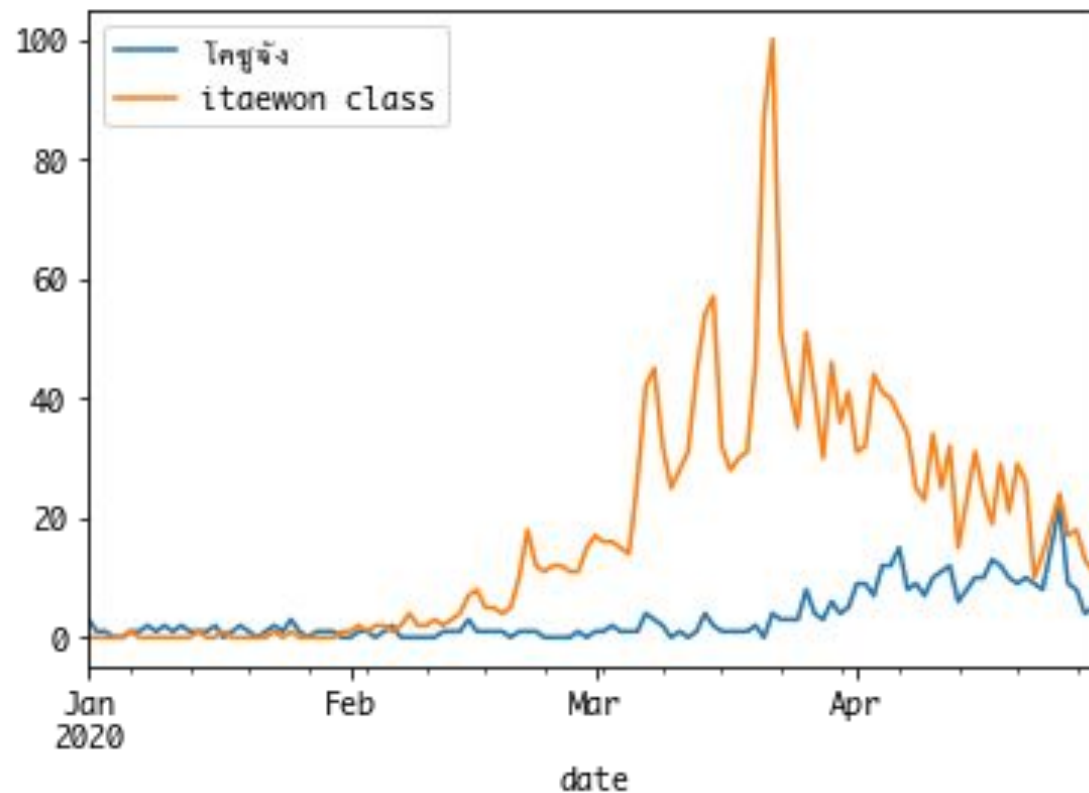# Case study: Did itaewon class cause โคชูจัง?

Granger Causality with Google Trend

Time Series Data

- 'itaewon class' and 'โคซูจัง' are a time series data that are collected from google trend since 2020-01-01 to 2020-04-30.

Non-stationary Vs. Stationary

- Time series data that can be used for forecasting must be stationary time series.

$$\text{Mean :} \quad E(Y_t) = \mu$$

$$\text{Variance :} \quad Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$$

$$\text{Covariance :} \quad E\left[(Y_t - \mu)(Y_{t+k} - \mu)\right] = \gamma_k$$

- Stationary time series has the properties as mean, variance and covariance are constant (same value) across time.

- Normally, time series data is a non-stationary.

- In statistics, using non-stationary time series for analysis can cause spurious regression and assumptions for analysis not being valid.

