Causfinder: An R package for Systemwise Analysis of Conditional and Partial Granger Causalities

Erdogan CEVHER
Department of International Organizations and Bilateral Relations
The Ministry of Science, Industry and Technology
Ankara, Turkey
erdogancevher@gmail.com

Abstract—Multivariate Granger causality analysis is the study where at least one of the sets of independent and dependent variables includes more than 1 variable when these variables are conditioned on third set of variables in the analyzed system. causfinder reveals – via systemwise approaching the G-causalities – all conditional and partial Granger causalities in the system in any desired pattern formed by various distributions of these variables to these three sets. It also reveals the character of variables from more independent variables to dependent ones with the degrees and directions of the G-causality relations.

Keywords— R, Conditional Granger causality, partial Granger causality, bootstrapping.

I. CONCEPTS, LITERATURE AND DEFINITIONS

Granger causality (shortened as “G-causality” or “GC”) concept has a long history and goes back to Wiener’s 1956 study [1]. After Granger’s formalization in 1969, its usage has spread unbelievably. Yes, one of the methods that whether a time series can be used in estimating another time series is to test “G-noncausality” among these time series (in many contexts, “Granger...” was frequently abbreviated as “G-”). Here, we note that it is standard to name the statistical tests based on H_0 null hypothesis, and therefore we added the “non-” prefix in entitling the test above. Normally, regressions reflect “only” correlations, but, the causality relations among variables can be found with specific tests as well. Let x and y be two time series. A time series x is said to “Granger causes” (shortly, G-causes) y if y can be predicted better using the past values of both y and x if compared with its prediction with only the past values of y [2]. That is to say, the causal influence of one time series on another can be conceived by the notion that the prediction of one time series is improved by incorporating knowledge about the other [3]. After whether x G-causes y is tested, if this causality relationship turns out to be true, then the relationship is shown as x → y, and if it turns out to be false (i.e., x does not G-cause y), the relationship is shown as x ↛ y.

Let’s think the claim that “the increases in the parents’ expenses for the education of their children during their lives (private teaching institutions, college tuitions, etc.) cause their children to have more successful results in their education life”. The correlation of education expenses to the education result of their children is positive, i.e., the children of the parents who are spending more on education obtain more successful results. Given any time moment, when the effects of confounding variables (revenue etc.) are controlled, if every sudden change in explanatory variable (education expenses) causes to the increases (when compared to the moments with no sudden change) in the result variable (success in education) just after the sudden changes in accordance with these sudden changes of explanatory variable, then the explanatory variable (education expenses) is the G-cause of the result variable (success in education).

In essence, since the Granger causality test does not exclude the “post hoc ergo propter hoc” (the event A occurred, then the event B occurred; so, the cause of B is A) fallacy, it is not a suitable test in the strictest sense (in the context of mathematical abstract logic) to test the causality relations among variables. This is true for every tests in econometrics that is entitled as pseudo “causality test” [4]. That said, however, “causality tests” in econometry reflect quite well the existing realities of the life, and that’s why, they are frequently used in many different fields such as economy, finance, natural science and medicine.

It is better not to dwell on classical Granger causality due to the existence of a giant literature on it. Since the aim of this paper is essentially just to introduce an R package for the practical usage of advanced and modern G-causality concepts, we will cover the detailed theoretical topics very superficially.

There are many cases in analyzing multivariate systems:

- In a 2-variable system where both of the variables are stationary, classical GC test [2] is achieved.

- In a 2-variable system, if the variables are cointegrated (in this case, both are nonstationary), a vector autoregressive (VAR) model is not formed by differencing the variables (in this case, a vector error correction (VEC) model is formed) and G-noncausality tests are not performed via t/F tests (classical G-noncausality test) [5].

- In a 2-variable system, when at least one of the variables is nonstationary, the G-causalities among variables is examined by Toda-Yamamoto G-noncausality...
test [6] since testing G-causality using F statistics (classical G-noncausality test) may result in spurious G-causality [7].

Since there is a large literature about analysis of G-causality among variables in 2-variable system, we omit this part for the time being, and pass to the systems where there are more than 2 variables. In systems with “>2” variables, since interaction and confounding effects completely change the nature of G-causality, such systems must be treated carefully. The standard framework for GC has been recently extended to the multivariate case, where predictor and dependent variables are no longer constrained to be univariate [8].

