during the peak influenza period during winter. As such, the issue surrounding vaccination is very much a public health issue or even a philosophical one. In this author's opinion, people should decide for themselves whether to get vaccinated or not. Regardless of this point, there is still a risk of infection. Therefore, people must carefully consider the subject before making a decision.

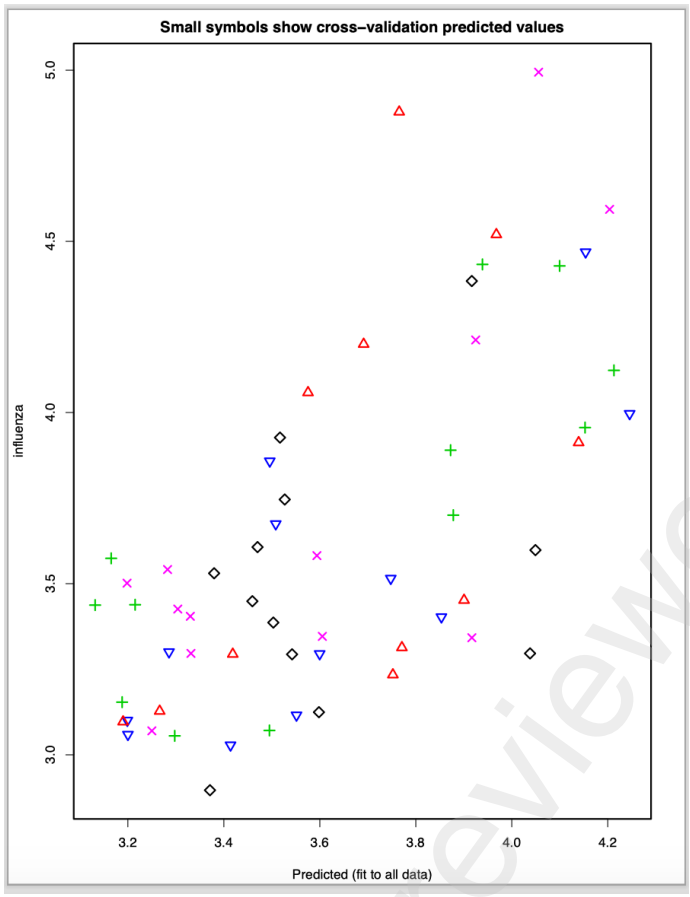
It is also important for a small-scale preliminary vaccination experiment to be performed before the aforementioned vaccination scheme is implemented worldwide.

Significant and Limitations

There are some limitations to this study. The first is one may apply similar techniques in looping in generat.path( ) for the second part of this author's programming code (the concept array of the array). Hence, a larger set of data will be generated and one may calculate the more precise value in estimating the predicted values and the accuracy of our model in prediction (forecasting). Last but not least, if one can use two models for prediction, then MAE is sure a better choice than RMSE.

The following diagram is the result of the 10-fold cross-validation in our prediction model as a generalized case while an example is provided in the next page. It should be noted that a predicted model is either under-fit, appropriate or over- fit according to the following lines of best fit:





Actually, by applying the supervised machine learning together with the linear regression, one may even build a more sophisticated model for our predictive (this is what author's forecasting does mean) one in human-weather interaction between influenza infected rate. It is somehow in such perspective very different from what the Bayes' theorem will predict the probability of people who take vaccine given that they were being infected by influenza (the final result of my Butterfly Effect Conjecture). Hence, the most important contribution of it (my proposed theory) is one may forecast the efficacy of the vaccine. Instead of traditional vaccine efficacy (which is a relative comparison rate) and effectiveness calculation, the (predicted) vaccine imperfection is computed by using the concept of conditional probability for the following years' about human being's rate of influenza infection. This obviously means the higher the (forecast) predicted probability (rate) of being infected even taken the vaccine (approaching to one), the lower efficacy it will be during the peak month of influenza disease.

Strictly speaking, one may redefine the traditional vaccine efficiency by well mixed with the confusion matrix of the influenza disease's false positive or negative ratio etc through the following case study:

By our usual definition of vaccine efficacy:

Suppose for a particular influenza vaccine, we have an overall efficacy of around 77.8%. The result is based on the doubly blinded and randomised controlled trial. The total tested participated persons are 244,190. The vaccinated group consisted of 122, 210 (with 240 infected) while the controlled or unvaccinated group had 121,980 tested persons (with 1060 infected).