Examples of Non-Stationary Time Series
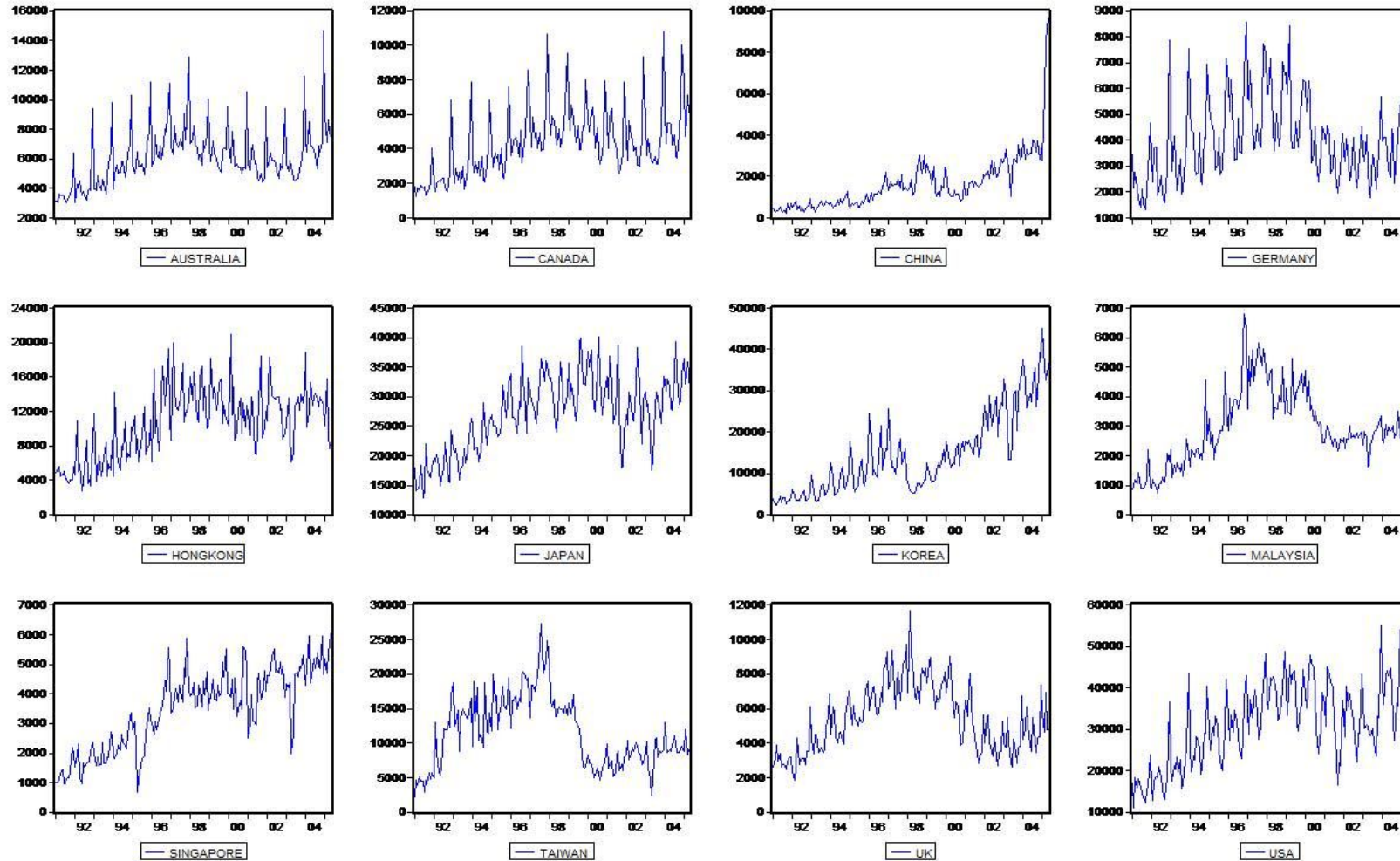
Examples of Stationary Time Series

