

EVALUATION OF THE WEATHER-INFLUENZA PATTERN WITH A REGRESSION MODEL APPROXIMATION TO CAUSALITY

Lam Kai Shun

M.Sc.(I.T. Edu), M.Sc.(Eng), B.Sc.

ABSTRACT: *Once at a time, this author starts a research in the relations between Australia's number of case in influenza and weather. The outcome has been hypothesised with a structural equation model (SEM). In this paper, the author tries to evaluate the model. It is true that one can apply the evaluation to both of the formative measurement model and structural model through certain suitable procedures. At the same time, this author approximates the model by linear regression method. The result is one can apply the regression to the Hayes' Process model and find out the wanted model with mediation and moderation effects. In addition, one can also use the Granger Causality Test to examine all of the hypothesised causal relationships between those independent variables such as temperature, wettest_1, concentration of carbon dioxide, strongest wind and coolest and the number case of influenza infected etc. The final outcome is that Hayes' model 91 is the best mediated one with carbon dioxide as the moderated factor. This author will also explain in details why we have the above prescribed Hayes' model 91 as the proposed regression model approximation to causality from the SPSS data analysis.*

KEYWORDS: weather-influenza, pattern, regression, model, approximation, causality,

INTRODUCTION

After this author's previous structural model in describing the relationship between the number of case in influenza and weather (Australia), one needs to evaluate it. Through the evaluation, one can modify the model and the most important thing is one can evaluate this author's HKLam Theory. In the following sections, this author will depict a brief review in how one should evaluate the formative measurement model, structural model together with the assessment of the causal relations from regression approximation. It is hope that all of the above evaluation processes, one can verify the truthiness of my proposed HKLam Theory. While the main results focus on the models selection (Hayes' model analysis) and their corresponding equations. These events explain why this author chooses Hayes' model 91 as the wanted mediation model.

Theoretically Background

Evaluation of the Formative Measurement Model

First of all, one is required to know what convergent validity is. It is indeed a measure that correlates with other measures within the same construct under different indicator (such as formative Vs reflective) in the case of a formative measurement model. This is known as redundancy analysis (Chin,1998). Or the formative measured construct acts as an exogenous latent variable that predicts an endogenous latent variable through some reflective indicators. This gives rise to the value 0.8 or higher (with 0.7 as minimum) for the path between Y (formative) and Y (reflective). Next, when there are high correlation values occurred between formative indicators, this is known as collinearity. If there are more than two involved indicators, then this is referred as multicollinearity. In order to access the level of collinearity, researchers are required to compute the value of tolerance

(TOL). This can be done through two steps:

1. Regress the first formative indicator with all same block remaining indicators, calculate the proportion of variance of the first indicator that associated with others (R^2);
2. TOL can then be calculated from the formula using $(1-R^2)$.

Indeed, the measure of collinearity is the variance inflation factor (VIF) which is just the reciprocal of the tolerance. In the case of PLS-SEM, Hair et al in 2011 told us that there may be a potential linear problem when the tolerance value is lower than 0.2 and with a VIF value higher than 5.

Finally, one should examine the indicators' outer weight and outer loading. When both of them are not significant, then the indicator is needed to be deleted.

Step 1: Formative construct's convergent validity
Vs (its correlation with)
alternative measure of the construct
where the resulting value should be higher than 0.8.



Step 2: Check the collinearity of each indicators by using their tolerance (VIF)'s values (should be higher than 0.2 but smaller than 5). Otherwise,

1. Eliminating the indicators;
2. Merging the indicators into a single index;
3. Creating some higher order constructs.



Step 3: Examine each indicator's outer weight and outer loading, using bootstrapping for the assess of their significance. The results diverse into three cases: (i) indicator's weight is significant, retain the indicator;

(ii) indicator's weight is not significant but with high item loading (> 0.5), generally retains the indicator;

(iii) both the outer weight and the outer loading are not significant, no need to retain the indicator and delete it from the model.

Figure 1: A summary of procedures to evaluate the formative measurement indicators

Evaluation of the Structural Model

Firstly, one needs to check the collinearity of those indicators. The procedure is the same as those mentioned in the section “Evaluation of the Formative Measurement Model”. Secondly, one is required to evaluate the structural model path coefficients. Actually, when the path coefficients are more closely to +1, this represents a stronger positive relationship with statistically significant. While for the path values closer to zero, it means weaker relationship. For the bootstrap standard error, when an empirical t value is larger than the critical value, then the t-value is significant with a certain amount of error probability. This may refer to most researchers’ usage of p values to assess significant levels. Practically, a p-value means the probability of getting a t-value when one observes conditionally from the supported null-hypothesis. Thirdly, another value for us to evaluate the model is the coefficient of determination (R^2 value). It is used to measure the predictive power of the model and is just the squared correlation between the actual and predictive values in endogenous construct. Indeed, R^2 ranges from 0 to 1 where higher level of its value implies a more accurate prediction. In scholarly marketing research, 0.75, 0.5 and 0.25 represent substantial, moderate or weak predictive power (Hair et al., 2011; Henseler et al., 2009). However, it is dangerous to select a model purely based on the R^2 . With the multiple regression, one may apply the adjusted coefficient of determination (R^2_{adj}). It avoids those bias in complex model. Fourthly, one can test how the endogenous constructs may be impacted by the omitted constructs. This is referred to the f^2 — the effect size. Technically, one can calculate the change in R^2 from the estimation of the Path Least Square model twice. While f^2 with values 0.02, 0.15 and 0.35 indicates small, medium and large effect (Cohen, 1988). For those values smaller than 0.02, this means there is no effect of the exogenous latent variable. Fifthly, the Stone-Geisser’s Q^2 value is a measure of the model’s out-of-sample predictive power i.e., the predictive relevance. In other words, Q^2 tells us how well the path model can predict the originally observed values. Actually, Q^2 value uses the blindfolding procedure and performs sampling for the omission of every d-th data point. Hence Q^2 computes those parameters for the remaining data points (Chin, 1998; Henseler et al., 2009; Tenenhaus et al., 2005). In fact, blindfolding is an iterative model re-estimation. Finally, one may compare the relative impact of predictive relevance through the measure to the q^2 effect size.

Assessment of the Causal Relationships

Theoretically, this author’s suggested HKLam Net-Seizing Theory can be expressed mathematically in the following ways:

3. The Bayesian Probability part: For all conditional probabilities to events, they can be expressed in terms of the corresponding Bayesian trees. While these trees can be expressed in terms of matrices.