Step IA: Baseline risk for the vaccine

ed group: (240 / 122210) or 0.196%Step IB: Baseline risk for the controlled group: (1060 / 121,980) or 0.86%Step II: The relative risk is: RR = 0.196 / 0.86 or approximate equals to 0.2279Step III: The traditional vaccine efficacy or VE is: (1 - RR)*100% = (1 - 0.2279)*100% or 77.21%

But how is if we well mixed with the false positive or negative etc of the confusion matrix under such definition of vaccine efficacy before the computation, then we may have the following:

Suppose before the VE computation, among the vaccinated population of 122,210 participated persons, we have the corresponding confusion matrix from the Bayes' theorem as below:

N = 122,210	**Predicted: Not infected**	**Predicted: Infected**	
Actual: Not infected	TN = 121,946	FP =12	121,958
Actual: Infected	FN =24	TP = 228	252
	121,970	240	

Thus, the accuracy of the infected test for the vaccinated population is:

(TP + TN) / (TN + FN + TP + FP) = (121,946 + 228) / (121,946 + 24 + 228 + 12)

= 122,174 / 122,210 = 0.999705 or 99.9705%

I.e. Among the 122,210 vaccinated and tested persons, there are 252 true infected persons

Thus the baseline risk for vaccinated group is: (252 / 122210) or 0.206%

Similarly, for the controlled or non-vaccinated group with population of 121,980 participated persons, we have the respective confusion matrix from the Bayes' theorem as shown in the following:

N = 121,980	**Predicted: Not infected**	**Predicted: Infected**	
Actual: Not infected	TN = 120881	FP = 106	120,987
Actual: Infected	FN = 39	TP = 954	993
	120920	1060	

TN is true negative, FP is false positive, FN is false negative, TP is true positive.

Thus, the accuracy of the infected test for the controlled or non-vaccinated

population is: (TP + TN) / (TN + FN + TP + FP) = (120,881 + 954) / (120881 + 39 + 954 + 106)

= 121,835 / 121980 = 0.9995737 or 99.95737%

i.e. Among the 121,980 controlled and tested persons, there are 993 true infected persons.

Thus, the baseline risk for the controlled or (non-vaccinated) group is (993 / 121,980) or 0.814%

Then the new and modified relative risk is: NMRR = (0.206 / 0.814) or approximate equals to 0.2531

The new and modified vaccine efficacy or NMVE is: (1 - RR)*100% = (1 - 0.2531)*100% = 74.69%

The difference between VE and NMVE is: -2.52% or the percent change is: (-2.52 / 77.21)*100% = -3.26% or

a gap of approximately decreasing 3.3 percent change from the original to the new and adjusted vaccine effica-

cy.

Here, this author defined the **new and adjusted vaccine efficacy (NMVE) by:**

$$\{1 - [(FN + TP)/N]_{\text{vaccinated group}} / [(FN + TP)/N]_{\text{controlled group}}\}*100\%$$

Indeed, this author remarks that in other words, there is an increasing 3.3 percent change from the new and adjusted to the original vaccine efficiency. These increasing and decreasing gap constitute a philosophy of percent change in the vaccine efficiency. Or in terms of an absolute value, there is a net gap of 3.3 percent change between the original and the new modified in the vaccine efficiency.

With the above confusion matrices, I may find the

Sensitivity of the vaccine: P(diagnosed sick| sick) = (121946 + 120881/121958 + 120987)

or approximately equals to 0.99.

Specificity of the vaccine: P(diagnosed not sick|not sick) = (954 + 228/993 + 252)

or approximately equals to 0.950.

To go a forward step, we humans may even update the traditional SIR model with the new and adjusted testing population (true or false & positive or negative) from the confusion matrices. **Hence the new and adjusted set of dependent variables S,I,R with respect to independent variable time t in SIR model is** (Smith et. al, 2004 & Kundu et. al, 2021 & Cohen, 2020)**:**

S = S(t) is the number of susceptible individuals,

I = I(t) is the number of infected individuals is updated with the following probabilities: $I = I(t)*\{[1-(FP_1/TP_1+FP_1)+(FN_1/TN_1+FN_1)]*100\%\}$,

$R = R(t)*\{[1-(FP_2/FP_2+TP_2)+(FN_2/FN_2+TN_2)]*100\%\}$ is the number of recovered individuals.

This author wants to remark that we may employ time series to predict future data. Then, we may

transfer these data through a software JASP function into the binary data and hence coded into a confusion matrix for the multiplication between those adjusted false positive or negative (probability) ratios. On the other hand, the confusion matrix may be transformed into the binary data and hence obtains the binary time series in order to get the corresponding future predicted data values. In addition, FP_1, TP_1, FN_1, TN_1, etc., are selected from the total testing population (true or false & positive or negative) of the Bayesian tree with the infected/uninflected sub-trees etc. The FP_2, TP_2,FN_2,TN_2 are chosen from the Bayesian sub-trees of the recovered/unrecovered population from the total number of truly infected patients.