Source: https://drsifu.wordpress.com/2012/11/27/time-series-econometrics/
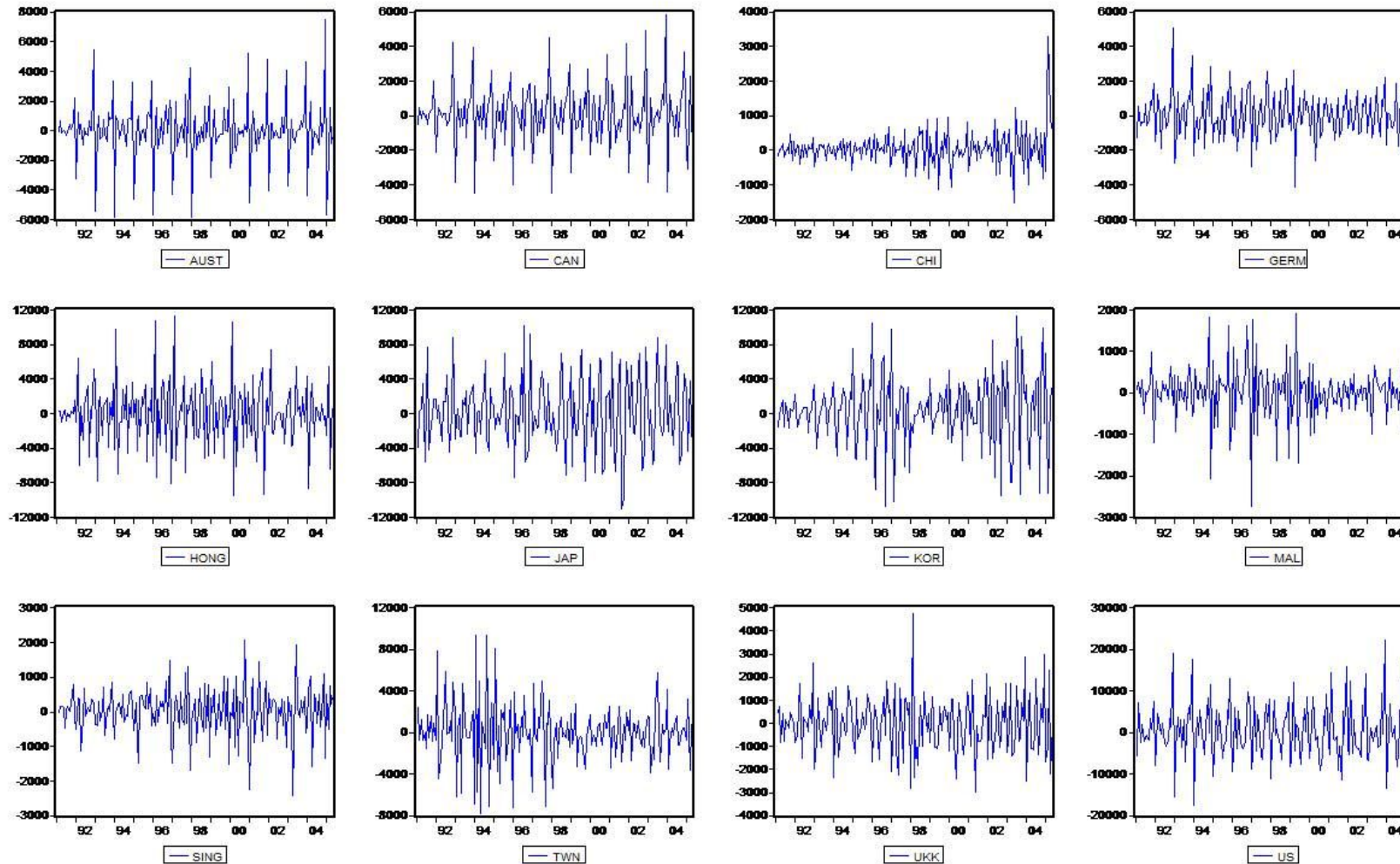
What is Spurious regression?