4. The Linear Mapping part: One can evaluate the linear mapping through the selection of a suitable linear transformation. Through the transformation, one can map it to the proposed casual relationship. This author notes that in order to verify a linear transformation, the transformation should have the following properties:

T: $U \longrightarrow V$

I) $T(U + V) = T(U) + T(V)$ and

II) $T(cU) = cT(U)$

Abstractly, for a collection of all linear maps, $T: V \longrightarrow V$, denoted by $\text{End}(V)$ is a (non-commutative) ring, where addition is a point-wise addition:

$(T_1 + T_2)(V) \longmapsto T_1(V) + T_2(V)$

And the respective multiplication is the composition of:

$(T_1 \cdot T_2) \longmapsto T_1(T_2(V))$

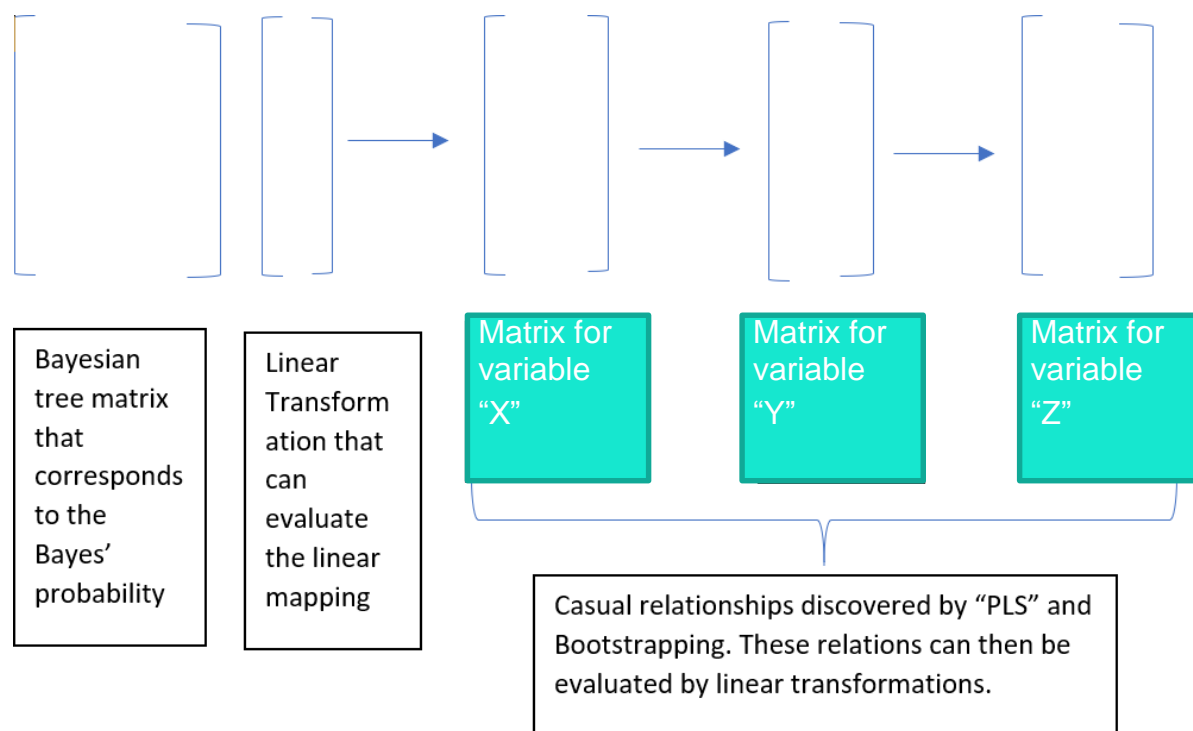
Now choose some particular $T \in \text{END}(V)$. By the universal property of the polynomial ring $k[x]$, one can define a ring homomorphism $k[x] \rightarrow \text{END}(V)$ by simply declaring that x should go to T . The result is the evaluation homomorphism $e_v T: k[x] \rightarrow \text{END}(V)$, $e_v T(p) = p(T)$

To be precise, one may have the following:

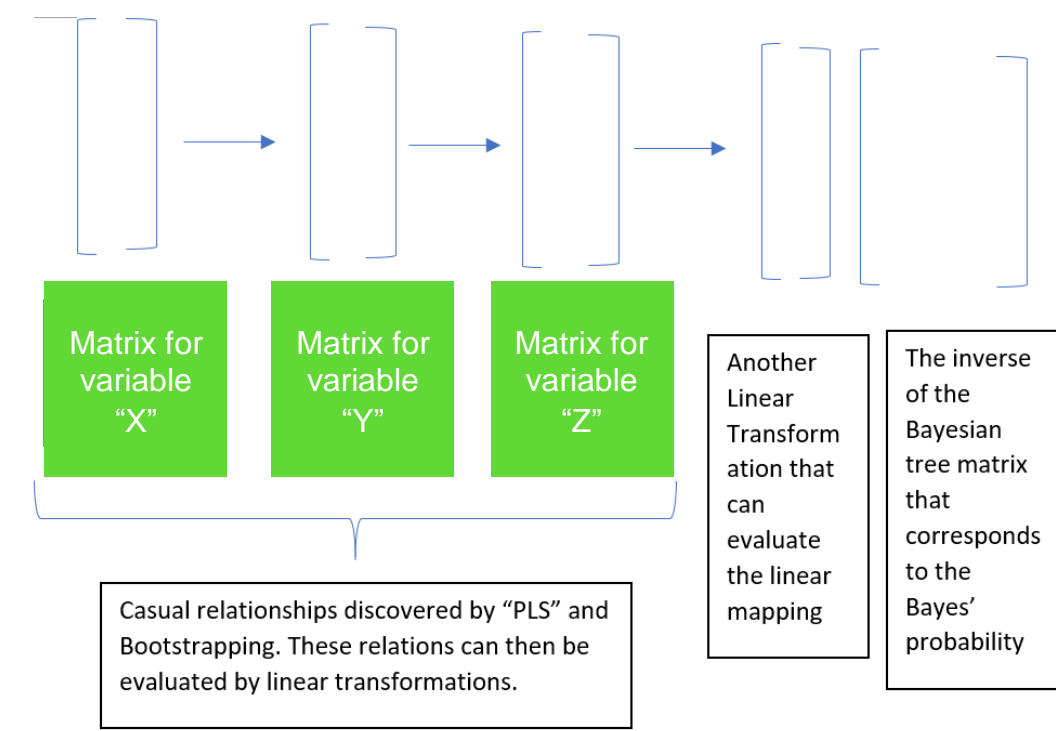
If V is a k -vector space, and $T \in \text{End}(V)$ is some linear map $T: V \rightarrow V$, then the polynomial:

$$p = a_0 + a_1x + \dots + a_nx^n \in k[x]$$

Evaluated at T is just: $a_0 + a_1T + \dots + a_nT^n \in \text{END}(V)$ where T^k is T composed with itself k times, and (aT) is the element defined by $(aT)(v) = a*(TV)$



Actually, partial least square is only used to find the relations between two matrices (X and Y). That is PLS is a latent variable approach which one can model the covariance structures of two spaces. In practice, a collection of data for different variables (like this author's data file — Australia_weather_influenza.xlsx) can be expressed in form of a matrix. Then one may perform Bootstrapping in statistics with the use of software SmartPLS. The purpose is to find out those causal relationship together with the expression of these relations in form of the required matrices.



This author notes that for the domino effects (or the later part of my proposed philosophy), it suffices to find out those causal relations. Theoretically, suppose there are variable matrices **X**, **Y** and **Z**, they can be expressed as follows:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F}$$

$$\mathbf{Z} = \mathbf{VR}^T + \mathbf{G}$$

Where **X** is an $n \times m$ matrix of predictors, **Y** is an $n \times p$ matrix of the corresponding responses (to **X**). **T** and **U** are $n \times 1$ matrices that are respectively, projections of **X** and projection of **Y**. **P** and **Q** are respectively, $m \times 1$ and $p \times 1$ orthogonal loading matrices; and matrices **E** and **F** are the error terms. Similarly, one may apply the same decomposition method to **Y** and **Z**. The aim is to maximise the covariance between **T** and **U** together with **V**.

