

RESEARCH ARTICLE

Forecast the Prediction or Predict the Forecast — A Precise View of Investigating Butterfly Effects

Lam Kai Shun Carson

Department of Mathematics, University of Hong Kong, Pokfulam, Hong Kong

Received: 28-07-2021; Revised: 13-08-2021; Accepted: 20-08-2021

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 outbreak 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 person is infected. This predictive vaccine efficacy is based on Bayes' probability, rather than the traditional definition of efficacy or effectiveness. This is the definition of "forecasting a prediction." Conversely, if Bayes' probability tree is understood, and the same categorized 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.

Key words: Forecast, butterfly, influenza infection

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 minimize 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

Address for correspondence:

Lam Kai Shun Carson

E-mail: h9361977@connect.hku.hk

events, one could determine the probability of a catastrophic earthquake.

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 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.^[1-8]

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. 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 haruspices, 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 were

loosened, allowing for more innovative solutions and decisions to be achieved. This is known as the Age of Scrutinizing 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 Gottfried 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 modeling system where there is a change, and a way to deduce the predictions of such models" (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. 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 model is that there is no need to have a thorough understanding of the governing mechanism (Wong, 2009).^[3] 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 modeling approach uses multiple linear regressions. More specifically, the outcomes of the statistical data can be viewed as probability distributions, rather than unique values only.

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

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,^[4]

“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 gray scale. They are not black and white and that you have levels of confidence in your beliefs about how the world works that are <100% but >0% and even more importantly as you go through the world and encounter new ideas and new evidence, that

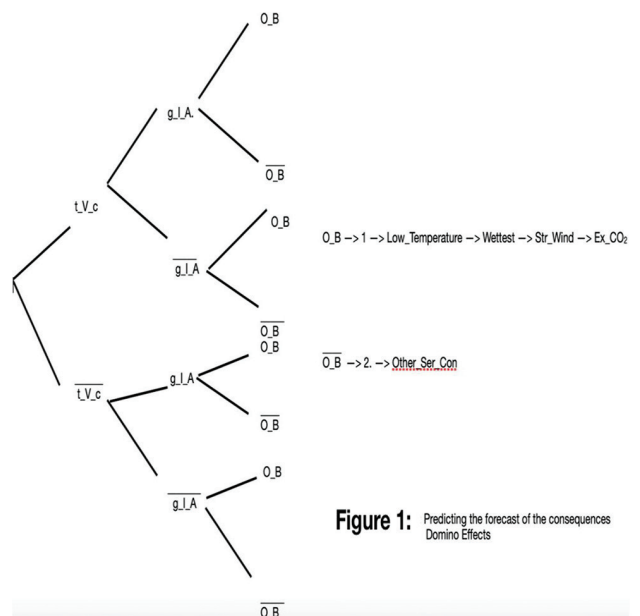
level of confidence fluctuates, as you encounter evidence for and against your beliefs.” That is the value of Bayes’ theorem and the proposed butterfly effect theory.

A More Generalized View of Proposed HKLam Theory (Or the Former 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 according to their properties in the following ways:

$$\begin{aligned}
 [A_1] &\rightarrow \dots \rightarrow [(A_1|A_1A_1\dots A_1), (A_1|A_2A_1\dots A_1), \dots, (A_1|A_nA_n\dots A_n)] \rightarrow RV_1 \rightarrow DE_1 \\
 [A_2] &\rightarrow \dots \rightarrow [(A_2|A_1A_1\dots A_1), (A_2|A_2A_1\dots A_1), \dots, (A_2|A_nA_n\dots A_n)] \rightarrow RV_2 \rightarrow DE_2 \\
 &\vdots \\
 [A_i] &\rightarrow \dots \rightarrow [(A_i|A_1A_1\dots A_1), (A_i|A_2A_1\dots A_1), \dots, (A_i|A_nA_n\dots A_n)] \rightarrow RV_i \rightarrow DE_i \\
 &\vdots \\
 [A_n] &\rightarrow \dots \rightarrow [(A_n|A_1A_1\dots A_1), (A_n|A_2A_1\dots A_1), \dots, (A_n|A_nA_n\dots A_n)] \rightarrow RV_n \rightarrow DE_n
 \end{aligned}$$

More specifically, each outcome event must be grouped with similar properties together with a suitable random variable (RV). Each of these variables is 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 generalized probability tree diagram, RV, and the consequence domino effects become:



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

Only if Part of the Proposed HKLam 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 categorized 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, RV and probability tree diagram will be drawn in the coming page as shown below:

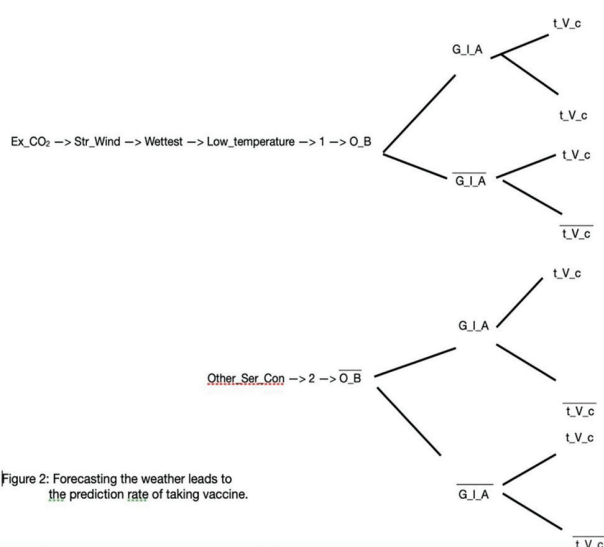


Figure 2: Forecasting the weather leads to the prediction rate of taking vaccine.

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°C) 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 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) determine 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).^[5]

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 to 2018, this study considered nine climate-related factors: Rate of carbon dioxide emission, number of infected, hottest day, coldest night, wettest

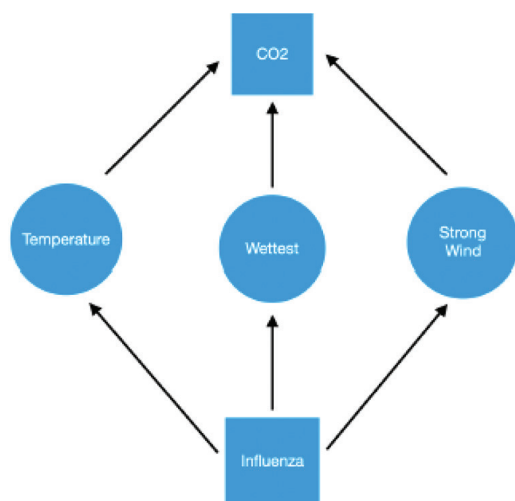
day, warmest overall, coolest overall, wettest overall, and strongest wind. A prototyped model describing the relationship between the weather and the infected will then be created (mainly using path coefficients). This can be done through linear regression using software such as PSPP.

The following data were obtained:

The predicted influenza infected rate over the coming 12 months is 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

Left figure: The suggested prototyped structural equation model



The SEM for influenza and temperature:
 $Influenza = -0.09 \text{ temp} + 5.22$

The SEM for influenza and SW_log, Str_wind:
 $Influenza = -0.02 \text{ Str_wind} + 6.56 \text{ SW_log} - 8.23$

The SEM for influenza and SW_log, Str_wind:
 $Influenza = 0.00 \text{ Wettest_1} + 4.02$

Table: Summary of the Influenza Infection Rate to Weather Model

Model	Constant	Variable	SEM
1	5.22	Temperature with value = -0.09	$Influenza = -0.09 \text{ temp} + 5.22$
2	-8.23	SW_log with value = 6.56 Str_wind with value = -0.02	$Influenza = -0.02 \text{ Str_wind} + 6.56 \text{ SW_log} - 8.23$
3	4.02	Wettest_1 = 0.00	$Influenza = 0.00 \text{ Wettest_1} + 4.02$

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 Modeling) 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 standardized. The aim is 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:

