Granger Causality - A possible Feature Selection Method in Marketing Mix Modeling (MMM)

Venkatraman R.
venkat@arymalabs.com
Aryma Labs Pvt. Ltd

Ridhima Kumar
ridhima@arymalabs.com
Aryma Labs Pvt. Ltd

Tannishtha Sen
tannishtha@arymalabs.com
Aryma Labs Pvt. Ltd

Abstract

Feature Selection (aka covariate/independent variable selection) in any regression problem is a crucial task. Incorporating features in the model that best predicts or explains the dependent variable is both an art and science. There are many feature selection methods.

In MMM, incorporating the right features in the model is a very tricky proposition. Traditionally, in MMM, feature selection is done using a combination of correlation approaches and domain informed judgments about the variables. However, this is not a robust approach since this can lead to Bias in the model.

As a workaround, we have been constantly experimenting with other feature selection methods. In this paper, we propose that Granger Causality could be a potential feature selection method. We illustrate how the Granger Causality (if carefully used) can aide in feature selection in an MMM setup.

Keywords: Marketing Mix Modelling, Granger Causality, Econometrics, Causal Inference

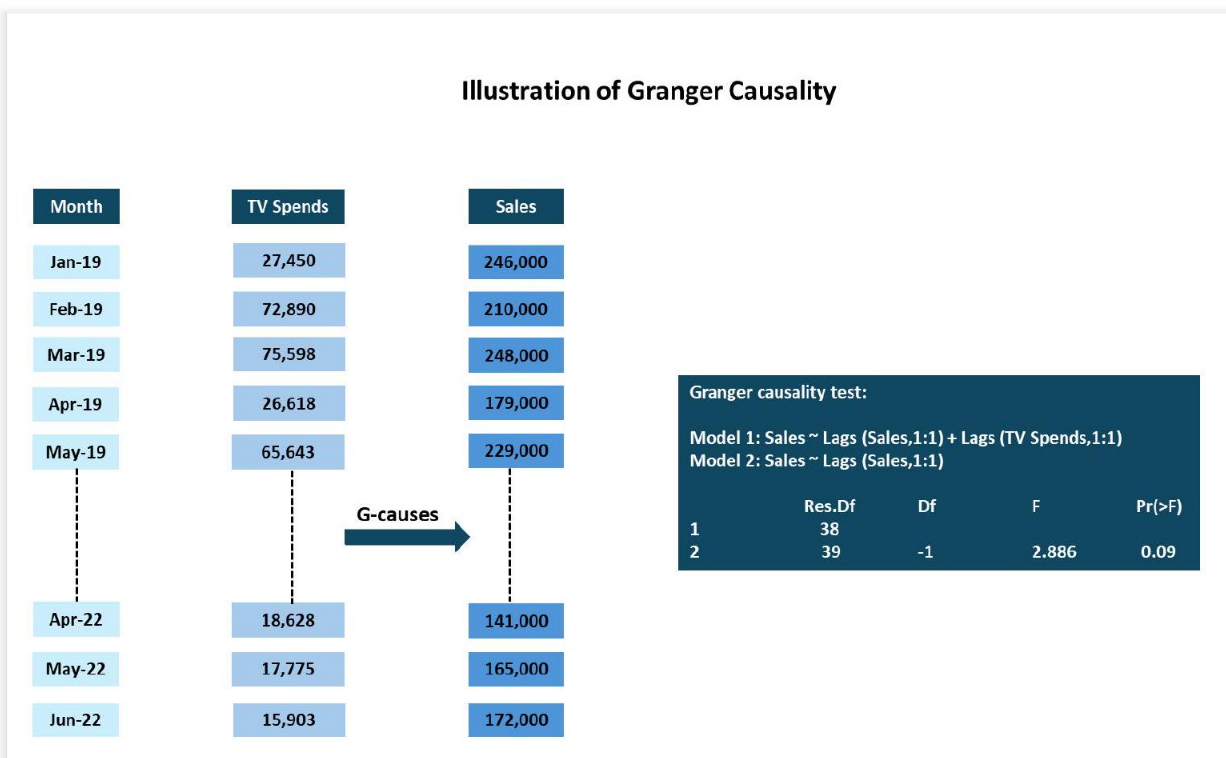

Figure 1: Illustration of Granger Causality

1. Introduction to Granger Causality

Granger causality test is a statistical hypothesis test for determining that whether one time series is useful in forecasting another time series. Granger causality was first proposed by the Nobel award-winning economist Clive W. J. Granger in [1] in the 1960.
Clive Granger proposed that causality in Economics can be tested by measuring the ability of predicting the future value of a time series with the help of past values of another time series.
In simple words, Granger Causality can be described as follows:
Suppose we have two time series X and Y. Then X is said to "Granger cause" Y, if predictions of Y based on the past values of X and past values of Y are better than the predictions of Y based only on the past values of Y.
Therefore, we see if X contains some useful information for the prediction of Y which is not present in the past values of Y.
Clive Granger's explanation [1] of Granger Causality is as follows:
"Suppose that we have three terms, $\mathrm{X_t}, \mathrm{Y_t}$, and $\mathrm{W_t}$, and that we first attempt to forecast $\mathrm{X_{t+1}}$ using past terms of $\mathrm{X_t}$ and $\mathrm{W_t}$. We then try to forecast $\mathrm{X_{t+1}}$ using past terms of $\mathrm{X_t}, \mathrm{Y_t}$, and $\mathrm{W_t}$. If the second forecast is found to be more successful, according to standard cost functions, then the past of Y appears to contain information helping in forecasting $mathrmX_{t+1}$ that is not in past $\mathrm{X_t}$ or Wt. In particular, $\mathrm{W_t}$ could be a vector of possible explanatory variables. Thus, $\mathrm{Y_t}$ would "Granger cause" $\mathrm{X_{t+1}}$ if (a) $\mathrm{Y_t}$ occurs before $\mathrm{X_{t+1}}$; and (b) it contains information useful in forecasting $\mathrm{X_{t+1}}$ that is not found in a group of other appropriate variables".