Next, one may try to estimate the factor and loading matrices **T**, **U**, **V** and **P**, **Q**, **R**. One may then construct the linear regression between **X** and **Y**, **Y** and **Z** as

$$\mathbf{Y} = \mathbf{XB} + \mathbf{B}_0 \text{ and } \mathbf{Z} = \mathbf{YD} + \mathbf{D}_0.$$

The above way is known as partial least squares method for column vector \mathbf{Y} and \mathbf{Z} or matrices \mathbf{Y} and \mathbf{Z} . Actually, for a series of domino effect, one will have:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{B}_0$$

$$\mathbf{Z} = \mathbf{YD} + \mathbf{D}_0 \text{ or } \mathbf{Z} = \{\mathbf{XB} + \mathbf{B}_0\} \mathbf{D} + \mathbf{D}_0$$

This means one can always express the series of domino effect in a sequence of recursive approximated manner or a partial least square regression.

To be precise, the Bayesian Matrix, say $[\mathbf{M}]$, can be expressed by the regression as:

$$[\mathbf{M}] [\mathbf{LT}] = \mathbf{X} + (\mathbf{XB} + \mathbf{B}_0) + \{(\mathbf{XB} + \mathbf{B}_0)\mathbf{D} + \mathbf{D}_0\} \quad \text{Eq (1)}$$

where $[\mathbf{LT}]$ is the associated linear transformation; while the converse is also true:

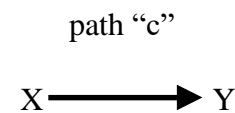
$$\mathbf{X} + (\mathbf{XB} + \mathbf{B}_0) + \{(\mathbf{XB} + \mathbf{B}_0)\mathbf{D} + \mathbf{D}_0\} = [\mathbf{M}] [\mathbf{LT}]$$

Hence, from the above mathematical expression, the causal relationships that found from my proposed Net-Seizing Theory, can be assessed by Baron and Kenny regression method (1986) — Testing for Mediation. The steps are listed as below¹:

Step I: Conduct a simple regression analysis with \mathbf{X} predicting \mathbf{Y} in order to test for path

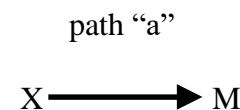
“c” alone. Or, one may have:

$$\mathbf{Y} = \mathbf{B}_0 + \mathbf{B}_1\mathbf{X} + \mathbf{e}_0$$



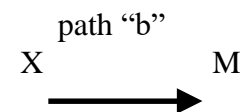
Step II: Conduct a simple regression analysis with \mathbf{X} predicting \mathbf{M} to test for path “a”. Or one may have:

$$\mathbf{M} = \mathbf{B}_2 + \mathbf{B}_3\mathbf{X} + \mathbf{e}_1$$



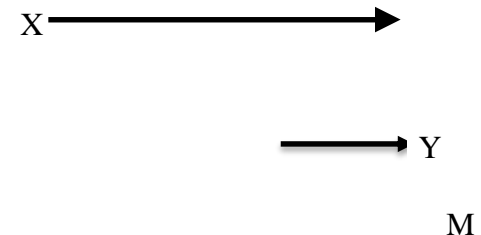
Step III: Conduct a simple regression analysis with \mathbf{M} predicting \mathbf{Y} to test the significance of path “b” alone. Or one may have:

$$\mathbf{Y} = \mathbf{B}_4 + \mathbf{B}_5\mathbf{M} + \mathbf{e}_2$$



Step IV: Conduct a multiple regression analysis with \mathbf{X} and \mathbf{M} predicting \mathbf{Y} . Or one may have: path “c” “

$$\mathbf{Y} = \mathbf{B}_6 + \mathbf{B}_7\mathbf{X} + \mathbf{B}_8\mathbf{M} + \mathbf{e}_3$$



¹ http://web.pdx.edu/~newsomj/semclass/ho_mediation.pdf

path “b”

In Step I to III, when one or more of these relationships are nonsignificant, then one may usually conclude that mediation are not possible or likely with exceptions from MacKinnon et al., 2007. If one assumes there are significant relationships from Step I to Step III, one can proceed to step IV. There is some form of mediations when path b remains significant after controlling for X. If M is controlled and X is no longer significant, the finding gives full mediation. When both X and M significantly predict Y, the finding provide partial mediation. When we go a further step, compare the above Eqt (1) with the the Eqt in Step IV, we get:

$$[M][LT] = (\mathbf{X} + \mathbf{XB} + \mathbf{XBD}) + (\mathbf{B}_0 + \mathbf{B}_0\mathbf{D}) + \mathbf{D}_0$$

$$= (\mathbf{I} + \mathbf{B} + \mathbf{BD})\mathbf{X} + \mathbf{B}_0\mathbf{D} + \mathbf{B}_0\mathbf{D}_0. \quad \text{—————(2)}$$

Obviously, from the regression equation in Step IV, the matrix (or vector) **D** in the term “**B₀D**” of equation (2) can be viewed as a wanted mediator. In other words, rather than the regression equation for prediction, we may thus construct a hypotheses Hayes model with a feasible moderator.

Indeed, there are defects of Baron and Kenny’s mediation test. Firstly, the test never assesses the significance of the indirect path or how X affects Y through the paths “a” and “b”. Secondly, the test suffers much from the Type II error of some true mediation effects (Mackinnon et al., 2007). Alternatively, there are two other approaches to test the mediation:

1. Judd & Kenny Difference of Coefficients Approach;
2. Sobel Product of Coefficient Approach,

Although there are drawbacks, this author will still employ Baron and Kenny’s regression method as a way to evaluate the finding causal relationships in chapter fifteen.

In brief, one can evaluate the Bayesian tree matrix by linking it to the casual relationship matrices through the linear transformation in forms of a polynomial matrices.

MAJOR RESULTS

A Regression Model Approximation to Causality

1. Granger Test for Causality

In order to test the proposed causal relationships, one may need to perform the Granger Causality Test. This author employs the software EViews for such Granger examination. One can find out the corresponding related diagram and hence the linked causal relation diagrams as shown in the next page

Pairwise Granger Causality Tests

Date: 01/10/20 Time: 10:44

Sample: 1 59

Lags: 2

Null Hypothesis:

D(INFLUENZA_1) does not Granger Cause D(TEMPERATURE)
D(TEMPERATURE) does not Granger Cause D(INFLUENZA_1)

Obs	F-Statistic	Prob.
56	1.383348	0.25998
	3.902697	0.02647

The temperature precede the number of influenza_1 case happening.

I.e., Temperature \longrightarrow Influenza_1

where " \longrightarrow " means a Granger cause

+

Pairwise Granger Causality Tests

Date: 01/10/20 Time: 10:50

Sample: 1 59

Lags: 2

Null Hypothesis:

D(INFLUENZA_1) does not Granger Cause D(CO2_1)
D(CO2_1) does not Granger Cause D(INFLUENZA_1)

Obs	F-Statistic	Prob.
56	2.934096	0.06221
	3.604373	0.03434

The amount of carbon dioxide in the atmosphere precedes the number of influenza_1 case happening.