The first handling of advanced Granger causality (shortened as “G-causality” or “GC”) analysis in R was performed by Roelstraete and Rosseel: the authors presented their work as the paper “FIAR: An R Package for Analyzing Functional Integration in the Brain” [9]. By “advanced” we mean the advanced and new concepts and definitions of Granger causality (such as “conditional G-causality”, “partial G-causality”, “conditional difference G-causality”, “partial difference G-causality”, “canonical G-causality”, “harmonic G-causality”, “global G-causality”, “componental G-causality” etc.) that are beyond Granger’s 1969 definition of G-causality [10]. For a small theoretical background of advanced GC, we refer the reader to the above paper.

In order to understand advanced concepts of G-causality, a new terminology to the already used definitions and concepts is necessary. There are mainly two important differences in usage of terms between any classical G-causality analysis and advanced and modern G-causality analysis: “multivariate” and “pairwise”. First, this distinction will be thoroughly revealed:

A. Classical Granger Causality Analysis

In a classical G-causality analysis, “multivariate” means there are more than 2 variables in the system or model in discussion for which G-causalities will be found. “Pairwise” means the G-causality analysis is performed by taking the variables in pairs (the number of the both of independent and dependent variable is just 1). For example, if the system’s variables are “x, y, z, w”, then “whether x G-causes y or not”, “whether x G-causes z or not”, etc. are analyzed without taking the effects of all the remaining variables in the system into account. Definitely, this latter is in fact not a G-causality: even though Granger first defined G-causality as the causality between 2 variables, he also emphasized that for finding G-causalities in systems with more than 2 variables, “all of the information in the system must be used for a G-causality analysis”. In search of “whether x G-causes y or not” that is isolated from the other variables, the effects of the variables of z and w are omitted and thereby some of the information in the system is not used. Hence, in essence, one can say that the pairwise G-causality of classical G-causality is in fact not a causality in the Granger sense!

B. Modern (Advanced) Granger Causality Analysis

In advanced G-causality analysis, “multivariate” means there are more than 1 variables in either of the “set of causers” (like-independents) and “set of caused by the causers” (like-dependents) in search of G-causality. Remember that in any Granger causality analysis, the variables are not pre-determined as independents and dependents; G-causality analysis reveals which ones are independent and which ones are dependent. For example, assume that there are 5 variables (x, y, z, q, w) in the system or model in which we search G-causalities among variables. Then, the (searched) G-causalities

- “x and y G-causes z conditional on q and w”, (pattern: 5,2,1)
- “x G-causes y and z conditional on q and w”, (pattern: 5,1,2)
- “x and y G-causes z and q conditional on w”, (pattern: 5,2,2)
- “x, y and z G-causes q and w”, (pattern: 5,3,2)
- “x and y G-causes z, q and w” (pattern: 5,2,3)

... are all “multivariate” G-causalities in an advanced G-causality analysis. Notice pattern is “the number of all variables in the system, the number of variables in the causers set, the number of variables in the caused-by-the-causers set”. “Pairwise” means there are only 1 variable in both “set of causers” and “set of caused by the causers”. Hence, continuing with the above example, the (searched) G-causality of “x and y G-causes z conditional on q and w” is not a pairwise G-causality. On the other hand, “x G-causes y conditional on z, q and w”, “q G-causes z conditional on x, y and w”, ... are all “pairwise” G-causalities.

It would be beneficial to mention the reflections of advanced G-causality analysis in R in parallel with the other software platforms. There have been some handlings of advanced G-causality concepts in MATLAB. Two important MATLAB packages that is worth mentioning are MVGC (Multivariate Granger Causality) and GrangerCausalityGUI.

GrangerCausalityGUI package is the resultant work of Jianfeng Feng group that was developed in accompany of some papers through 2008-2013 [11]. Seth’s GCCA (Granger Causal Connectivity Analysis) appeared in 2009 [12]. Various studies were conducted by using GCCA. In 2011, Roelstraete and Rosseel wrote first R package that handles advanced G-causality analysis [13]. Their work revealed a bug in GCCA and contributed further to the studies in the field [14]. The feedback and the experience obtained via GCCA resulted in MVGC in May 2014 [15]. That is to say, MVGC is the new version of GCCA. Definitely, MVGC brought a new breath to advanced G-causality analysis.