This author notes that the negative values of the path — standardized 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 modeling 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 series. To overcome randomness, a logarithmic function will be used to standardized the data. By calculating 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 trialing 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, together with the accuracy of the model, through K-Fold Cross validation. The followings 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$influenza, predicted=infectedPred))
correlation_accuracy <- cor(actuals_preds)
head(actuals_preds)
min_max_accuracy <- mean(apply(actuals_preds, 1, min) / apply(actuals_preds, 1, max))
mape <- mean(abs((actuals_preds$predicted - 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)
```

N.B. Indeed, we may replace the above `lmMod <- lm(influenza ~ temperature +...)` by `lmMod <- lm(influenza ~., data =trainingData)` which could give out a more accurate model and less error for all of the related variables in the given training data set. Rather than predicting the effect of weathering to the number of cases in influenza or even the next species of influenza through the banking database, people may apply the model in constructing an environmental friendly business building that the users can save the total energy consumption in air conditioning. 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 5 years, where these 2 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

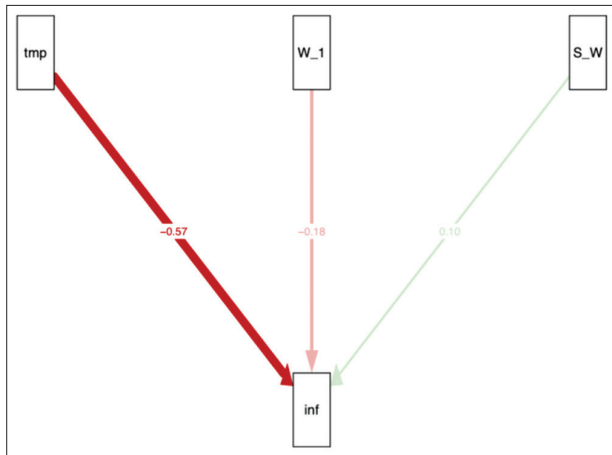


Figure 1: Network analysis for the causal relations among (temperature, wettest, and strong wind) and influenza

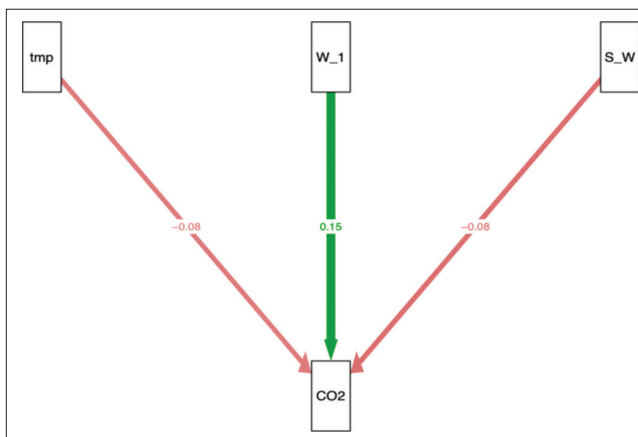


Figure 2: Network analysis for the causal relations among (temperature, wettest, and strong wind) and carbon dioxide emission

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 penalizes 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, minimizing 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 analysis [Figures 1 and 2], which are both visualizations of the data, indicates a correlation between variables or indirectly implies a causal relationship that succeeds 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 [Figure 2]. While there is also a weak causal outcome of 0.08 factor — increasing carbon dioxide emissions — it also means a slightly decrease in temperature [Figure 2]. Finally, the results also show there is a strongly casual relation that a decrease in temperature will increase the number of humans infected with the flu (-0.57) relative to the level of humidity (-0.18) and strong wind (0.1), as shown in Figure 1. As a result, the possible causal relationships become: Humidity \rightarrow CO₂ \rightarrow drop in temperature \rightarrow 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 \rightarrow climate (cold + decrease in humidity + strong wind) \rightarrow 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 wind speed 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 are 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 behavioral causal theory, the partial least squares (PLS) method can be applied for structural equation modeling to build and test it. To test this study's proposed causal relationship; the researcher will first design a series of corresponding experiments (investigating 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.^[6,9] 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 (i.e., the PLS algorithm, bootstrapping together with mediation analysis) to test the data which are obtained from Australia government web-site and hence use the software Smart PLS to test those hypothesis (suggested casual relationships) as will be mentioned below.

