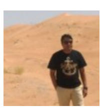
Using Granger Causality Test to Know If One Time Series Is Impacting in Predicting Another?

medium.com/swlh/using-granger-causality-test-to-know-if-one-time-series-is-impacting-in-predicting-another-6285b9fd2d1c

July 31, 2020



Granger causality test is used to determine if one time series will be useful to forecast another variable by investigating causality between two variables in a time series. The method is a probabilistic account of causality; it uses observed data sets to find patterns of *correlation*. One good thing about time series vector autoregression (VAR) is that we could test 'causality' in some sense. This test is first proposed by Granger (1969), and therefore we refer to it as the Granger causality.

It is based on the idea that if X causes Y, then the forecast of Y based on previous values of Y AND the previous values of X should best result in the forecast of Y based on previous values of Y alone.

Granger causality should not be used to test if a lag of Y causes Y. Instead, it is generally used on exogenous (not Y lag) variables only. In simple terms '*X* is said to Granger-cause *Y* if *Y* can be better predicted using the histories of both *X* and *Y* than it can by using the history of *Y* alone'

When performing Granger Causality Test we need to consider two assumptions:

1. Future values cannot the past values.
2. A notably distinct information is contained in about effect which will not be available elsewhere

> Python implementation of **statsmodel** package for the Granger test.

Lets do the code:

```
import pandas as pdimport numpy as npimport matplotlib import seaborn as snsimport
randomimport matplotlib.pyplot as pltfrom dateutil.parser import parsefrom scipy
import signalfrom scipy.interpolate import interp1dfrom scipy import statsfrom
statsmodels.tsa.stattools import adfuller, kpss, acf, pacf, grangercausalitytestsfrom
statsmodels.nonparametric.smoothers_lowess import lowessfrom statsmodels.tsa.seasonal
import seasonal_decomposefrom statsmodels.graphics.tsaplots import plot_acf,
plot_pacffrom sklearn.metrics import mean_squared_error%matplotlib inline# Upload
Data in a DataFramedf = pd.read_csv('IxR_Data.csv', parse_dates=
['Value_Date'])df.set_index(['Value_Date'])# Get few recordsdf.head()
```

```
In [3]: df.head()
executed in 14ms, finished 15:06:30 2020-07-21
Out[3]:
```

	Value_Date	IxD	RxB
0	2020-07-20	109.50	424.0
1	2020-07-17	110.45	420.0
2	2020-07-10	107.05	405.0
3	2020-07-03	101.20	409.5
4	2020-06-26	104.35	412.0

> We will do some EDA …. Not relevent at this to show here. But Just wanted to show you that how the series are behaving in data time series

```
ax = sns.lineplot(x="Value_Date", y="IxD", data=df)ax1 = sns.lineplot(x="Value_Date",
y='RxB', data=df)
```

```
ax = sns.lineplot(x="Value_Date", y="IxD", data=df)
ax1 = sns.lineplot(x="Value_Date", y='RxB', data=df)
```

executed in 642ms, finished 15:06:33 2020-07-21

We can see from the above chart that both *Time Series* move (*not so)* simultaneously, and lead to the inter relationship can be use for the prediction of any variable.

> Now we are going to use Gregner Test on the data series

```
res = grangercausalitytests(df[['IxD', 'RxB']], maxlag=15)
```