causfinder is an R package that systematically examines advanced G-causalities in a general-to-specific way as was done in GrangerCausalityGUI, GCCA and MVGC. That is to say, causfinder systematizes FIAR package of Roelstraete and Rosseel. To concretize, assume we have a system of 6 variables (x, y, z, q, t, w). To find whether “x,y” (conditional, partial, etc.) G-causes “z,q,t” (conditional on w”, in FIAR, one places the variables in suitable order, and FIAR reveals the result. In causfinder, the advanced G-causality analysis is performed in
systemwise manner. Hence, the variables are assigned to the ordered numbers (from 1 to 6 in this example); every number shows a different variable; and this assignment is kept throughout the analysis. To find a specific G-causality, one only needs to give the relevant pattern. Since there are 6 variables in the system, 2 variables in the causers (independents), 3 variables in the caused-by-the-causers set, the pattern is (6, 2, 3). causfinder lists all the G-causals in this pattern. That is to say, it reveals whether “x,y” (conditional, partial, etc.) G-causes “z,q,t” conditional on w”, but besides this, it also reveals “x,z” (conditional, partial, etc.) G-causes “y,q,t” conditional on w”, “x,q” (conditional, partial, etc.) G-causes “y,z,t” conditional on w”, “x,w” (conditional, partial, etc.) G-causes “z,q,t” conditional on y”, etc. causfinder also plots all these G-causals. Hence, one can easily see which one(s) of the variables in the system can be taken as independent and which one(s) of the variables in the system as dependent in a certain pattern. When pairwise G-causals (in the sense of advance G-causality analysis) are used, causfinder may list all of the variables in the system from more independents to dependents one.

To reveal the robustness of causfinder, firstly, we found it necessary to repeat the analysis in FIAR of Roelstraete and Rosseel and get the same result they obtained by using the same dataset. In presenting their papers, the then-version of FIAR was 0.3. Later, FIAR 0.5 appeared. The grangerdata dataset in FIAR 0.3 and FIAR 0.5 is different. Since the article of Roelstraete and Rosseel was performed via FIAR 0.3, and FIAR reached to 0.5 version eventually, we exported grangerdata of FIAR 0.3, and we loaded FIAR 0.5, and retrieved the exported grangerdata, and made the relevant analysis in FIAR 0.5 (latest version of FIAR). We observed that the functions used in FIAR 0.3 are more robust and extendable for further analysis. Then we employed causfinder to find the same results in FIAR 0.3, but this time from top-to-bottom analysis. We renamed grangerdata of FIAR 0.3 paper as granger.df in causfinder (.df” extension is to denote data frame; we used this notation in all the other datasets in causfinder).

II. ROBUSTNESS CONTROL

Here is the analysis performed via causfinder (reader is suggested to glimpse at Roelstraete and Roseel’s paper):

A. grangerdata of FIAR 0.3

\[
\begin{align*}
R & \textbf{> library(causfinder)} \\
R & \textbf{head(granger.df)}
\end{align*}
\]

\[
\begin{array}{cccccc}
x & y & z & q & w \\
[1,] & 3.75092 & 0.26922 & 1.2932 & 0.31352 & -0.288403 \\
[6,] & -6.82409 & 1.65985 & -3.1487 & -5.22565 & 0.016327
\end{array}
\]

\[
\begin{align*}
R & \textbf{> AOrderG(granger.df, max = 10)} \\
[1] & 3
\end{align*}
\]

The above optimal minimal VAR lag order can be checked with bird’s eye view via VARomlop (vector autoregressive optimal minimal lag order plotter):

\[
\begin{align*}
R & \textbf{> VARomlop(granger.df,10)}
\end{align*}
\]

Though SBC is generally preferred in VAR order selection processes, we continue with AIC to reflect the analysis in FIAR. Bird eye’s view shows optimal minimal lag order of 3 is suitable for AIC-sticker.

When the variables are put to a data frame in the order x, y, z, q, w, the conditional G-causality from x to y and z conditional on q and w is (the variable order in the data frame is important both for FIAR and causfinder):

\[
\begin{align*}
R & \textbf{> conditionalGb(granger.df, nx=1,ny=2,order=3)} \\
$orig & [1] 0.6457208 \\
$prob & [1] 0
\end{align*}
\]

\[
\begin{align*}
R & \textbf{> conditionalGb(granger.df, nx=1,ny=2,order=3, boot=TRUE, blol=70)} \\
& \text{STATIONARY BOOTSTRAP FOR TIME SERIES} \\
& \text{Average Block Length of 70} \\
& \text{Call:} \\
& \text{tsboot(tseries = data, statistic = conditionalGranger, R = bs, l = blocksize specified, sim = "geom", nx = m ndependents, ny = m in dependents, order = m orderspecified) Bootstrap Statistics :}
\end{align*}
\]

\[
\begin{align*}
R & \text{original bias std. error} \\
& t1* 0.6457208 \quad -0.4498475 \quad 0.04592638
\end{align*}
\]