I.e., CO2 \longrightarrow Influenza_1

where " \longrightarrow " means a Granger cause

+

Pairwise Granger Causality Tests

Date: 01/10/20 Time: 11:06

Sample: 1 59

Lags: 2

The Granger cause relation between COOLEST and Influenza_1:

Pairwise Granger Causality Tests

Date: 01/10/20 Time: 12:57

Sample: 1 59

Lags: 5

Null Hypothesis:

D(INFLUENZA_1) does not Granger Cause D(STRONGEST_WIND)
D(STRONGEST_WIND) does not Granger Cause D(INFLUENZA_1)

Obs	F-Statistic	Prob.
53	1.074871	0.38803
	4.900275	0.00127

the relation between Strongest_wind and Influenza_1

I.e., Strongest_wind precedes the Influenza_1 or

Strongest_wind \longrightarrow Influenza_1

where " \longrightarrow " means a Ganger cause

+

Null Hypothesis:

D(CO2_1) does not Granger Cause D(TEMPERATURE)
D(TEMPERATURE) does not Granger Cause D(CO2_1)

Obs	F-Statistic	Prob.
56	16.6114	3.E-06
	9.44135	0.0003

I.e., CO2 precedes the temperature nappening

Temperature precedes the CO2 happening

CO2 \longrightarrow Temperature

Temperature \longrightarrow CO2

where " \longrightarrow " means a Granger cause

+

Pairwise Granger Causality Tests

Date: 01/10/20 Time: 11:18

Sample: 1 59

Lags: 2

Null Hypothesis:

D(WETTEST_1) does not Granger Cause D(TEMPERATURE)
D(TEMPERATURE) does not Granger Cause D(WETTEST_1)

Obs	F-Statistic	Prob.
56	0.358226	0.70065
	4.303410	0.01874

The relation between temperature and wettest_1,

I.e., Temperature precedes the wettest_1 case

happening or

Temperature \longrightarrow wettest_1

Pairwise Granger Causality Tests

Date: 01/10/20 Time: 12:09

Sample: 1 59

Lags: 5

Null Hypothesis:

D(INFLUENZA_1) does not Granger Cause D(WETTEST_1)
D(WETTEST_1) does not Granger Cause D(INFLUENZA_1)

Obs	F-Statistic	Prob.
53	2.245001	0.06738
	3.100681	0.01800

The connection between Wettest_1 and Influenza_1

I.e., Wettest_1 precedes the Influenza_1 case

happening or

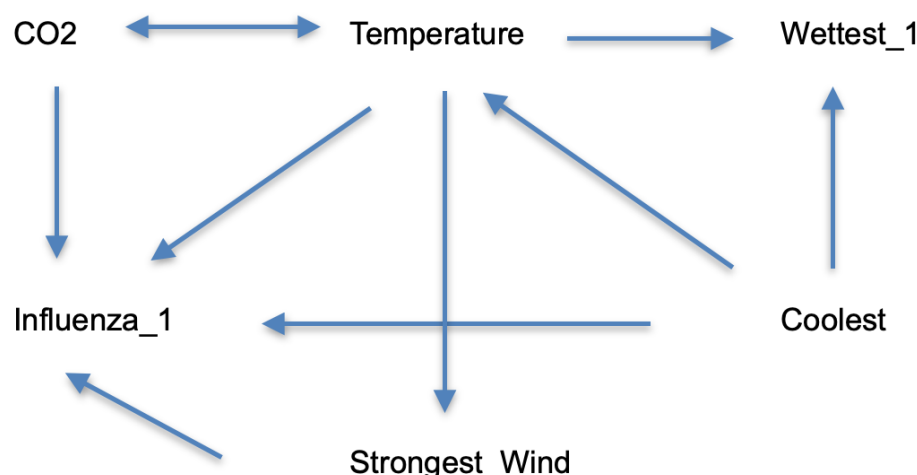


Figure 2: Granger

There are several

1. Coolest —
2. CO2 <—> Temperature —> Wettest_1 —> Influenza_1
3. CO2 <—> Temperature —> Strongest_Wind —> Influenza_1

Indeed, one can first convert the data set into time series by statistical programming software R first. Then one can apply the maximum likelihood estimation for the fitting of the above Granger Causal relation to build the corresponding models for the prediction. As the process is similar to the following mediation analysis by software SPSS add-on Hayes PROCESS, this author decides to omit the approximation.

2. Mediation analysis for the Causal Relationships found by SmartPLS (Lam, 2019)

Formally speaking, in order to test a causal relationship, one should perform causal analysis. Indeed, causal analysis = regression analysis (or any test) + theory (and hypothesis) or logical analysis

Usually, in statistics and economics, causality is often tested by regression (i.e. the present influenza relationship research case). Actually, one should perform exploratory causal analysis which is known as data causality or causal discovery. It is the use of statistical algorithms to infer associations in observed data set. In addition, these data sets are potentially causal under strict assumptions. Therefore, according to the results of chapter fifteen, there are potentially several cases of causal relationships. According to the aforementioned Baron and Kenny's Testing of mediation method, one obtains the following partial least square regression results (by R programming):

Step 1

```
> model = lm(influenza ~ wettest, data = australia_weather_influenza)
> summary(model)

Call:
lm(formula = influenza ~ wettest, data = australia_weather_influenza)

Residuals:
    Min       1Q   Median       3Q      Max
-0.6999 -0.3504 -0.1150  0.2660  1.3127

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.7699516  0.1066826  35.338  <2e-16 ***
wettest     -0.0001773  0.0001022  -1.734   0.0884 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4926 on 57 degrees of freedom
Multiple R-squared:  0.05008,    Adjusted R-squared:  0.03342
F-statistic: 3.005 on 1 and 57 DF, p-value: 0.0884
```

From the statistical data, one discovers that the p-value of the relationship is 0.0884 with the model equation equals to:

$$\text{influenza} = 3.7699516 - 0.0001773 \times \text{Wettest}$$

Hence one will reject the null hypothesis: number of influenza infected is unaffected by Wettest.

Step II

```
> model = lm(temperature ~ wettest, data = australia_weather_influenza)
> summary(model)

Call:
lm(formula = temperature ~ wettest, data = australia_weather_influenza)

Residuals:
    Min       1Q   Median       3Q      Max
-7.5544 -2.7871  0.1083  3.2289  5.0736

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.631e+01  7.359e-01  22.162 <2e-16 ***
Wettest       1.705e-03  7.053e-04   2.418  0.0188 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.398 on 57 degrees of freedom
Multiple R-squared:  0.09301, Adjusted R-squared:  0.0771
F-statistic: 5.845 on 1 and 57 DF, p-value: 0.01884
```

One observes that the p-value of the relationship is

0.01884 with the model equation equals to:

Temperature = $1.631e+01 + 1.705e-03 \cdot \text{Wettest}$

Hence one will reject the null hypothesis:

Temperature is unaffected by Wettest.

Step III

```
> model = lm(influenza ~ temperature, data = australia_weather_influenza)
> summary(model)

Call:
lm(formula = influenza ~ temperature, data = australia_weather_influenza)

Residuals:
    Min       1Q   Median       3Q      Max
-0.73352 -0.24386 -0.04267  0.24618  1.14992

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.21750    0.26193   19.919 < 2e-16 ***
temperature -0.08998    0.01449   -6.209 6.53e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3904 on 57 degrees of freedom
Multiple R-squared:  0.4034, Adjusted R-squared:  0.393
F-statistic: 38.55 on 1 and 57 DF, p-value: 6.532e-08
```

One finds that the p-value of the above relationship is $6.532e-08$ with the model equation equals to:

Influenza = $5.21750 - 0.08998 \cdot \text{Temperature}$

Hence one will reject the null hypothesis:

Influenza is unaffected by Temperature.

Step IV:

```
> model = lm(influenza ~ wettest + temperature, data = australia_weather_influenza)
> summary(model)

Call:
lm(formula = influenza ~ wettest + temperature, data = australia_weather_influenza)

Residuals:
    Min       1Q   Median       3Q      Max
-0.7376 -0.2486 -0.0272  0.2711  1.1425

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.214e+00  2.643e-01  19.729 < 2e-16 ***
Wettest      -2.627e-05  8.577e-05  -0.306  0.761
temperature -8.855e-02  1.534e-02  -5.772 3.56e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3935 on 56 degrees of freedom
Multiple R-squared:  0.4044, Adjusted R-squared:  0.3832
F-statistic: 19.01 on 2 and 56 DF, p-value: 4.988e-07
```

One observes that the p-value is $4.988e-07$ with the model equation equals to:

influenza = $5.214e+00 - 2.627e-05 \cdot \text{Wettest} - 8.855e-02 \cdot \text{Temperature}$

Hence, one will reject the null hypothesis: influenza is

unaffected by the combined effects of Wettest and

Temperature.