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):

	Original Sample...	Sample Mean (...)	Standard Devia...	T-Statistics (QO...	P-Values
carbonDioxide -> temperature_ -> influenza	-0.050	-0.002	0.094	0.527	0.598
carbonDioxide -> wetttest -> temperature_ -> influenza	0.046	0.038	0.049	0.950	0.342
wetttest -> temperature_ -> influenza	0.333	0.362	0.074	4.529	0.000
wind_ -> wetttest -> temperature_ -> influenza	0.127	0.131	0.063	2.002	0.046
carbonDioxide -> wetttest -> influenza	0.050	0.042	0.051	0.887	0.374
wind_ -> wetttest -> influenza	0.138	0.127	0.063	2.182	0.029
carbonDioxide -> wetttest -> temperature_	0.083	0.068	0.082	1.009	0.313
wind_ -> wetttest -> temperature_	0.229	0.225	0.108	2.112	0.035

1. Wetttest → temperature → influenza with $P = 0.00$ or one rejects the null hypothesis that: The impact of wetttest on the number of case of influenza infected is unaffected by temperature; or there is a causal relationship
2. Wind → wetttest → temperature → influenza with $P = 0.046$ or one rejects the null hypothesis that: The impact of wind on the number of case of influenza infected is unaffected by wetttest and temperature; or there is a causal relationship
3. Wind → wetttest → temperature with $P = 0.035$ or one rejects the null hypothesis that: The impact of wind on the temperature is unaffected by wetttest; or there is a causal relationship
4. Wind → wetttest → influenza with $P = 0.029$ One rejects the null hypothesis that: The impact of wind on the number of case of influenza infected is unaffected by wetttest.

The above simulated results imply that wind and wetttest 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:

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

1. Carbon dioxide → temperature → influenza with ($P = -0.05$) 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 $P = 0.046$ 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 $P = 0.083$

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 effects. 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 Boot- strapping, 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 and Schermelleh-Engel, 2015: 586). Indeed, PLS-SEM has advantages over (Covariance-Based) CB-SEM is that:

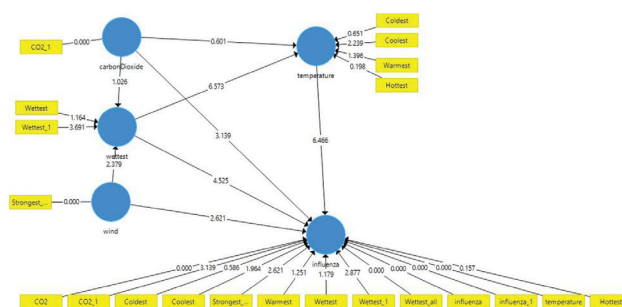
1. Sample size is small while high valued structural path coefficients are needed;
2. Applications have little available theory;
3. Predictive accuracy is paramount;
4. Correct model specification cannot be ensured; But the disadvantages are:
 - a. Problem of multicollinearity if not handled well;

- b. Single headed arrow cannot model undirected correlation;
- c. A lack of complete consistency in scores on latent variables may result based component estimation, loadings, and path coefficients;
- d. 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 cannot 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.

In general 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:



a one-to-one linear mapping to the corresponding domino effects individually. The situation likes human's deductive reasoning that an outcome is based on another argument and determines a cause-effect relationship. (b): The reverse of the HKLam Theory — Domino consequence to the Bayesian tree. (c) The reverse of the HKLam Theory — Other Domino consequence to the Bayesian tree