The distribution of \(H_0\) in bootstrapping process of the above multivariate conditional Granger causality can be plotted:

\[
\begin{align*}
R & \textbf{> bootstrapplot(granger.df,5,1,2,order=3, nrawobs=2000)}
\end{align*}
\]
In order to use the summary of functions (conditionalGgFp, conditionalGgFp, partialGgFp, partialGgFp) of causfinder, first let’s find a value of a parameter (“maxoi”) that will be needed. maxoi is the maximum number of the order of integration that makes all the nonstationary variables (if any) stationary ones. Since all of the variables in a system can be stationary, the default value of maxoi is 0 in functions of causfinder. maxoi can be determined via adfcs by applying to each variable in the system. As is known in all of the lag selection procedures in econometrics, same (common) sub-sample must be used to determine the correct optimal minimum lag. adfcs calculates optimal minimum lag of autoregressive processes by using this fact. Here is the stationarity analysis of variables in granger.df:

```r
R> adfcs(granger.df[,1])@test
....
R> adfcs(granger.df[,5])@test
```

The results of augmented Dickey-Fuller (ADF) tests via adfcs* show all of the variables in granger.df (i.e., x, y, z, q, w) are stationary. Hence, maxoi is 0. By the way, as a rule of R, if max parameter is not supplied to adfcs and type of the AR regression is wanted to be entered to adfcs, then parameter name must be written in order to prevent wrong assignment of type to max parameter in adfcs (max is the 2nd parameter, type is the 3rd parameter in the function).

The above value of “x conditional G-causes to y and z conditional on q and w” is only one of the G-causalities in the pattern (5,1,2) (number of variables, number of causers (like-independents), number of caused by the causers (like-dependents)). By the way, since p-value is 0 (<0.05), this conditional G-causality value is significant (i.e. there is a conditional G-causality in the specified direction). There are however dim(combn(5,1))[[2]] * dim(combn(4,2))[[2]] = 30 G-causalities in (5,1,2) pattern. Let’s assign each variable to a number from 1 to the number of variables in the system; and then protect this assignment along the analysis: 1=x, 2=y, 3=z, 4=w, 5=q. The 30 G-causalities (that we will search) of the pattern (5,1,2) can be found via gctemplate:

```r
R> gctemplate(5,1,2)
```

For example, “5 2 4 1 3” in 29th line is used for searching “q (conditional, partial etc.) G-causes y and w conditional on x and z”.

The conditional G-causality (“x G-causes y and z conditional on q and w”) that was calculated above can be found as a single item of the systemwise G-causality analysis via conditionalGblup and conditionalGgFp. conditionalGblup gives conditional G-causalities with bootstrapping, lower and upper bounds for the bootstrapping statistics and the plot of the results. conditionalGgFp computes all conditional G-causalities of a multivariate time series in a certain pattern (here; (5,1,2) will be used), graphs them, and tabulates the F statistics and p values. Here is the same result from systemwise (top-to-bottom) analysis: (it took less than 1.5 minute: the number of bootstrapping samples is left in its default value “bs=100”; all the 30 conditional G-causalities are calculated; with fly-on-the-air variables, R holds more than 30*100*h variables!):

```r
R> conditionalGblup(granger.df,1,2,3,2000,0,blol=70)
```

The rounded value 0.65 (of 0.6457208) is printed in the graph for visual appearance. Since 0.65 is outside of the lower and upper limits (0.11 and 0.3 respectively), the existence of strong (blue is far from red-band) “G-causality from x to y and z conditional on q and w” is deduced from systemwise analysis as well. Looking at the graph, one can understand the character of the variables in the system; i.e., which one(s) are more independent if compared with the other(s).

The existence of strong conditional G-causality from x to y and z conditional on q and w can be obtained via conditionalGgFp as well:

```r
R> conditionalGgFp(granger.df,niindependents=1,niindependents=2,nobsraw=2000,maxoi=0,order=3)
```
The significance level of alpha=0.10 is drawn to give an idea in the following graph. One can adopt any desired significance level. As is seen from figure, the graphs of F and p values behave oppositely, when one increases the other decreases, and so for the decreasing case. High F values (equivalently low p values) indicate the existence of strong G-causalities. Since three graphs/tables will be presented, user is expected to click Graphics Device during the execution of the command.