- Spurious regression is a problem that arises when regression analysis indicates a strong relationship between two or more variables but in fact they are totally unrelated.

- Regression characteristics expected to be Spurious Regression.
  - R^2 is typically very high.
  - t-statistic value most often is significant.
  - Durbin-Watson statistic (DW) is low.
  - R^2 of the regression is greater than the Durbin-Watson Statistic.

Unit Root Test

- Therefore, it needs to check whether Time series is stationary.

- Hypothesis:
  - Null Hypothesis (H_0): time series has a unit root, meaning it is non-stationary.
  - Alternate Hypothesis (H_1): time series does not have a unit root, meaning it is stationary.

- Unit Root Test is a test for checking stationary of data that are various methods:
  - Dickey Fuller (DF)
  - Augmented Dickey and Fuller (ADF)
  - Etc.

```
Augmented Dickey-Fuller Test on "โคชูจัง"
--------------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level      = 0.05
Test Statistic          = -0.3921
No. Lags Chosen         = 5
Critical value 1%       = -3.489
Critical value 5%       = -2.887
Critical value 10%      = -2.58
=> P-Value = 0.9114. Weak evidence to reject the Null Hypothesis.
=> "โคชูจัง" is Non-Stationary.


Augmented Dickey-Fuller Test on "itaewon class"
--------------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level      = 0.05
Test Statistic          = -1.2427
No. Lags Chosen         = 7
Critical value 1%       = -3.49
Critical value 5%       = -2.887
Critical value 10%      = -2.581
=> P-Value = 0.6549. Weak evidence to reject the Null Hypothesis.
=> "itaewon class" is Non-Stationary.
```
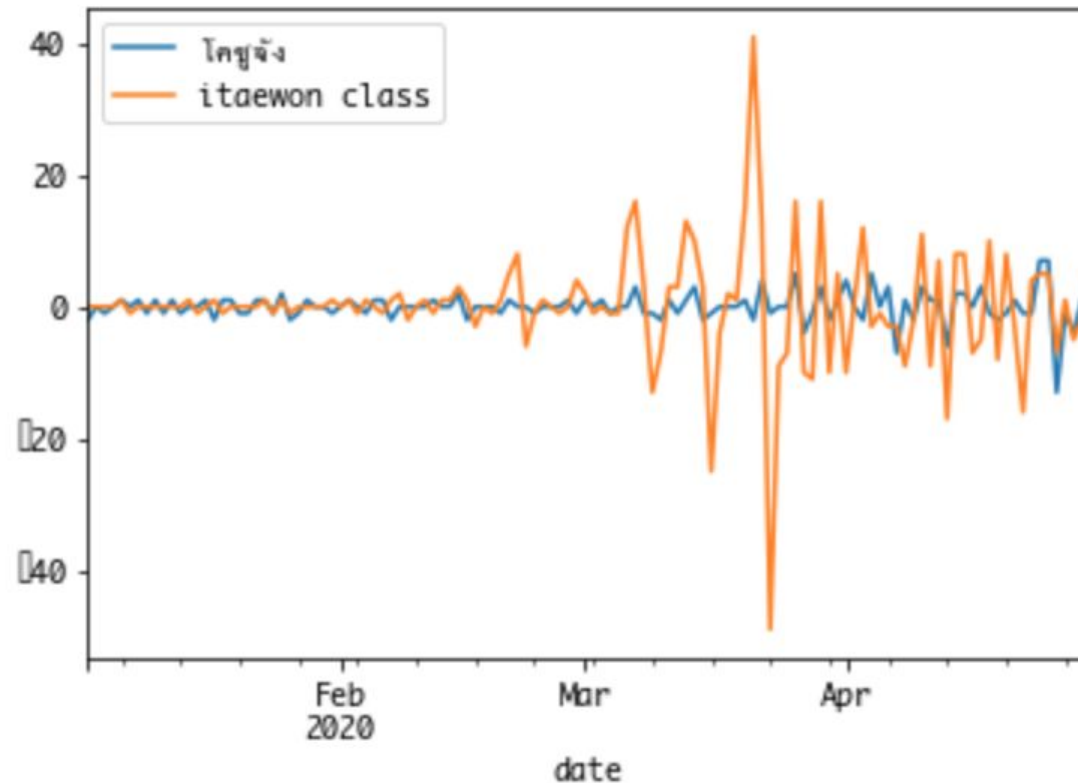
- p-value > 0.05: Fail to reject the null hypothesis (H_0).
- So, "โคชูจัง" and "itaewon class" have a unit root. They are non-stationary.