It should be noted that the first domino event "Dom" corresponds to RV and events E1, E2, and 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]). While the other domino event "Other DOM" corresponds to RV1 with the relevant probability for other events (e.g., [E3', E2', E1']). The situation likes inductive reasoning that an inference is made with uncertainty; while the conclusion is likely but not guaranteed. Finally, the sequence of the domino events suffice 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 Modeling.

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:^[7]

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.

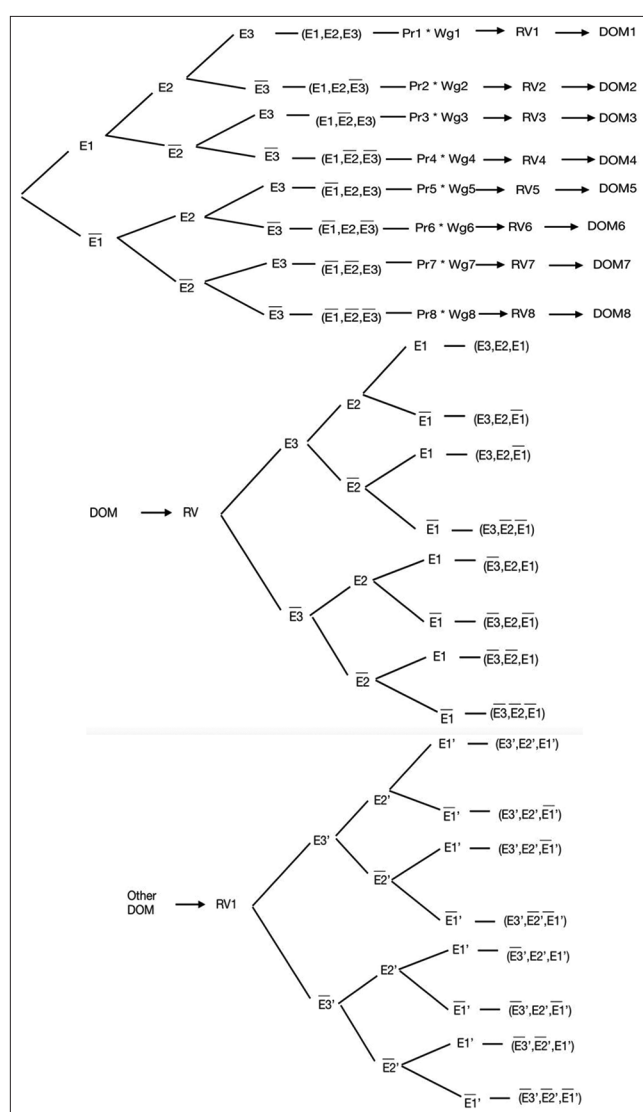


Figure 3: (a) It should be noted that E1, E2, and E3 are different independent events where it is possible that the outcome probability $Pr_1 = Pr_2 = Pr_3 = Pr_4 = Pr_5 = Pr_6 = Pr_7 = Pr_8$. However, the weight multiplied must be different. This will create a one-to-one linear mapping to the corresponding domino effects individually. The situation likes human's deductive reasoning that an outcome is based on another argument and determines a cause-effect relationship. (b): The reverse of the HKLam Theory — Domino consequence to the Bayesian tree. (c) The reverse of the HKLam Theory — Other Domino consequence to the Bayesian tree

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 6 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

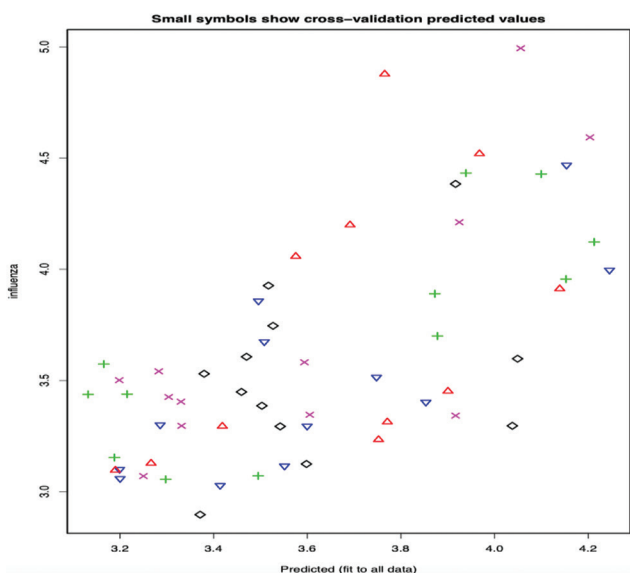
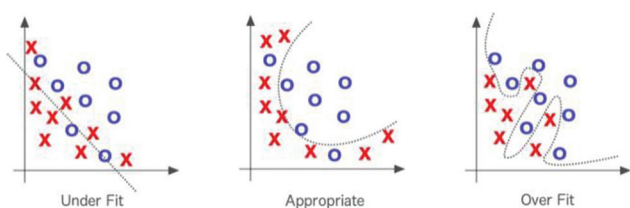
get vaccinated or not. Regardless of this point, there is still a risk of infection. Therefore, people must carefully consider the subject before making a decision.

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 generate 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 ten-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).^[8] 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 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 model efficacy of the fitting to the data it will be during the peak month of influenza disease.

REFERENCE

1. Krammer F, Smith GJ, Fouchier RA, Peiris M, Kedzierska K, Doherty PC, et al. "Influenza". *Nat Rev Dis Prim* 2018;4:3.
2. A Brief History of Forecasting, Foresight R, Trends and Forecasting Strategies. Available from: <http://foresightr.com/2016/05/06/a-brief-history-of-forecasting>.
3. Wong, H.K., Wong, R.T., Seroyer, C. (2009), *The First Days of School: How to be an Effective Teacher*. California: Harry K Wong Publication.
4. *Predicting the Future with Bayes Theorem*. Farnam Street Media Inc.; 2021. Available from: <https://fs.blog/2018/09/bayes-theorem>.
5. Łuksza, M., Lässig, M. (2014), A predictive fitness model for influenza. *Nature*, 507(7490), 57-61.
6. *The Reason for the Season: Why Flu Strikes in Winter*, Blog, December 1, 2014. Available from: <https://sitn.hms.harvard.edu/flash/2014/the-reason-for-the-season-why-flu-strikes-in-winter>.
7. Available from: <https://www.mnn.com/earth-matters/climate-weather/blogs/how-urban-design-makes-cold-weather-cities-less-bleak>.
8. Lam KS. Big data investigation into the causes and treatment of caries in kindergarteners. *Schl J Phys Math Stat* 2020;7:10.
9. Lam KS. Does Butterfly Charging Friendly to our Environment? An Experimental Research; 2021. Available from: https://www.papers.ssrn.com/sol3/papers.cfm?abstract_id=3886349. [Last accessed on 2021 Jul 14].