In practice, the confusion matrix is actually a classification model of the machine learning. Hence, one may employ the last few years' data of being infected even taking vaccine to- gether with those healthy given vaccination for the next year's (forecast) prediction. The method in calculating these past years' data — TN, TP, FN, FP and the percent change etc can be found by us- ing Bayes' theory with respect to the population investigated in an ideal case and a direct sized population investigation for practice. Then one may use these values in the PSPP software for the future years' data prediction. Or one will ultimate perform this author's true expression of "forecast the prediction". This will save millions of people since scientists can explore the tendency of the vaccine's efficacy in the future years. Moreover, suppose there will be an extra expenditure of government's spending 500 dollars for not taking vaccine if getting infected and 100 dollars of not infected in medical treatments. Then one may calculate the expected value in the expenditure for the coming years' prediction. The requirements are one should know the total population of the outcome area together with those necessary conditional probability data. In other words, one may calculate the expected expenditure that our government may spend in the predicted influenza disease outbreak. This is good for the planning of our yearly financial budget. Another major application of my but-

terfly conjecture (theorem) besides vaccine efficacy prediction is in the earthquake. By using modelling (SEM), big data together with Bayes' theory etc as aforementioned in this paper's previous sections, scientists may give an early warning before an earthquake according to those phenomena that may occur in advance. The details are left to those civil engineers and geographical professionals. The discussion is out of the scope of this paper.

Remarks: 1. This author notes a Bayesian Trap may occur during the testing of a common disease. If you consider a yearly checkup (test) from a doctor's request in a particular disease, the probability that you get the disease is 0.1%. The test is with 99% accuracy (i.e., with 1% error). Now imagine that there is a small group of investigation with 1000 people. There may be only one person (0.1%) who will have the disease with the positive test. At the same time, other 10 people out of the 900 more healthy (or the error percentage) will also give a positive test. Finally, there will be a total of 11 people who will get a positive test result but with only one is the real patient. In re- ality, there is about 9% of people who are healthy given that their test are positive. The above situa- tion is called a Bayesian Trap.

2. One may use the R programming to construct a naive Bayesian classifier for any disease (e.g., influenza disease). The aim of it is to combine Bayes' model with decision rule like the hypothesis which is the most probable outcomes. Hence, one may establish the corresponding predictive philosophy about the disease.

3. Another application of my proposed Butterfly effects theory is in the field of reversing paralysis. One may consider the Bayesian tree as our human brain while the domino parts as the handicapped legs or hands. They are middled with the microchip processor for linking the brain signals and the paralysed human legs or hands. By placing a microchip processor in the human body which connects the brain signal and stimulate the paralysed legs or hands. Paralysed patients may have the chance of being recovered. The technology is now tested by the Swiss nation. However, the contro-

versy is that human may be turned into semi-machine-human and this might be immoral. Cer- tainly, if the damaged nervous can be reborn, then the problem seems to be solved completely. But such bio-technology needs time to develop. This author believes in the mean time, the suggested ma- chine-aided technology may be the best method for helping our paralysed person to be recovered.

4. One may suggest the following steps for solving the Bayesian Traps:

First of all, if one assume that the component of the claim is true, one should focus on it and identify those evidences one expect and love to find. This is done after one clearly articulates the causal mechanism.

Next, each item of evidence is assigned two probabilities, one for get the disease and one for Type I Error. Ideally, one is looking for evidence with high rate of getting the disease and low Type I Error.

Plug the probabilities for getting the disease and Type I Error into the Bayes Formula to determine the posterior confidence for each item of evidence.

Focus evidence gathering firstly on evidence with the highest posterior confidence values, for example with values of 0.85 and above. If, having searched for this evidence, it is not found, move to the next level down (0.7 – 0.85). Again, if insufficient evidence is found, move to the next level down (0.5 – 0.7). Continue this process until you find the evidence you need, and leave the rest! The point is that you don't need to gather all the evidence in your list, just the evidence that vali- dates your claim with the highest level of confidence.

5. There are several challenges over infectious disease epidemiology, surveillance and control. The first challenge is the sharing of data across early warning tools which support risk assessment and predictive models. The second challenge is to have a good legal frameworks for public health data sharing so that potential research can be unlocked without causing impact to citizens privacy. The third challenge is the strict regulation over the IT industry with regards to the manipulation of relevant data in everyday usage.