Figure 4. Systemwise plot of conditional Granger causalities with F and p values

| TABLE I. SYSTEMWISE MATRIX OF CONDITIONAL GRANGER CAUSALITIES WITH F STATISTICS |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 0.646 | 0.003 | 0.002 | 0.001 | 0.002 |
| 0.963 | 0.002 | 0.002 | 0.002 | 0.003 |
| 0.491 | 0.004 | 0.002 | 0.117 | 0.090 |
| 0.174 | 0.001 | 0.002 | 0.002 | 0.003 |
| 0.104 | 0.004 | 0.002 | 0.141 | 0.097 |
| 0.291 | 0.004 | 0.002 | 0.002 | 0.007 | 0.067 |

The direction of the G-causalities is from columns to rows, and conforms to the pattern given by gctemplate. For example, the F values of conditional G-causality “from x to y and z conditional on q and w” and “from x to y and q conditional on z and w” are 0.646 and 0.963 respectively.

| TABLE II. SYSTEMWISE MATRIX OF CONDITIONAL GRANGER CAUSALITIES WITH P VALUES |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 0 | 0.477 | 0.468 | 0.491 | 0.483 |
| 0 | 0.485 | 0.482 | 0.463 | 0.469 |
| 0 | 0.494 | 0.483 | 0.004 | 0.022 |
| 0 | 0.459 | 0.482 | 0.485 | 0.474 |
| 0 | 0.01 | 0.461 | 0.464 | 0.001 | 0.015 |
| 0 | 0.465 | 0.475 | 0.017 | 0.017 | 0.069 |

p-values of conditional G-causalities are read from columns to rows as is in reading of F values.

The plots and tables of conditionalGblup and conditionalGgFP make it easy to find the order of variables from independents to dependents. Anyway, there is another function in causfinder, namely, gcintensity. From matrix4 slot of conditionalGblup, gcintensity automatically calculates the degree of independence of the variables in the system. “4” in matrix4 is to denote that 4 columns are presented in conditionalGblup to derive the related plot. Here is the way to list the variables in the system from more independents to dependents ones in the (5,1,2) pattern:

R> gcintensity(conditionalGblup(granger.df,1,2,3,2000,0,block=70)$matrix4, 5,1,2)

$GCmatrix
[1,] 0.3440522154
[2,] 0.5227375551
[30,] 0.0162249097

$GCIntOfVariables
[1,] 0.98415356
[2,] 0.07148067
[3,] 0.01336220
[4,] 0.17616651
[5,] 0.05852829

From gcintensity output, the order of variables from more independents to dependents (and the degrees of independences) is found: x, q, y, w, z. Here, we note that the type of patterns like (5,1,2), (5,1,3), (5,2,2), etc. are useful especially in neuroscience when the functional integration in the brain is analyzed. But when we turned back to economics, finance, etc. we guess that the G-causalities of the pairwise GC patterns, i.e. (....,1,1) patterns that causfinder reveals will be more used by the researchers.
B. The Determinants of Foreign Direct Investment (FDI) of Turkey

Till now, we covered causfinder and showed its robustness from the grangerdata dataset of FIAR. We also prepared another dataset; the determinant of “foreign direct investments (FDI) in Turkey”. The 6 variables in this dataset are: trade openness (aciklik), political stability (istikrar), exchange rate (kur), lnFDI (lnDYSY), lnGDP (lnGSYIH), number of annual flights (ucus). This dataset has two versions:

First Version: V6Nonstationary43ObsOL.df:

- 6 variables, raw data (both nonstationary and stationary variables may exist), 43 observations (1970-2012) with the added observation labels to the data frame:

R> head(V6Nonstationary43ObsOL.df)

<table>
<thead>
<tr>
<th>OBS</th>
<th>ACIKLIK</th>
<th>ISTIKRAR</th>
<th>KUR</th>
<th>LNDSY</th>
<th>LNGSYIH</th>
<th>UCUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>0.06690</td>
<td>0.0100</td>
<td>0.46044</td>
<td>0.01415</td>
<td>0.07525</td>
<td></td>
</tr>
<tr>
<td>1971</td>
<td>0.09002</td>
<td>0.0070</td>
<td>0.38066</td>
<td>0.10943</td>
<td>0.09031</td>
<td></td>
</tr>
<tr>
<td>1972</td>
<td>0.07601</td>
<td>0.0090</td>
<td>0.37612</td>
<td>0.10397</td>
<td>0.11146</td>
<td></td>
</tr>
<tr>
<td>1973</td>
<td>0.08799</td>
<td>0.0065</td>
<td>0.46394</td>
<td>0.10563</td>
<td>0.13501</td>
<td></td>
</tr>
<tr>
<td>1974</td>
<td>0.10897</td>
<td>0.0190</td>
<td>0.41588</td>
<td>0.10793</td>
<td>0.11806</td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td>0.09463</td>
<td>0.0190</td>
<td>0.47362</td>
<td>0.11082</td>
<td>0.11899</td>
<td></td>
</tr>
</tbody>
</table>

From this data set, we obtain another one where all variables are stationary. Since the order of integration of variables in V6Nonstationary43ObsOL.df is all one (found via adfcs; see table below), we take first differences for the 6 variables.