2. Granger Causality $\neq$ Causality

The word 'Causality' in Granger Causality is actually a misnomer. Granger causality tests whether X forecasts Y rather than X causes Y. Granger causality is better described as "precedence" or as Clive Granger himself claimed "temporally related". Precedence alone is not a sufficient condition for causality. Many Statisticians remark that Granger Causality suffers from post hoc fallacy [7] and hence it can't be deemed a causal framework. Anyhow, the fact that Granger Causality is not a full-fledged casual framework does not diminish its utility in cases where it is used to check the predictability/forcastability of one time series with respect to another.

3. Methodology of Granger Causality

Suppose we have Sales and marketing spends as the two time-series. We want to test whether the marketing spends Granger-causes (G-causes) Sales. We then design the following experiment.
We consider two models to determine the impact of past sales and marketing spends on the prediction of current month's sales:

Model 1: A model that determines how well the past sales data predict the current month's sales.

Model 2: A model that utilizes both prior sales data and prior marketing spends data to predict current month's sales.

If a statistical test comparing these two models indicates that the inclusion of marketing spends provides a better fit, we can infer that marketing spends Granger-cause Sales. The concept of Granger causality suggests that a variable X Granger causes Y if it contains statistically significant information about future values of Y.
The models are represented as follows:

- **Reduced Model:**

$$\text{Sales}_t = \theta_0 + \sum_{i=1}^{p} \alpha_i \cdot \text{Sales}_{t-i} + \varepsilon_t$$

- **Full Model:**

$$\text{Sales}_t = \theta_0 + \sum_{i=1}^{p} \alpha_i \cdot \text{Sales}_{t-i} + \sum_{i=1}^{p} \beta_i \cdot \text{media spend}_{t-i} + u_t$$

The hypothesis for our test is defined as:

- H_0: Past marketing spends do not Granger cause sales.
- H_1: Past marketing spends Granger cause sales.

The Granger-F statistic is computed as follows:

$$\text{Granger F-Stat} = \frac{\frac{RSS_0 - RSS_1}{p}}{\frac{RSS_1}{t-2p-1}} \sim F_{p,t-2p-1}$$

where RSS_0 is the residual sum of squares of the reduced model and RSS_1 is the residual sum of squares of the full model. If the F-statistic is significant at the relevant significance level, we conclude that past marketing spends "G-causes" Sales.
Note: The appropriate level of lag could be specified through an information theoretic approach like AIC.