Step I: Wettest → Influenza cases

Step II: Wettest → Temperature

Step III: Temperature → Influenza cases

Step IV: Wettest + Temperature → Influenza cases

Hence, the proposed causal relationship — wettest and influenza is actually a fully mediated one with temperature as the immediate mediator. This event indirectly implies that wettest and influenza is actually a causal relationship.

2. The second one is the evaluation of the causal relationship “wind and influenza”: Wind → Wettest → Temperature → Influenza

The following listed is the partial least square result (Hayes Process Model Macro) that

obtained through software SPSS:

A. Wettest and Wind

```

OUTCOME VARIABLE:
Wettest

Model Summary
      R      R-Sq      MSE      F      df1      df2      p
      .2795      .0781 375399.450      4.8301      1.0000      57.0000      .0320

Model
      Coeff      se      t      p      LLCI      ULCI
constant -322.2481      532.0316      -.6057      .5471 -1387.6297      743.1335
Stronges      7.5870      3.4522      2.1977      .0320      .6741      14.4999

Standardized coefficients
      Coeff
Stronges      .2795

Covariance matrix of regression parameter estimates:
      constant      Stronges
constant 283057.659 -1815.9068
Stronges -1815.9068      11.9175
  
```

We observe that the model equation for

Wind and Wettest is:

$$\text{Wettest} = -322.2481 + 7.5870 \cdot \text{Wind}$$

B. Temperature and (Wind together with Wetest)

```

OUTCOME VARIABLE:
  temperat

Model Summary
      R      R-sq      MSE      F      df1      df2      p
      .3124      .0976    11.6920     3.0272     2.0000    56.0000    .0564

Model
      Coeff      se      t      p      LLCI      ULCI
constant    14.7740    2.9787    4.9599    .0000     8.8069    20.7411
Stronges     .0107     .0201     .5317    .5971    -.0295     .0509
Wetest       .0016     .0007     2.1581    .0352     .0001     .0031

Standardized coefficients
      Coeff
Stronges     .0703
Wetest       .2853

Covariance matrix of regression parameter estimates:
      constant    Stronges    Wettest
constant    8.8727    -.0579     .0002
Stronges    -.0579     .0004     .0000
Wetest      .0002     .0000     .0000

*****

```

We observe that the model equation for Temperature from (Wind and Wetest) is:

$$\text{Temperature} = 14.7740 + 0.0107 \cdot \text{Wind} + 0.0016 \cdot \text{Wetest}$$

C. Influenza and (Wind and Wetest and Temperature)

```

OUTCOME VARIABLE:
  influenz

Model Summary
      R      R-sq      MSE      F      df1      df2      p
      .4475      .2002    250048440     4.5897     3.0000    55.0000    .0061

Model
      Coeff      se      t      p      LLCI      ULCI
constant  43685.5285  16526.0736    2.6434    .0107  10566.3151  76804.7418
Stronges   26.7064    93.0281     .2871    .7751  -159.7273   213.1401
Wetest     -2.1845    3.5578    -.6140    .5417   -9.3144    4.9455
temperat -2054.0522   617.9784   -3.3238    .0016  -3292.5168  -815.5875

Standardized coefficients
      Coeff
Stronges     .0361
Wetest      -.0803
temperat    -.4219

Covariance matrix of regression parameter estimates:
      constant    Stronges    Wettest    temperat
constant  273111110  -1177928.9  12766.3258  -5642138.7
Stronges  -1177928.9  8654.2225  -82.1602  -4074.2063
Wetest    12766.3258  -82.1602  12.6576  -609.2217
temperat -5642138.7  -4074.2063  -609.2217  381897.287

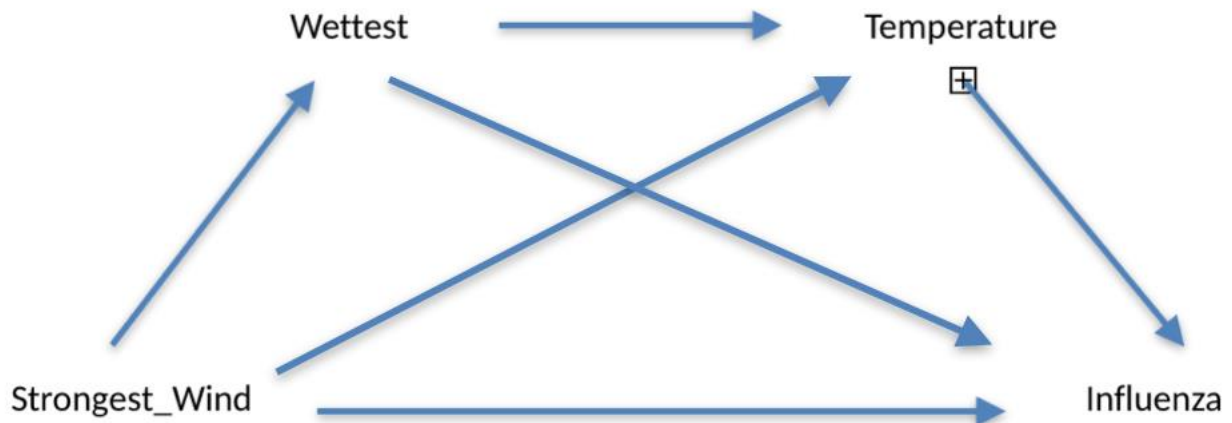
***** TOTAL EFFECT MODEL *****

```

We observe that the model equation for Influenza from (Wind, Wetest and Temperature) is:

$$\begin{aligned} \text{Influenza} = & 43685.53 + 26.71 \cdot \text{Wind} - \\ & 2.185 \cdot \text{Wetest} - 2054.05 \cdot \text{Temperature} \end{aligned}$$

All of the above data shows that there should be a Hayes PROCESS model 6 established as like the following:



This author notes that although the bootstrap confidence interval straddles zero which means that the mediation is not significant (or actually a border case), it doesn't imply that thing we are estimating is zero. Thus, this author finally concludes wettest and temperature are the mediators for wettest and influenza.