**TABLE III.** ADF STATISTICS OF THE RAW VARIABLES

<table>
<thead>
<tr>
<th>Variable</th>
<th>Level (t variate)</th>
<th>c</th>
<th>Prob. Value</th>
<th>m</th>
<th>t</th>
<th>1st Differ (t variate)</th>
<th>c</th>
<th>Prob. Value</th>
<th>m</th>
<th>t</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>aciklik</td>
<td>0.06690</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>istikrar</td>
<td>0.06690</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>kur</td>
<td>0.06690</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>lnDYSY</td>
<td>0.06690</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>lnGSYIH</td>
<td>0.06690</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>ucus</td>
<td>0.06690</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Note: Variables: trade openness (aciklik), political stability (istikrar), exchange rate (kur), lnFDI (lnDYSY), lnGDP (lnGSYIH), number of annual flights (ucus). In ADF regression, the results were given in the order of “both drift and time trend”, “drift without time trend” and “no drift, no time trend”. Whether the drift and time trend coefficients in the ADF regressions is significant was specified with “s” (significant) that is given after the values of ADF statistics; if the coefficients were significant (Pr>|t|<0.05) “s” was written; if not, nothing was put to the related cell. The inconclusive ADF tests (the ones with the coefficient of the 1st lag of the dependent variable in the right of ADF regression is not “s” in the left; the dependent variable appear with the differentiated form) was specified with “inconclusive”. In such cases where ADF test is inconclusive, all the statistics and data (ADF statistics, the significance of drift and time trend term, probability value, optimal minimum lag order) were disregarded since they do not make sense in the inconclusive situation. “omlo” is optimal minimum lag order.

R> V6Nonstationary43ObsOL.df
R> V6Stationary42ObsOLf.df <- data.frame(matrix(NA, nrow = 42, ncol = 7))

R> V6Stationary42obsOLf.df

# The data frame V6Stationary42obsOLf.df is formed by the 1st differences of the variables of V6Nonstationary43ObsOL.df:

```r
R> V6Stationary42obsOLf.df <- data.frame(diff(V6Nonstationary43ObsOLfdf[2], differences=1),diff(V6Nonstationary43ObsOLfdf[3], differences=1),diff(V6Nonstationary43ObsOLfdf[4], differences=1),diff(V6Nonstationary43ObsOLfdf[5], differences=1),diff(V6Nonstationary43ObsOLfdf[6], differences=1),diff(V6Nonstationary43ObsOLfdf[7], differences=1))
```

V6Stationary42obsOLf.df <-

data.frame(V6Stationary42obsOLfdf[2:43,1],
            diff(V6Nonstationary43ObsOLfdf[2],
                 differences=1),diff(V6Nonstationary43ObsOLfdf[3],
                 differences=1),diff(V6Nonstationary43ObsOLfdf[4],
                 differences=1),diff(V6Nonstationary43ObsOLfdf[5],
                 differences=1),diff(V6Nonstationary43ObsOLfdf[6],
                 differences=1),diff(V6Nonstationary43ObsOLfdf[7],
                 differences=1))

```
R> dim(V6Stationary42obsOLfdf) # 42 7
```

R> colnames(V6Stationary42obsOLfdf) <- c("obs" ,"aciklik1d", "istikrar1d", "kur1d","lnDysy1d", "lngsyi1d", "ucus1d")

R> V6Stationary42obsOLf.df

Since the negative values in the data frame may give rise to complications in calculations, in order to prevent the possible future problems, all the values in the data frame were converted to positive values. If there is no negative value or 0 in a column in a data frame, then there is no need to make a conversion in that column. In order to make all the value in the data frame positive; in each of the columns (variables) where there are negative values, the minimum value in that column is subtracted from all of the values in the column, and then a small value (for example, 0.3) is added to all the values in the column. Thereby, a column where all of the values are positive is obtained.

```r
```

R> V6Stationary42obsOLf.df

http://www.ijsat.com

International Journal of Science and Advanced Technology (ISSN 2221-8386)

Volume 4 No 10 October 2014

http://www.ijsat.com

"FDI" of Turkey. The 6 variables in the data frame may give inconclusive situation. "omlo" is optimal minimum lag order.
In this version, all variables are stationary (with positive values variables). The suffix “1d” is to denote 1st difference of variables.

