A Speedy and Seamless Stationarity Analysis via causfinder Package in R

Erdogan CEVHER[a]

[a] The Ministry of Science, Industry and Technology, Eskisehir Yolu, 7.Km, Cankaya, Ankara 06510, Turkey

erdogancevher@gmail.com

Abstract. Stationarity analysis is a must for various studies in many fields that employ time series analysis. adfcs function in causfinder package in R performs Augmented Dickey-Fuller (ADF) test that takes into account the usage of same (i.e., common) sub-sample for all of the lag orders for the autoregressive process when stationarity is investigated. As is known, in all of the lag selection procedures in econometrics, same sub-sample must be used to determine the correct optimal minimum lag. We bouqoueted adfcs functions in adfcstable function whose functional value is a table that reveals all of the needed stationarity analysis of all the variables in a given system in a couple of seconds. In the returned ADF table, the results of all of the three standard cases ("both drift and time trend", "drift without time trend" and "no drift, no time trend") are presented for all of the variables in question. Whether the drift and time trend coefficients in the ADF regressions is significant is specified. adfcstable reveals the inconclusivities of ADF tests (the coefficient of the 1st lag of the dependent variable in the right of ADF regression is not "<0"; in the left, the dependent variable appear with the differenced form) whenever there appears such cases. It also presents optimal minimum lag order for the ADF regressions. We used three datasets from various fields: a dataset of functional integration of brain, a dataset for the determinants of foreign direct investment in Turkey, and a dataset for the determinants of current account deficit of Turkey.

Keywords: stationarity, augmented Dickey Fuller test, causfinder, R.

1 THEORY

1.1 Stationarity

A time series y_t is said to be "stationary" ("weak stationary", "second-order stationary"; "covariance stationary") if it satisfies all the following properties together:

(i) the mean of y_t is constant over time
(ii) the variance of y_t is constant over time
(iii) the covariance between two values of y_t does not depend on the real time the two observed, but depends only on the time length between these two values.

In other words, if the statistical properties of y_t is constant over time, i.e., for two different time intervals, the sample means and sample covariances of y_t are the same over time, then y_t is stationary. Expressed in mathematical syntax;
y_t is stationary if, the followings holds for all values of y_t and for all time periods:

$$\forall t \qquad E(y_t) = \mu \qquad \text{(constant mean)}$$
$$\forall t \qquad \mathrm{var}(y_t) = E[(y_t - \mu)^2] = \sigma^2 \qquad \text{(constant variance)}$$
$$\forall t \qquad \mathrm{cov}(y_t, y_{t+s}) = \mathrm{cov}(y_t, y_{t-s}) = E[(y_t - \mu)(y_{t-s} - \mu)] = \gamma_s$$

(the covariance depends on s, not on t).

Here, μ, σ^2 and γ_s are fixed finite numbers that do not depend on time t (for γ_s, different constant for different s).

1.2 Augmented Dickey-Fuller (ADF) test

Since ADF test is covered extensively in literature, we summarize it very briefly: There are three standard cases ("both drift and time trend", "drift without time trend" and "no drift, no time trend") in ADF regression equations. Define $\gamma \equiv \rho - 1$ and $(\Delta y)_t \equiv y_t - y_{t-1}$.

i) No drift (constant, intercept), no time trend:

$$y_t = \rho y_{t-1} + \nu_t$$

$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + \nu_t$$

$$(\Delta y)_t = (\rho - 1) y_{t-1} + \nu_t$$

$$(\Delta y)_t = \gamma y_{t-1} + \nu_t.$$

H_0: $\rho = 1$ $(\gamma = 0)$ corresponds to nonstationarity (unit root) of y_t whereas H_1: $\rho < 1$ $(\gamma < 0)$ corresponds to stationarity of y_t (See Fig. 1).

ii) Drift (α; constant, intercept; α can be nonzero), but no time trend:

$$\begin{aligned} y_t &= \alpha + \rho y_{t-1} + \nu_t \\ y_t - y_{t-1} &= \alpha + \rho y_{t-1} - y_{t-1} + \nu_t \\ (\Delta y)_t &= \alpha + (\rho - 1) y_{t-1} + \nu_t \\ (\Delta y)_t &= \alpha + \gamma y_{t-1} + \nu_t. \end{aligned}$$