> Results:

```
Granger Causality
 number of lags (no zero) 1
 ssr based F test: F=13.3271 , p=0.0003 , df_denom=519, df_num=1
 ssr based chi2 test: chi2=13.4041 , p=0.0003 , df=1
 likelihood ratio test: chi2=13.2349 , p=0.0003 , df=1
 parameter F test: F=13.3271 , p=0.0003 , df_denom=519, df_num=1


 Granger Causality
 number of lags (no zero) 2
 ssr based F test: F=13.4540 , p=0.0000 , df_denom=516, df_num=2
 ssr based chi2 test: chi2=27.1687 , p=0.0000 , df=2
 likelihood ratio test: chi2=26.4840 , p=0.0000 , df=2
 parameter F test: F=13.4540 , p=0.0000 , df_denom=516, df_num=2


 Granger Causality
 number of lags (no zero) 3
 ssr based F test: F=8.9401 , p=0.0000 , df_denom=513, df_num=3
 ssr based chi2 test: chi2=27.1863 , p=0.0000 , df=3
 likelihood ratio test: chi2=26.4994 , p=0.0000 , df=3
 parameter F test: F=8.9401 , p=0.0000 , df_denom=513, df_num=3


 Granger Causality
 number of lags (no zero) 4
 ssr based F test: F=6.7321 , p=0.0000 , df_denom=510, df_num=4
 ssr based chi2 test: chi2=27.4037 , p=0.0000 , df=4
 likelihood ratio test: chi2=26.7047 , p=0.0000 , df=4
 parameter F test: F=6.7321 , p=0.0000 , df_denom=510, df_num=4


 Granger Causality
 number of lags (no zero) 5
 ssr based F test: F=5.8029 , p=0.0000 , df_denom=507, df_num=5
 ssr based chi2 test: chi2=29.6440 , p=0.0000 , df=5
 likelihood ratio test: chi2=28.8268 , p=0.0000 , df=5
 parameter F test: F=5.8029 , p=0.0000 , df_denom=507, df_num=5


 Granger Causality
 number of lags (no zero) 6
 ssr based F test: F=5.0143 , p=0.0001 , df_denom=504, df_num=6
 ssr based chi2 test: chi2=30.8620 , p=0.0000 , df=6
 likelihood ratio test: chi2=29.9760 , p=0.0000 , df=6
 parameter F test: F=5.0143 , p=0.0001 , df_denom=504, df_num=6


 Granger Causality
 number of lags (no zero) 7
 ssr based F test: F=4.3764 , p=0.0001 , df_denom=501, df_num=7
```

```
ssr based chi2 test: chi2=31.5520 , p=0.0000 , df=7
likelihood ratio test: chi2=30.6250 , p=0.0001 , df=7
parameter F test: F=4.3764 , p=0.0001 , df_denom=501, df_num=7


Granger Causality
number of lags (no zero) 8
ssr based F test: F=3.8112 , p=0.0002 , df_denom=498, df_num=8
ssr based chi2 test: chi2=31.5303 , p=0.0001 , df=8
likelihood ratio test: chi2=30.6027 , p=0.0002 , df=8
parameter F test: F=3.8112 , p=0.0002 , df_denom=498, df_num=8


Granger Causality
number of lags (no zero) 9
ssr based F test: F=3.4022 , p=0.0005 , df_denom=495, df_num=9
ssr based chi2 test: chi2=31.7947 , p=0.0002 , df=9
likelihood ratio test: chi2=30.8501 , p=0.0003 , df=9
parameter F test: F=3.4022 , p=0.0005 , df_denom=495, df_num=9


Granger Causality
number of lags (no zero) 10
ssr based F test: F=3.0085 , p=0.0011 , df_denom=492, df_num=10
ssr based chi2 test: chi2=31.3696 , p=0.0005 , df=10
likelihood ratio test: chi2=30.4478 , p=0.0007 , df=10
parameter F test: F=3.0085 , p=0.0011 , df_denom=492, df_num=10


Granger Causality
number of lags (no zero) 11
ssr based F test: F=2.9947 , p=0.0007 , df_denom=489, df_num=11
ssr based chi2 test: chi2=34.4912 , p=0.0003 , df=11
likelihood ratio test: chi2=33.3791 , p=0.0005 , df=11
parameter F test: F=2.9947 , p=0.0007 , df_denom=489, df_num=11


Granger Causality
number of lags (no zero) 12
ssr based F test: F=2.7394 , p=0.0013 , df_denom=486, df_num=12
ssr based chi2 test: chi2=34.5641 , p=0.0005 , df=12
likelihood ratio test: chi2=33.4453 , p=0.0008 , df=12
parameter F test: F=2.7394 , p=0.0013 , df_denom=486, df_num=12


Granger Causality
number of lags (no zero) 13
ssr based F test: F=2.6577 , p=0.0013 , df_denom=483, df_num=13
ssr based chi2 test: chi2=36.4818 , p=0.0005 , df=13
likelihood ratio test: chi2=35.2360 , p=0.0008 , df=13
parameter F test: F=2.6577 , p=0.0013 , df_denom=483, df_num=13
```

```
Granger Causality
number of lags (no zero) 14
ssr based F test: F=2.4680 , p=0.0022 , df_denom=480, df_num=14
ssr based chi2 test: chi2=36.6399 , p=0.0008 , df=14
likelihood ratio test: chi2=35.3812 , p=0.0013 , df=14
parameter F test: F=2.4680 , p=0.0022 , df_denom=480, df_num=14
```

```
Granger Causality number of lags (no zero) 15 ssr based F test: F=2.3488 , p=0.0029 , df_denom=477, df_num=15 ssr based chi2 test: chi2=37.5225 , p=0.0011 , df=15 likelihood ratio test: chi2=36.2014 , p=0.0017 , df=15 parameter F test: F=2.3488 , p=0.0029 , df_denom=477, df_num=15
```

> Lets analyze the results:

Simple notion:

X(t) granger causes Y(t) , if the past values of X(t) helps in predicting the future values of Y(t).

> Y(t) is not just the function of Y(t -1) *lag of 1 but* also the function of X(t - 1)

`number of lags (no zero) [xx]`

Shows the Lags used in finding causility.

`p values of all (`

Both p-values of every lag is same, show both-directional Granger-cause.

`Granger Causality number of lags (no zero) 2 ssr based F test: F=13.4540`

Lag 2 show the highest *F test* value out of all the lags. *F-test* checks that the lagged values of X *jointly* improve the forecast of Y (or vice versa). It can conclude that for predicting Y with two predictors X*1* and X*2* where X*2* is same X1 with small noised values.

Few Quick Points

- Testing for Granger-causality using F-statistics when one or both time series are non-stationary can lead to false causality. If both the time series are stationary then differencing, de-trending or other techniques must first be employed before using the Granger Causality test.
- F statistic can be used only when variables are cointegrated.
- We say that Granger-causes y when the is rejected.
- The for the test is that lagged X-values do not explain the variation in Y. Put simple, it assumes that X(t) doesn't Granger-cause Y(t).
- The Toda–Yamamoto approach is used to test causality for co-integrated time-series data.

- Cointegration analysis is a useful tool in order to examine if there exists a long run equilibrium relationship between two or more time series. ()
- Causation is can be One-Direction, Both-Direction or NO-Direction