In this version, all variables are stationary (with positive values variables). The suffix “1d” is to denote 1st difference of variables.

# The 1st column is of observation labels
R> V6Stationary42ObsOLf.df[,2:7]
openness1d stability1d exchangerate1d lnFDI1d lnGDP1d flights1d

1 0.38082 0.30 0.42552 2.47983 0.50245 0.38674
2 0.38569 1.25 0.42552 2.68815 0.53382 0.39283
…………
42 0.36691 1.23 0.54657 2.4773 0.42402 0.52122

The optimal minimum lag order for the vector autoregressive (VAR) model of 6-variable system can be determined by either of ARorderG and VARomlop:

R > ARorderG(V6Stationary42ObsOLf.df[,2:7])
# [1] 5
R > VARomlop(V6Stationary42ObsOLf.df[,2:7])

Choosing the optimal block length of the future bootstrapping operations (if desired) can be performed by b.star:

R > mean(b.star(V6Stationary42ObsOLf.df[,2:7]))
# [1] 3.286664

Pairwise analysis can be performed by taking both of the number of independent and dependent variables as 1. We want to emphasize once more: the word “pairwise” used here in nor in the sense of classical G-causality where variables in a system are taken pairwise and the effects of the other variables are ignored. The word “pairwise” used here (in advanced G-causality sense) means:

- “both of the number of elements in the sets of independent and dependent variables are 1”,
- and the total number of independents and dependents cannot exceed 2.

- the effect of the remaining variables in the system on these 2 variables (independent and dependent) are calculated/taken into account,
- this confounding effect is subtracted from the effect of independents and dependents (plus the effect of the variables on which the effect of independent to dependent was conditioned) to each other,
- and eventually the G-causality from independents to dependents is obtained isolated from the (possible confounding effects of) the conditioning variables.

All of the pairwise conditional G-causalties in 6-variable system can be found via conditionalGgFp (suffix: g:graphics; F:F statistic; p:p-value) function of the causfinder package (Since we work with the stationary variables, the maximum order of integration that makes all the variables in the system stationary (“maxoi”) is 0; AR order=5; the 0.10 significance line was drawn to give an idea; any level of significance can be chosen at this stage):

R> conditionalGgFp(V6Stationary42ObsOLf.df[,2:7],ninde pendents=1,ndependents=1,nobsraw=42,maxoi=0, order=5)

The direction of G-causalties (F statistics) is from columns to rows: e.g. the (F statistic of the) conditional G-causality of aciklik1d to istikrar1d conditioned on the remaining others is 1.137. The conditional G-causality of aciklik1d to ln dysy1d conditioned on the remaining others is 0.311’dir. The conditional G-causality of ucus1d to ln dysy1d conditioned on the remaining others is 0.042.
After the G-noncausality test, the decision whether there is a causality or not in the searched direction is made according to the p value of the F statistic: If \( p>0.05 \) then “\( H_0: \) non-causality” is hold, if \( p<0.05 \) then “\( H_0: \) non-causality” is rejected and there is a G-causality among the variables in the tested direction.

The conditional GC intensity of ucus1d:
\[
R> \text{sum(conditionalGgFp(V6Stationary42ObsOLf.df[,2:7], \[21:25,2\])]} \# 0.9843924
\]

In the light of the above information, the order of the 6 variables in the system in terms of affecting the other variables conditional on the remaining is as follows:

lngsyih1d (0.86) > lndysy1d (0.92) > ucus1d (0.98) > aciklik1d (1.11) > istikrar1d (1.13) > kur1d (1.43).

This result of the analysis is completely consistent with the realities of the economy: during 1970-2012, the most influential variable in the economy is gross domestic product (GDP); this is followed by FDI, then flights per capita (as the proxy variable of the infrastructure), then trade openness, political stability, and finally exchange rate.

When the effect-receiving of the 6 variables in the system is analyzed:

lngsyih1d: \( 0.351 + 0.407 + 0.317 + 0.447 + 0.5 = 2.022 \)

lndysy1d: \( 0.365 + 0.282 + 0.467 + 0.218 + 0.271 = 1.603 \)

ucus1d: \( 0.257 + 0.292 + 0.29 + 0.232 + 0.135 = 1.206 \)

aciklik1d: \( 0.131 + 0.298 + 0.149 + 0.114 + 0.121 = 0.813 \)

kur1d: \( 0.071 + 0.026 + 0.251 + 0.101 + 0.035 = 0.484 \)

istikrar1d: \( 0.07 + 0.062 + 0.077 + 0.066 + 0.058 = 0.333 \)

In the light of the above information, the order of the 6 variables in the system in terms of receiving effects from the other variables conditional on the remaining is as follows:

istikrar1d (0.33) > kur1d (0.484) > aciklik1d (0.813) > ucus1d (1.206) > lngsyih1d (1.603) > Indysy1d (2.022).