Unit Root problem

- Unit Root problem can solve by taking different.
- After taking different at level(1), can plot graph as following graph.

```
Augmented Dickey-Fuller Test on "โคชูจัง"
-------------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level      = 0.05
Test Statistic          = -8.9465
No. Lags Chosen         = 4
Critical value 1%       = -3.489
Critical value 5%       = -2.887
Critical value 10%      = -2.58
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.

Augmented Dickey-Fuller Test on "itaewon class"
-------------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level      = 0.05
Test Statistic          = -5.4473
No. Lags Chosen         = 6
Critical value 1%       = -3.49
Critical value 5%       = -2.887
Critical value 10%      = -2.581
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.
```

- p-value <= 0.05: Reject the Null hypothesis (H_0).
- So, "โคชูจัง" and "itaewon class" have no unit root. They are Stationary.

Lag length

- A time lag is a delay between an economic action and a consequence.

- Very often, the dependent variable responds to an independent variable with a lapse of time.

- For Granger causality test, it needs to define an optimal lag for testing.

- The optimal lag is selected from considering p-value of following criterias:
  - ssr based F test (ssr_ftest)
  - ssr based chi2 test (ssr_chi2test)
  - likelihood ratio test (lrtest)
  - parameter F test (params_ftest)

Optimal lag
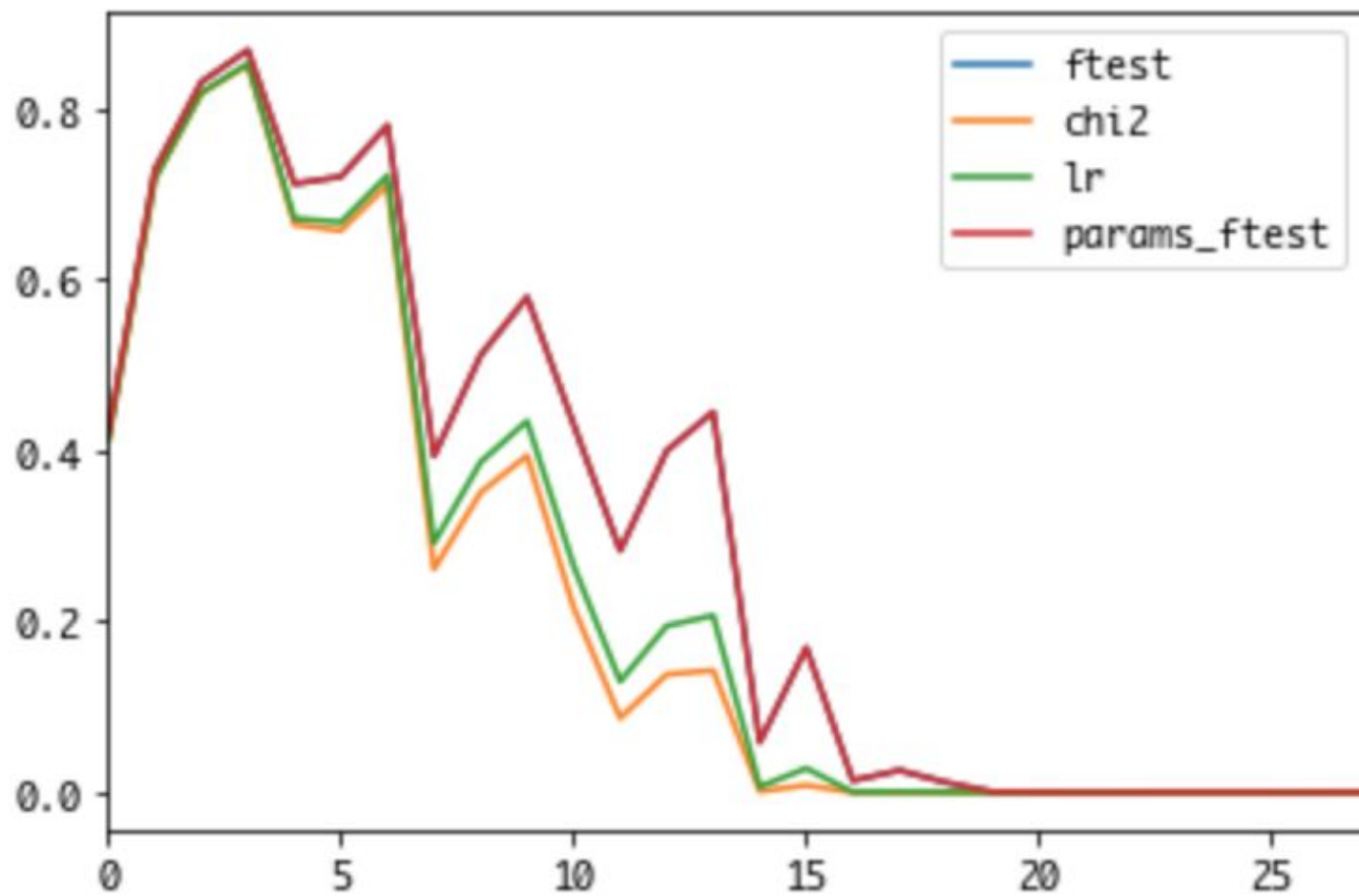
Granger causality

- Granger causality is a statistical concept of causality that is based on a prediction.

- Granger's Causality test on time-series data gives evidence that variable A Granger-causes B.

- If A Granger-causes B, then past values of A should contain information that helps predict B above and beyond the information contained in past values of B alone.

- Type of causality
  - Uni Directional Causality
    - A →B
  - Bi-directional Causality
    - A ↔ B
  - No directional Causality

Granger causality - Step

1. Develop a base case, autoregressive model, using a dependent variable and its lagged values as an independent variable.

2. Develop a test case, multivariate model, by adding a second lagged independent variable that you want to test.

3. Calculate the R-Square (the square of resident error) for two models and run F-test and t-test to check if the residuals are significantly lower when you added tested the second variable.