4. Novel Approach: Granger Causality as a feature selection method

We propose that Granger Causality can be used as a feature selection method in MMM. In a typical MMM framework, we usually have the KPI as Sales/Leads and the different independent variables are marketing spends and other micro and macroeconomic variables. Thus, it will allow us to answer questions about the influence of Display ads on search e.g. in [2], quantifying the impact of traditional marketing and online consumer activity e.g. in [4] and quantify the effects of Word of Mouth and Traditional Marketing e.g. in [6].
Through, Granger causality, we can identify whether a time series is helping to forecast another time series. Let us consider two time series in an MMM setup such as TV Spends and Sales. From Granger causality test, we will be able to identify whether the past value of TV Spends contains information that will help to forecast Sales than just the past values of Sales alone.
Lastly, we remark that our approach is along the lines of the appraoches in [6] and [5], which also try to utilise Granger Causality for variable selection.
Figure 2 is an illustration of a Granger causality test, which indicates that TV Spends "Granger-causes" Sales at 10% level of significance.
So, this test can be applied by taking all the available independent variables one by one. Based on the test results, we can select the variables that indicate significant Granger causality between the variable and Sales. Thus, the variables so selected could be incorporated in the model and the relevant goodness of fit checked.

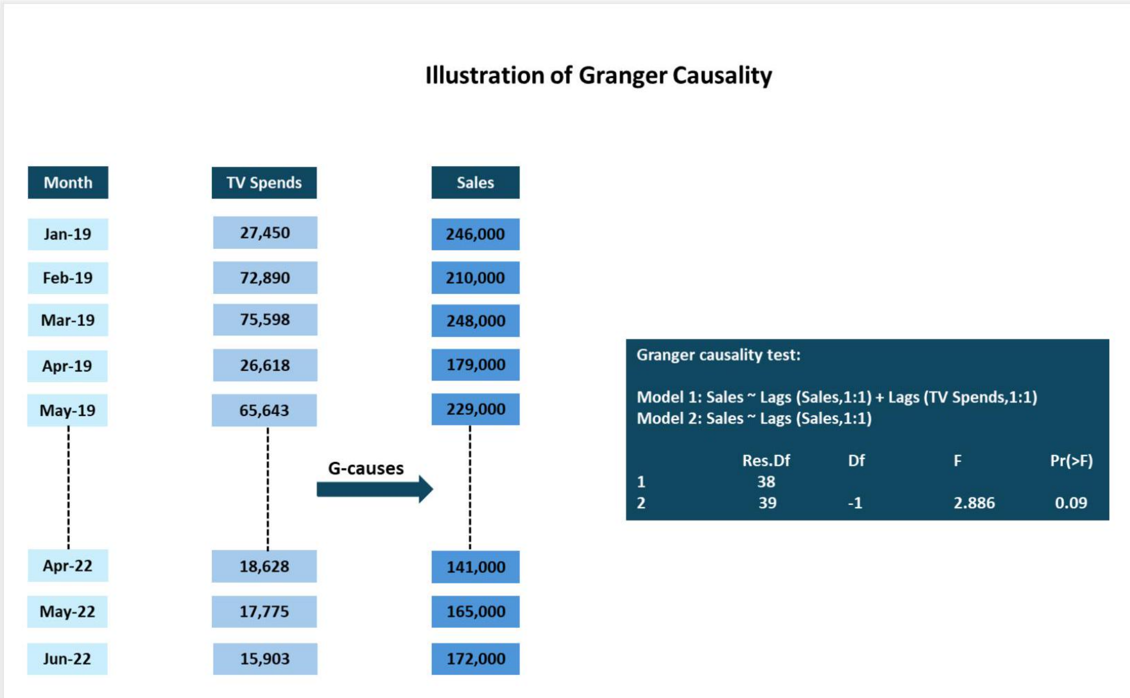

Figure 2: Illustration of Granger Causality

5. Approach in detail:

We tested using Granger causality as feature selection method post hoc, i.e., the MMM models were already specified with ’correct’ variables. We chose 10 MMM models already built and delivered to our clients, the independent variables in these models were chosen through our proprietary methods. The KPI in each of the MMM was Sales and the different independent variables included TV spends, digital spends, competition spends and impact of holiday etc.
Granger causality tests were carried out by taking each independent variables from the MMM models, and it was tested whether these variables helped in forecasting the dependent variable, which is Sales.
Performing the Granger Causality test for all the models, it was observed that the overlap between the variables in the MMM model and the variables that came statistically significant in the Granger Causality test were decently high.
The variable overlap percentages of the 10 models are shown in Table 1.
Figure 3 is an illustration of a Model 1 which has 12 variables. From the Granger causality tests, it was observed that 8 variables were ’Granger causing’ Sales. So, there is an overlap of 67% in the variables obtained through MMM model and Granger Causality tests.
As, the overlap % of variables is decently high, we can possibly utilise Granger Causality as a feature selection methodology.