To order the 6 variables from more exogenous (independents) to indigenous (dependents), the results are collected together (Table 4). Therefore, the ordering of the 6 variables from exogenous (independents) to indigenous (dependents) is: lngsyih1d (the most exogenous), lndysy1d, ucus1d, aciklik1d, istikrar1d, kur1d (the most endogenous). In Table 4, we multiplied the 2nd column with “-1” when obtaining the 3rd column to make the evaluation from the same baseline (the smaller the total of \( p \) values the bigger the independence (from the perspective of affecting the others)). One could also multiple the 1st column with “-1” when obtaining the 3rd column to make the evaluation from the same baseline (this time: the bigger the total of \( p \) values the bigger the independence (from the perspective of receiving the effects from the others)).
Therefore, the effect powers of the determinants of FDI in Turkey (when analyzed in the system of 6 variables) from more influential to less influential are as follows:

kur1d (0.317) > aciklik1d (0.351) > istikrar1d (0.407) > lnGSYIH1d (0.447) > ucs1d (0.5).

From here, this inference can be made: the most important variable (among the analyzed variables) that affects the FDI of Turkey is the exchange rate (kur). Trade openness (aciklik) follows the exchange rate in that. Another attracting point is that the effect/contribution of the political stability to FDI is larger than that of the bigness of the economy (GDP; GSYIH). That is to say, foreign investors prefer the political stability in Turkey more than total bigness of the economy when investing to Turkey.

Foreign investors attach the infrastructure in Turkey less importance if compared with the other variables when investing.

One can perform econometric analyses on the subsystems of the main 6-variable system as well. The 6-variable-system can be restricted to 4-variable-system in which, for example, the trade openness, exchange rate, lnFDI and lnGDP variables are included in the system and political stability and annual flights per capita are excluded. The optimal minimum lag order for the vector autoregressive model of such a 4-variable system can be determined by either of ARorderG and VARomlop:

R> ARorderG(V6Stationary42ObsOLf.df[,c(2,4,5,6)])
# [1] 7
R> VARomlop(V6Stationary42ObsOLf.df[,c(2,4,5,6)])

Choosing the optimal block length of the future bootstrapping operations can be performed by b.star:

R > mean(b.star(V6Stationary42ObsOLf.df[,c(2,4,5,6)]))
[1] 3.233398

Pairwise analysis can be performed by taking both of the number of independent and dependent variables as 1. Since we work with the stationary variables, the maximum order of integration that makes all the variables in the system stationary (“maxoi”) is 0. Hence, the pairwise conditional Granger causalities of this system (with AR order =7) is:

R > conditionalGgFp(V6Stationary42ObsOLf.df[,c(2,4,5,6)], nindependents=1,ndependents=1,nobsraw=42,maxoi=0, order=7)

The plots of F statistics and p values are obtained as well as their tabulated matrices just as in the case of the full system.

III. CONCLUSION

The variables and stationarity analysis are definitely almost part of every econometric research. The causality relationships among the variables of the systems of the researchers definitely are a must for them. causfinder not only provides the values of the GC statistics, directions, etc. in a G-causality analysis, but also it gives the degrees of causations. causfinder became the 1st tool in R for the systemwise analysis of conditional and partial G-causalities. We believe that advanced Granger world will be discovered with it more easily.
ACKNOWLEDGMENT

* I thank Fabian SCHEIPL (Dept. of Statistics, LMU Munich, Germany) for his helpful suggestions during the construction of adfcs.

I thank Funda YURDAKUL (Dept. of Econometrics, Gazi Univ., Turkey) for her helpful suggestions for the interpretation of combined p values in the last part of the paper besides her overall evaluation.

I thank all of the R community for their cooperation and support during my preparation of this package. Besides this, I want to thank a lot to those who helped me in theoretical sides of the study I conducted.

REFERENCES


E.Cevher, (Istanbul, 1976) received his BS degree in Mathematics in Bogazici University in 1999.

After graduation, he worked in the Mathematics Dept. of Dogus University in 1999-2002.


He is currently working in the Ministry of Science, Industry and Technology in Ankara, Turkey since 2004.