4. Redo step 1 to 3, but reverse the direction. By comparing the tests significance or p-value, you can see if A Granger-causes B more than B Granger-causes A.
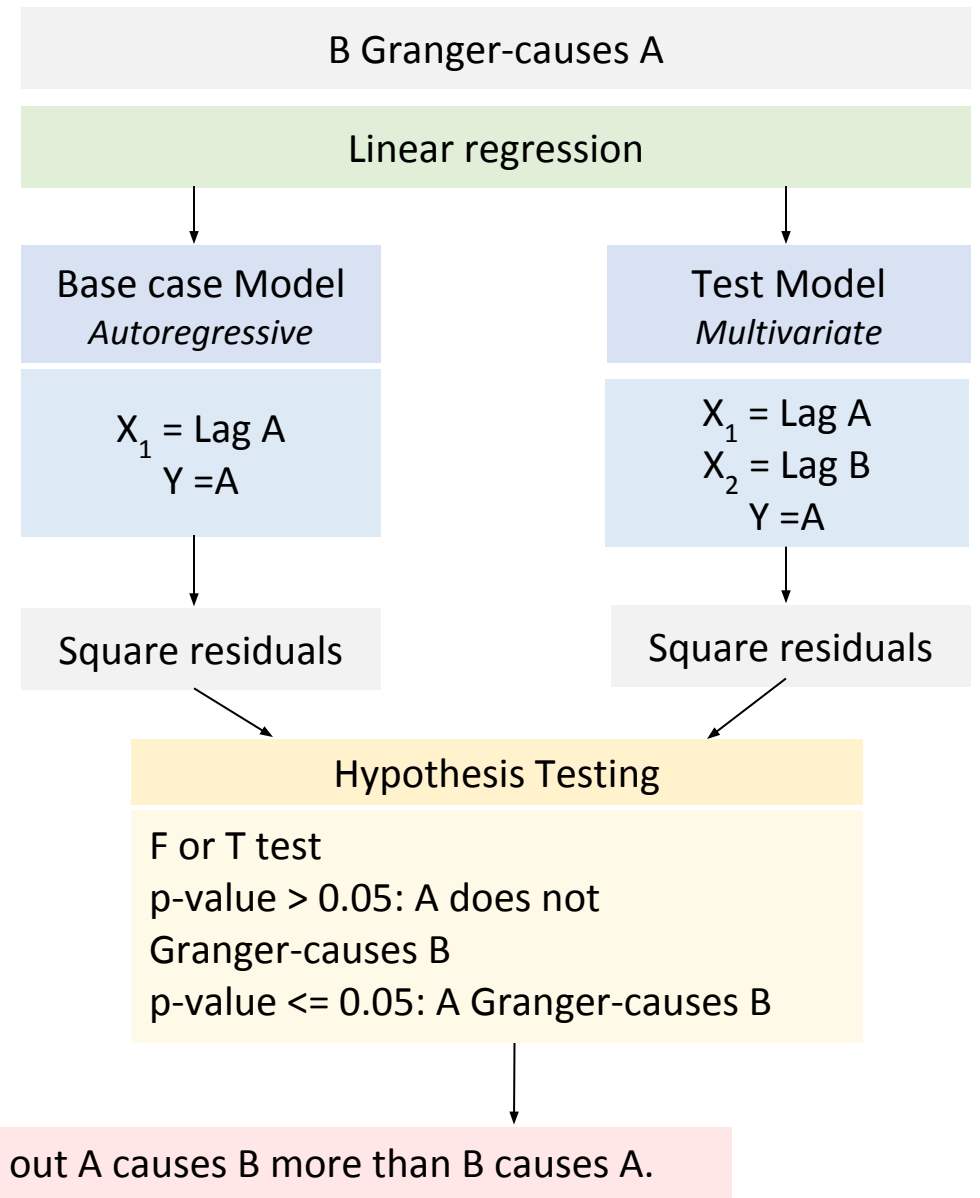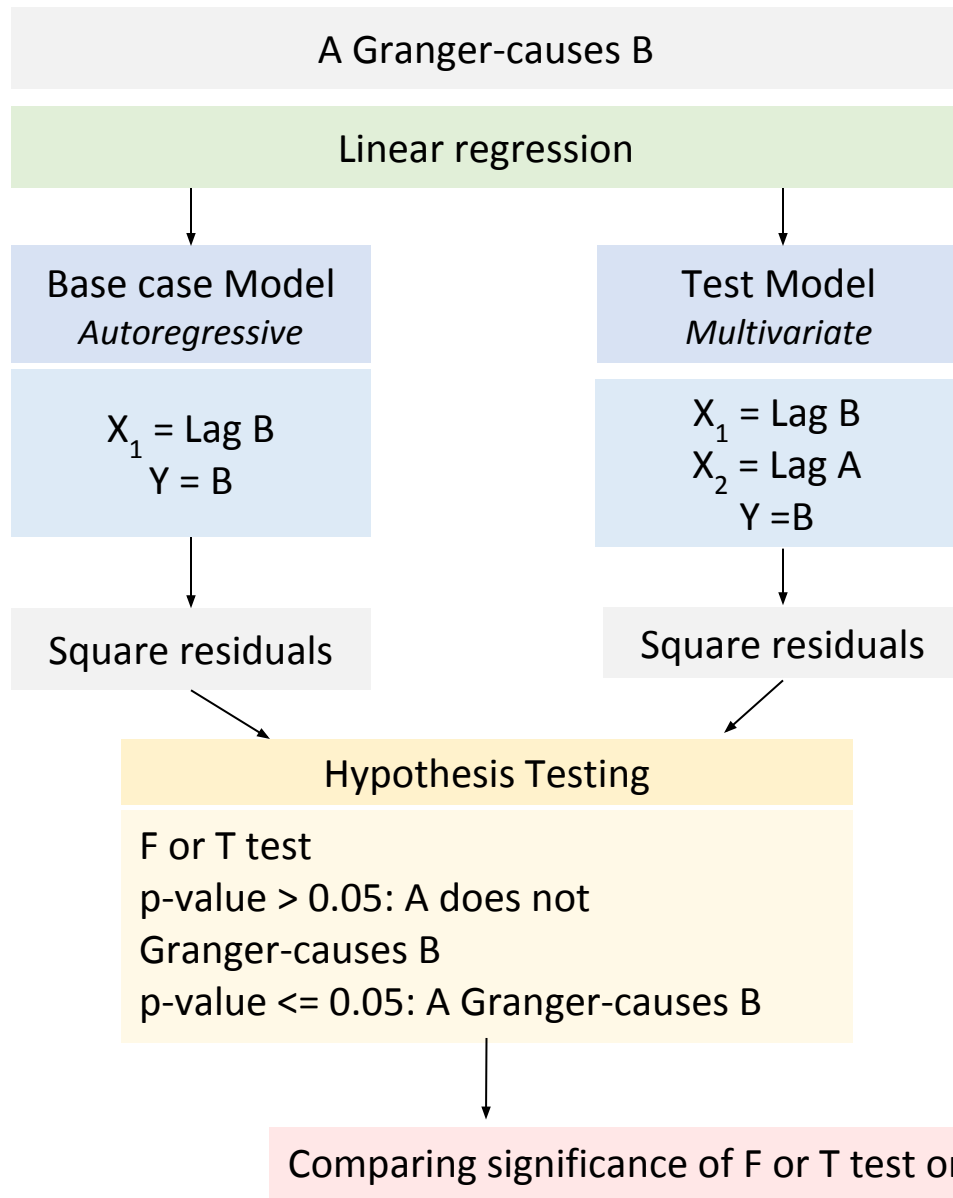
Granger causality - Interpret result

- Hypothesis:
  - H_0: $\beta_1 = \beta_2 = \ldots = \beta_m = 0$: no relation
  - H_1: at least one non-zero : have relation
- p-value > 0.05: Accept the null hypothesis (H_0), A does not Granger-causes B.
- p-value <= 0.05: Reject the null hypothesis (H_0), A Granger-causes B.

Granger causality results

	โคชูจัง_x	itaewon class_x	test
โคชูจัง_y	1.0000	0.0229	ssr_ftest
itaewon class_y	0.0253	1.0000	ssr_ftest
โคชูจัง_y	1.0000	0.0000	ssr_chi2test
itaewon class_y	0.0000	1.0000	ssr_chi2test
โคชูจัง_y	1.0000	0.0007	lrtest
itaewon class_y	0.0004	1.0000	lrtest
โคชูจัง_y	0.0000	0.0229	params_ftest
itaewon class_y	0.0253	0.0000	params_ftest

- p-value <= 0.05: Reject the Null hypothesis (H_0)
- "itaewon class" causes "โคชูจัง" and "โคชูจัง" causes "itaewon class"
- So,"itaewon class" and "โคชูจัง" are Bi-directional Causality