6. Statistical Concerns:

6.1. Stationarity:

The application of Granger causality assumes that the time series under study are stationary. Non-stationary series can be utilised by using a windowing technique which assumes that sufficiently

Model	Accuracy
Model 1	67%
Model 2	54%
Model 3	89%
Model 4	80%
Model 5	75%
Model 6	86%
Model 7	56%
Model 8	83%
Model 9	57%
Model 10	64%

Table 1: Comparison of model accuracies

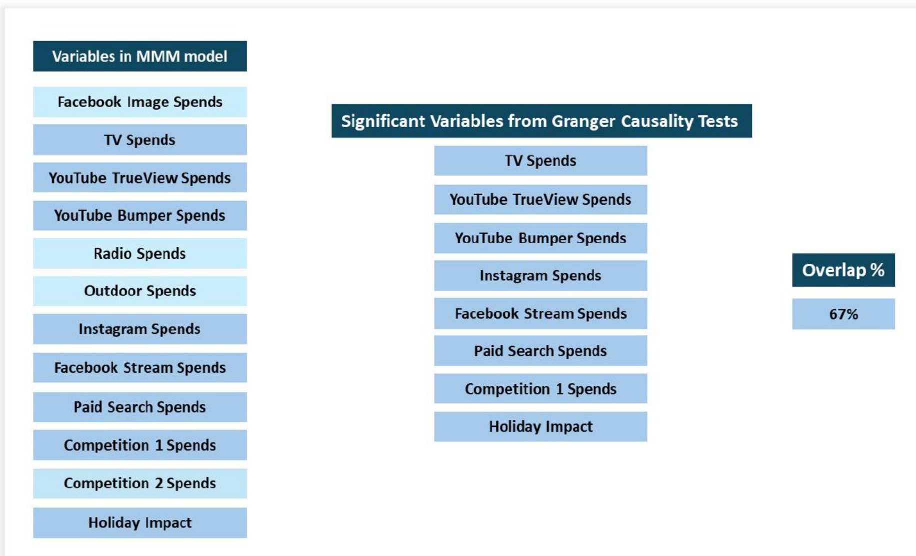

Figure 3: Selection of Variables through Granger Causality

small windows of a non-stationary series are locally stationary.

6.2. Linearity:

The original formulation of G-causality can only give information about linear features of signals. Extensions to nonlinear cases now exist, however these extensions can be more difficult to use in practice and their statistical properties are less well understood.

6.3. Multiplicity Issue:

Because we would carry out a series of tests, there is high chance of inflation of type 1 error. This problem becomes more prevalent in rolling window Granger Causality tests. However, there are some remedies to this mentioned in [3].

8. Note of Caution:

Through this approach, we are only proposing a possible feature selection method. We are not advocating that Granger Causality be a de facto feature selection method in MMM. Feature selection in MMM is very nuanced, one in which domain knowledge plays a crucial role. Further facets like carry over and diminishing returns need to be considered. In the above study, the Granger causal variables were compared with the transformed media variables. Overall, we merely want to aid or supplement existing feature selection methods. We believe that Granger Causality could be a possible feature selection method.

7. Future Scope:

In this paper, we only tested Granger Causality post hoc. It remains to be seen, if the high overlap could be replicated in a comparative study where; the Granger causality tests are applied to all the independent variables and the relevant significant variables chosen vs the independent variables chosen through conventional ways.

References

[1] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society, pages 424–438, 1969.

[2] Pavel Kireyev, Koen Pauwels, and Sunil Gupta. Do display ads influence search? attribution and dynamics in online advertising. International Journal of Research in Marketing, 33(3):475–490, 2016.

[3] Xing Ling. Additive P-Value Combination Test. PhD thesis, Michigan Technological University, 2023.

[4] Shuba Srinivasan, Oliver J Rutz, and Koen Pauwels. Paths to and off purchase: quantifying the impact of traditional marketing and online consumer activity. Journal of the Academy of Marketing Science, 44:440–453, 2016.

[5] Youqiang Sun, Jiuyong Li, Jixue Liu, Christopher Chow, Bingyu Sun, and Rujing Wang. Using causal discovery for feature selection in multivariate numerical time series. Machine Learning, 101:377–395, 2015.

[6] Michael Trusov, Randolph E Bucklin, and Koen Pauwels. Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. Journal of marketing, 73(5):90–102, 2009.

[7] Wikipedia contributors. Post hoc ergo propter hoc — Wikipedia, the free encyclopedia, 2024. [Online; accessed 6-May-2024].