Furthermore, if we add the concentration of carbon dioxide as the moderator that lays between temperature and influenza, one may obtain the results below:

```

***** DIRECT AND INDIRECT EFFECTS OF X ON Y *****
Direct effect of X on Y
  Effect    se      t      p    LLCI    ULCI
  8.6924   94.3905   .0921   .9270  -180.6324  198.0172

Conditional and unconditional indirect effects of X on Y:

INDIRECT EFFECT:
Stronges -> Wetttest -> influenz
  Effect    BootSE    BootLLCI    BootULCI
  -9.8542   19.5547   -62.4032   13.8536

INDIRECT EFFECT:
Stronges -> temperat -> influenz
  CO2_1    Effect    BootSE    BootLLCI    BootULCI
  398.9220  -29.5673   64.3348  -127.4806   136.0163
  403.6400  -21.9630   49.1415  -100.2869   102.1029
  408.9040  -13.4787   37.3614  -83.1024    73.7515

Index of moderated mediation:
  Index    BootSE    BootLLCI    BootULCI
CO2_1     1.6118     4.1492    -9.1840     8.6201
---

INDIRECT EFFECT:
Stronges -> Wetttest -> temperat -> influenz
  CO2_1    Effect    BootSE    BootLLCI    BootULCI
  398.9220  -33.5439   31.4431  -124.8538   -3.0546
  403.6400  -24.9169   24.9493  -97.3552   -2.1555
  408.9040  -15.2915   23.3582  -83.9560    3.3906

Index of moderated mediation:
  Index    BootSE    BootLLCI    BootULCI
CO2_1     1.8285     2.5126    -1.3135     8.5707
---

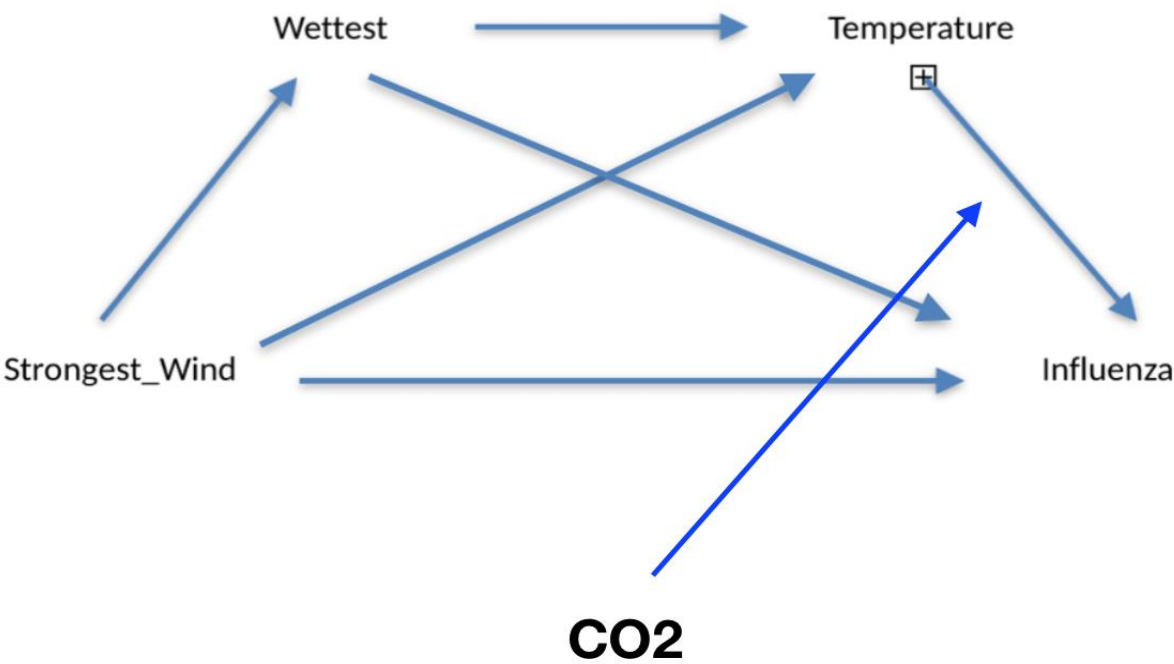
```

The aforementioned outcome tells us that the index of moderated mediation is referring to the “weight for the moderator in a linear function relating to the size of indirect effect of X on Y to the moderator” (Hayes, 2018,

p.491).

When the index is not zero, this will mean the indirect effect relates linearly with the moderator. Hence, one can claim there is a moderated mediation (Hayes, 2018). On the contrary, if the bootstrap confidence interval does not includes zero, this event implies the indirect effect will not relate linearly with the moderator (Hayes, 2018). Therefore, we conclude that carbon dioxide is the moderator to temperature and influenza. This event is because the above data shows that both the index is greater than zero and the confidence interval contains a zero.

Then, one may obtain the Hayes PROCESS model 87 as shown below:



In addition, if one changes the independent variable from wettest_1 to temperature (while wettest_1 and strongest_wind become the mediators), we will get the following results:

```
***** DIRECT AND INDIRECT EFFECTS OF X ON Y *****
```

Direct effect of X on Y

Effect	se	t	p	LLCI	ULCI
-1898.6041	646.8798	-2.9350	.0049	-3194.9887	-602.2194

Conditional and unconditional indirect effects of X on Y:

INDIRECT EFFECT:

Effect	BootSE	BootLLCI	BootULCI
-290.2719	220.5588	-776.4535	87.7065

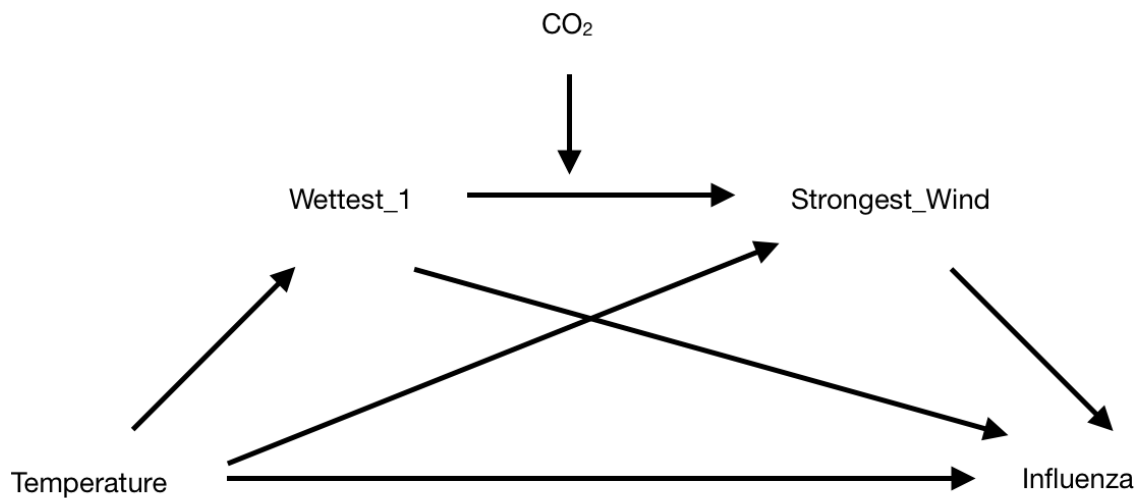
INDIRECT EFFECT:

Effect	BootSE	BootLLCI	BootULCI
-.0614	63.2151	-111.2673	155.1546

INDIRECT EFFECT:

Effect	BootSE	BootLLCI	BootULCI	
CO2 1	Effect	BootSE	BootLLCI	BootULCI
398.9220	14.1275	44.2673	-61.9476	116.5677
403.6400	36.6441	62.5905	-42.0169	194.8251
408.9040	61.7665	109.6117	-58.7893	370.4109

It is obviously observed from the above data that the index of moderated mediation has increased greatly from 2 to nearly 5 if we made the amendment of the change to temperature as the independent variables. Then the concentration of carbon dioxide becomes the moderator between wettest_1 and strongest_wind. This means that the indirect effect depends heavily on the moderator carbon dioxide. According to Hayes in 2018, the index of moderated mediation is slope of the equation formed by indirect effect. If it is equal to zero (i.e., flat) then the indirect effect is not related to the moderator. However, if the index is strongly greater than zero, the indirect effect depends heavily on the moderator. Thus, carbon dioxide acts as the wanted moderator. Therefore, we conclude that the Hayes PROCESS model 91 is the most suitable one for describing the relationship between different weather variables and the number of case in influenza in Australia when compared to the previous one. The final and conclusive model is shown in below:



CONCLUSION

To conclude, the above test and model tell us that wettest_1 and the temperature are the most feasible causal relationship to the number case of influenza. The results are obtained from the mediation analysis (Baron and Kenny's Testing and the Haye's PROCESS modelling for SPSS). Similarly, one can show that the sequenced domino effects is indeed a list of causal relations. Practically, one should find out all of the possible regression models (in this case, one should employ the carbon dioxide and strongest_wind as the independent variables respectively for the dependent variables number of case in influenza, other variables such as temperature and Wettest_1 will be used as mediator together with suitable moderators). Then, one should compare the r-square, r-square(adjusted) and r-square(predicted) for each of these calculated models. The aim is to find out the best goodness fit in these models from these r values and hence selects the best fitted model for the wanted causal relations. While at the same time, the concentration of carbon dioxide acts as the moderator that lays between Wettest_1 and strongest_wind provide that temperature as the independent variable (model 91). On the other hand, the concentration of carbon dioxide also acts as the moderator between the mediator temperature and the dependent variable number of case in influenza (model 87). The most significant discovery in this paper is that the role of carbon dioxide is indeed a moderator. It gives a moderated effect to other mediators. This author remarks that a moderator is difference from a mediator in that moderator only affects the strength of the concerning variables. While the mediator can explain the relationship between two variables. The existence of CO₂ as a moderator implies that the gas has a conditional effect to the number of cases of influenza infected. Hence, there may be a thermal degradation from CO₂ to CO at around 20°C (Asperen et al., 2015). But the

symptoms of carbon monoxide poisoning are similar to the common flu infected. In addition, the dissolved carbon dioxide can have an influence during the production of recombinant hemagglutinin component that induced from an influenza vaccine by insect cells (Meghrou et al., 2015). These are the reasons for the conditional effects of CO₂ in the number of case of influenza infected. Therefore, I suggest there should be a reduction in the emission of pollutants like CO₂ before and during the peak months of common flu infection. All in all, the above Hayes Process model 91 can help us verify the truthiness of this author's HKLam Net-Seizing Theory. In other words, all of the causal relations can be find by a mediation analysis. In addition, all of the causal relationships can be examined by Granger Causality Test followed by Maximum Likelihood estimation for fitting into the models. Hence the second part of the philosophy — domino effects can be expressed as a causal relationship is proposed to be correct.

Remarks:

There are lots of causality around our world. One of the case is the causal relationships during the procedure of making proteins among the living organisms that like the following:

DNA + RNA polymerase → pre-mRNA → mRNA → Ribosome → polypeptide (protein molecule)

According to the results of this paper, one may apply the similar technique which can generate the wanted regression model to approximate the above protein production causality process. Then, one may finally mass-produce the deserved protein or even the so-called “protein factory”. When those virus enters our cells, most virus have their own connexion respectively. Then we can use the corresponding shaped protein to attach on the surface of the virus, this event can prevent the virus to enter our body cells. To go a further step, the methylation for DNA and mRNA may act as a moderator, mediator or control in the disease of (Kidney) cancer etc^{2,3}.

Another application is in the field of ocean science. In the year 2014 of U.S. California state, due to the increase in global temperature, there were a sudden large amount of starfish finding to be died. Since starfish is the natural enemy of purple sea urchin, then there are huge sum of them appearing. They eat lots of kelp which decrease the number of other related sea living organisms like abalone. Hence, the final outcome is a harmful effect to the U.S. economy. The aforementioned phenomenon can also use this paper's technique to generate or approximate the wanted regression model for such kind of food chain causality.

Suppose we have the following set of data w.r.t the moderator regression equation:

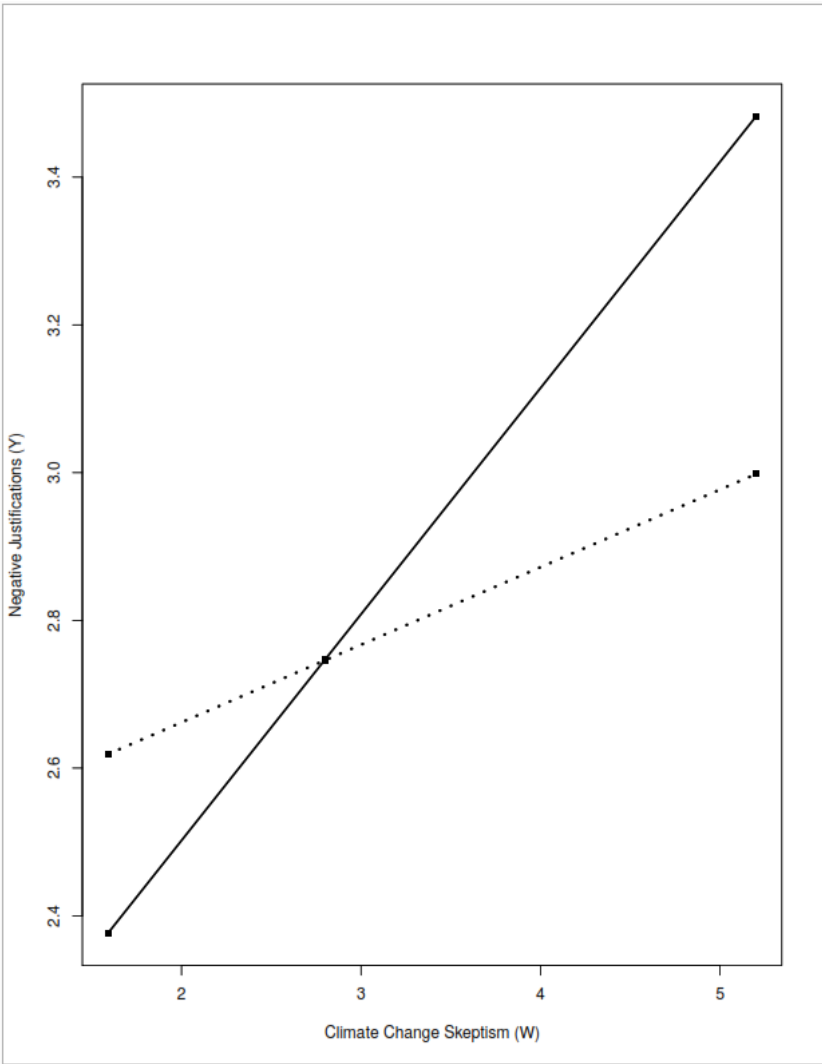
$$Y = 2.452 - 0.562X + 0.105W + 0.201XW$$

X (FRAME)	W(SKEPTIC)	Y
0	1.592	2.619
1	1.592	2.377
0	2.800	2.746
X (FRAME)	W(SKEPTIC)	Y
1	2.800	2.747
0	5.200	2.998
1	5.200	3.48

² <https://www.nature.com/articles/s41467-019-13527-1>

³ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6500370/>

After fitting the above data into the R programming software (R-Studio), we get the following diagram:



The above diagram depicts how the framing condition interacts with the climate change skepticism. This event can also be applied to the moderator regression as mentioned in this paper before. The implication can then be extended to the goal-directed multi-party interaction in the case of robot moderator⁴.

Furthermore, according to Yaron in 2020, the level of nitrogen dioxide are acting as a contributing factor to the number of death cases that caused by coronavirus (COVID-19)⁵. Hence, from both of the case in influenza (CO2 acts as a mediator in the regression model of weather pattern) and the NO2 concentration in the various geographical regions to the one as mentioned earlier, we may conclude that air pollutants have their definite roles in different kinds of infectious disease. My suggested way of reducing the cases of infected by these diseases is to give serious warnings during weather report and aids high risk groups to purchase electrical air-filters (which can be effectively in reducing NO2 and CO2 pollutants etc) among their living environments. One may also need to avoid unnecessary exploration in those critically polluted outdoor areas.