6. Indeed, one may extend the forecasting idea of both environment together with the predicting antigenic variants of H1N1 influenza virus based on a stacking model (Yin, at el., 2018). Hence,

one can find out the most feasible type of flu disease in the coming years.

7. After knowing the type of disease, the government may order the suitable vaccine for the next year, establish suitable warning tools and start the corresponding disease surveillance and control.

8. Indeed, the aforementioned predictive algorithm can also be applied in the protein-protein interaction prediction. This can prevent the accumulation of PPI which leads to several kinds of cancer if suitable drugs are developed.

9. One may employ the mixed methodology (qualitative ways) to determine those symptoms of liver cancer.

10. One may build the neural network for thinking through multi-step reasoning, iterative attention over abstracted, disentangle concepts that provides another way for the building of a predictive model instead of supervised learning and linear regression [6].

6. http://www.youtube.com/watch?v=-2JRiv3Mycs

Reference:

https://insights.careinternational.org.uk/development-blog/avoiding-the-data-trap-blog-3-an-ancient-monk-s-solution-for-confidence

https://www.analyticsindiamag.com/what-is-a-naive-bayes-classifier-and-what-significance-does-it-have-for-ml/

https://www.r-bloggers.com/naive-bayes-classification-in-r-part-2/

http://reliawiki.org/index.php/Simple_Linear_Regression_Analysis Linear regression analysis and prediction

https://www.quora.com/How-do-you-build-a-linear-regression-model-in-machine-learning
Creating model from linear regression analysis using Python

http://r-statistics.co/Linear-Regression.html using programming r

https://www.statmethods.net/advstats/factor.html Factor load

http://web.missouri.edu/~huangf/data/mvnotes/Using_R_for_path_analysis.html http://www.rpubs.com/tbihansk/302732

https://math.stackexchange.com/questions/1804362/interpretation-of-correlation-coefficient

Path analysis

https://beta.health.gov.au/resources/collections/childhood-immunisation-coverage-data-phn-and-sa3

https://www.immunisationcoalition.org.au/news-media/2019-influenza-statistics/

https://data.gov.au/search?q=observations-rainfall

https://data.gov.au/dataset/ds-dga-7e5598ef-7724-4e77-a7cc-6da741f72247/details?q=observations-rainfall

http://www9.health.gov.au/cda/source/pub_influ.cfm http://www.bom.gov.au/climate/current/month/nsw/archive/201702.sydney.shtml http://faculty.cas.usf.edu/mbrannick/regression/Pathan.html

Temperature and weather details http://www.bom.gov.au/climate/current/statement_archives.shtml https:// www.health.nsw.gov.au/Infectious/Influenza/Pages/2018-flu-reports.aspx

https://www.health.nsw.gov.au/Infectious/Influenza/Pages/reports.aspx

http://www.health.gov.au/internet/main/publishing.nsf/Content/cda-surveil-ozflu-flucurr.htm#current

http://www.bom.gov.au/state-of-the-climate/ http://www.bom.gov.au/climate/current/ statement_archives.shtml

https://www.environment.gov.au/climate-change/climate-science-data/greenhouse-gas- measurement/publications#quarterly

Australia https://www.co2.earth/monthly-co2 World

L.Marta, L.Michael, 2014, A predictive fitness model for influenza, Macmillan Publishing, doi:10.1038/nature 13087

Mu.E. Jianhong, MC.A. Bruce, Wu. Ximing and W.P. Michael, 2014.,

Climate Change and the Risk of Highly Pathogenic Avian Influenza Outbreaks in Birds; British Journal of Environment & Climate Change, 4(2): 166-185, 2014

Kostkova.P., (2018) Disease surveillance data sharing for public health: the next ethical frontiers., Life Science, Society and Policy, https://doi.org/10.1186/s40504-018-0078-x

Yin, R., Tran. H.V., Zhou, X., Zheng, J., and Kwoh, C.K., (2018) Predicting antigenic variants of H1N1 influenza virus based on epidemics and pandemics using a stacking model., PLOS ONE. https://doi.org/10.1371/journal.pone.0207777

David Smithand Lang Moore. The SIR Model for Spread of Disease - The Differential Equation Model, Convergence (December, 2004)

Bhattacharyya, R., Kundu, R., Bhaduri, R. Et al. Incorporating false negative tests in epidemiological models for SARS-CoV-2 transmission and reconciling with seropreva- lence estimates. Sci Rep 11, 9748 (2021). https://doi.org/10.1038/s41598-021-89127-1

Cohen, A.N. (2020) False Positive in PCR Tests for COVID-19 available from: icd10monitor.com/false-positives-in-pcr-tests-for-covid-19