It may be true that both the model 87 and 91 as described above should contain some moderators and control

⁴ <https://www.aaai.org/ocs/index.php/FSS/FSS15/paper/view/11658>
⁵ Yaron, O. (2020), Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality, Science of the Total Environmental Journal, Elsevier publication

variables. However, the main focus of this paper is — I have shown the ultimate result that CO₂ has a significant role (mediator) in the weather-influenza regression model. To be precise, the concentration of CO₂ mediates the wettest and strong wind. This event can finally affect the number of cases of influenza infected. Hence, both the moderator and the control variables are in light of their importance in the final model. Similarly, it may be also true that one should perform further heterogeneity analysis in order to find out data distributions. Then we may go ahead to purify those data outsiders through the improvement of a better influenza drug such as the extension to the effective period of a vaccine. However, as I have mentioned before, I only want to discover how our environment may affect the cases of influenza infected. Therefore, heterogeneity analysis is out of the scope of the present discussion. On the other hand, if we can invent an air purification machine which can immediately generate the indoor air heterogeneity model and eventually eliminate those pollutants like CO₂ and NO₂, then this can be a great achievement in the treatment of COVID-19 and the prevention of influenza infection. One may even go a further step, when the machine can compute the minimum path (by the Euler Lagrange Equation) during the instantaneous time of squirt from a COVID-19 patient and filter (catch) those virus, one can even drop down the transmission rate of such disease.

For the later part of my HKLam theory, one may establish a required linear regression model for the necessary prediction wanted. Indeed, it is also true that for a random chance phenomenon, we may stimulate the same situation through the monte-carol one by the following steps⁶:

1. Establish a suitable probability distribution;
2. Accumulate the distribution in step 1;
3. Set the random number intervals;
4. Generate the random numbers;
5. Find the answer of question that prescribed by using the above four steps.

Actually, my HKLam theory differs from the Monte-Carol stimulation that it has the ability to catch the butterfly effect occurring in the nature phenomena like the earthquakes or weather prediction etc. My theory has the properties of net-seizing — it can be in a relaxing status or in a tension whenever necessary to be applied for the object (most likely the butterfly effect) that has fallen in it (Lam March, 2020). Then one can maintain a high flexibility of controlling chaos or even create an equilibrium between linear dominos and disorder chaos. As in the first part of my HKLam Theory is concerning the Bayesian tree, one may apply the Monte-Carol simulation to compute both of the forwarding and the converse of Bayesian Inference (or conditional probability distributions) for the corresponding cancer mutations as I have mentioned in Lam Dec, 2019. In addition, we can minimise the predictive error function by modified Bayesian optimisation that supplemented with dialectic method (Lam May, 2016) for the best machine learning rate. Practically, the modified Bayesian optimisation is the surrogate loss function with the SPO+ loss one⁷. Since, the above (real data fitting) predictive and optimisation procedure belongs to the in-depth study of computer science. At the same time, we can update the above stated procedure for the model predictive controller which is widely used for the advance process control in different industrial applications. In realities, the heterogeneity system as mentioned before can be matched with the model predictive controller for the best predictive purification of indoor atmosphere with minimum number of virus appeared.

Practically, I will outline the philosophy of air purification. One may program our previous machine for stimulation of the indoor atmosphere. We may do this event by first applying the Bayesian distribution for the COVID-19 patients' squirt. Then the machine will predict all of the feasible paths for the virus transmission. The following step is the machine's calculation to find the minimum of the path according to the Euler Lagrange equations. We may then optimise the least of number of COVID-19 virus that appeared in the path and

⁶ <https://www.nature.com/articles/s41524-018-0137-0>

⁷ <https://arxiv.org/abs/1710.08005>

cannot lead to the infection. Next, the machine will perform the heterogeneity analysis to select those pollutants. Finally, one may apply my HKLam Theory to capture most of the virus (COVID-19) or pollutants (NO₂) on the minimum path etc. Hence, we may filter our air. The outcome is one can minimise the infection or the death rate and can be compared with the risk management before making any important decisions (Lam March, 2020). However, this paper will not concern fully how the machine be made. Only the conceptual parts of the computer programming codes and electronics (Lam Jan, 2019 & June, 2019 & March, 2020) will be discussed. The implementation (reality) details are indeed out of the scope of my present papers' study.

I note further that, in partial least square-structural equation model (PLS-SEM), rather than the linear relationships between constructs, there are also cases of non-linear relationships for those latent variables. One of the possible way to solve the problem is through the logarithmic transformation. In other words, we may apply the base 10 logarithm to every observation. Hence, for the quadratic effects, we may have for example:

$$Y_2 = 2Y_1 + Y_1^2$$

For the cubic effects, we may have for example:

$$Y_2 = 2 Y_1 + Y_1^2 - Y_1^3$$

Although I cannot exclude the possibilities of quadratic or cubic effects in the present weather-influenza pattern model, it seems from the major result (CO₂ acts as the mediator-moderator) that these effects do not show great influence to the role of CO₂.

According to Joseph, et al., 2018 (p.137), one may reduce the temperature variable into three types — hottest, warmest and the coolest. Thus, temperature may be a categorical moderator variable (non-theoretically) that splits the data set into three groups. Then we may need to estimate three separate models. That is we may need to further perform heterogeneity analysis between these three groups. However, as I only want to find out how the pollutants like CO₂ may affect the weather-influenza infection pattern, the average temperature data is applied in the model (a kind of methodology or even a philosophy) instead of an in-depth permutation test among the stated three groups.

Finally, I remind that the second generation activity theory is very similar to the model 91 that I have discovered for the weather-influenza case pattern. When one goes into a depth, one may discover that there are interactions underlying among complexity theory, chaos theory and activity theory^{8,9}. One may further expand the second generation activity theory to the third generation or even the fourth one for the complexity of our weather system (or even to model the atmospheric changes). Hence, one can use the net-seizing properties in HKLam theory to manipulate it.

Reference:

- Asperen, H.V., Warneke, T., Sabbatini, S., Nicolini, G., Papale, D., Notholt, J., (2015) The role of photo- and thermal degradation for CO₂ and CO fluxes in an arid ecosystem. *Biogeosciences*, 12, 4161–4174 doi:10.5194/bg-12-4161-2015
- Joseph, H.F.JR, Sarstedt, M, Ringle, C.M, Gudergan, S.P., (2018) *Advanced Issues in Partial Least Squares Structural Equation Modeling*, SAGE Publication
- Joseph, H.F.JR, Hult, G.T.M, Ringle, C.M, Sarstedt, M, (2017) *A Primer on Partial Lease Squares Structural Equation Modeling*, SAGE Publication
- Hayes, F.A, (2018) *Introduction to Mediation, Moderation, and Conditional Process Analysis — A Regression-Based Approach*, GUILIFORD Publication
- Meghrou J, Khramtsov N, Buckland BC, Cox MM, Palomares LA, Srivastava IK, Dissolved carbon dioxide determines the productivity of a recombinant hemagglutinin component of an influenza vaccine produced by insect cells, Wiley Publication. <https://www.ncbi.nlm.nih.gov/pubmed/25943562>

⁸ <https://onlinelibrary.wiley.com/doi/10.1111/j.1540-4781.2012.01279.x>

⁹ Jennifer, C., Steven, J.D, (2015) *Researching Medical Education*, Wiley Publication