Null and alternative hypothesis are the same as the first case (See Fig. 1).

iii) Both drift (α; constant, intercept; α can be nonzero) and deterministic time trend (λt; λt can be nonzero):

$$\begin{aligned} y_t &= \alpha + \rho y_{t-1} + \lambda t + \nu_t \\ y_t - y_{t-1} &= \alpha + \rho y_{t-1} - y_{t-1} + \lambda t + \nu_t \\ (\Delta y)_t &= \alpha + (\rho - 1) y_{t-1} + \lambda t + \nu_t \\ (\Delta y)_t &= \alpha + \gamma y_{t-1} + \lambda t + \nu_t. \end{aligned}$$

Null and alternative hypothesis are the same as the first case (See Fig. 1).

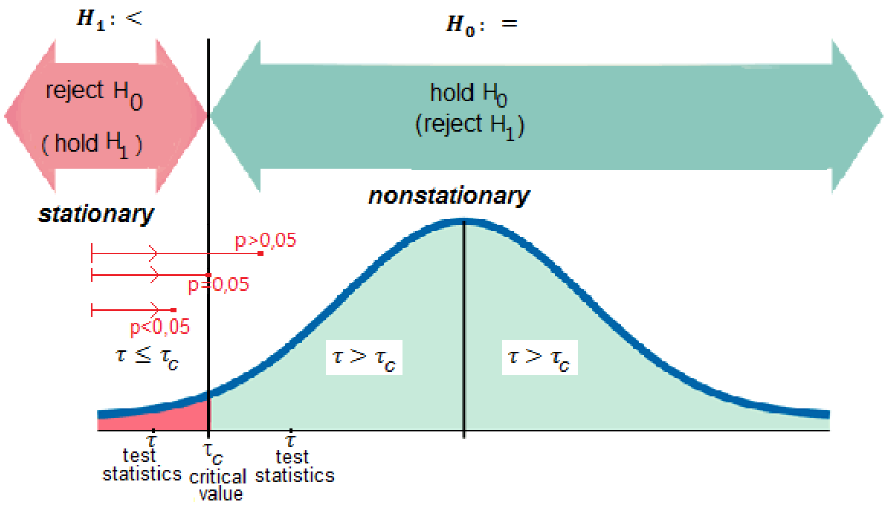

Fig. 1. The stationarity and nonstationarity regions of DF and ADF tests.

2 APPLICATION

2.1 The package causfinder

causfinder is an R package that we developed specifically for systemwise analysis of conditional and partial Granger causalities. The first version of the package causfinder appeared in October 2014 (Cevher, 2014). By the way, there is no spelling error in naming the package; naming was intentionally selected like that based on the inspiration from the packages “causnavigator”, “knitr” (neater) etc. in R.

causfinder can perform all the needed stationarity analysis of all the variables in a given system in a couple of seconds. In the returned ADF table, the results of all of the three standard cases (“both drift and time trend”, “drift without time trend” and “no drift, no time trend”)) are reported for all of the variables in the system.

2.2 The Meaning of “Speedy” and “Seamlessness” in the Context of the Paper

Speediness: There are various software tools that perform ADF test for a variable in a given system. Our systems contain more than one variables in most of the cases. Performing ADF test for each of the variables in the system separately and also another separate operations for all the three standard cases of ADF test may really become very burdensome. Consider a neuroscience experiment where there are 100 variables. In this case, 100x3=300 ADF tests to clarify very neatly the stationarity investigations in such platforms would become a bottleneck at some point (here, “3” for three distinct ADF regression cases mentioned in earlier headings). causfinder performs all the 300 ADF tests with just one command at once. So, for 5 variables, consider the practicality of only 1 operation instead of 5x3=15 operations.

Seamlessness: In all of the lag selection procedures in econometrics, same sub-sample must be used for all the optimal lag candidates to determine the correct optimal minimum lag order. In some software programs, same sub-sample is not used, hence, we deem them completely erroneous. There are some software packages where this fact (the usage of same sub-sample) is neglected during reporting phase. We deem it important to present the ADF test results taking into account this mathematical reality.

2.3 adfcs function

adfcs function performs Augmented Dickey-Fuller (ADF) test that uses common (sub-)sample for all of the lags when a selected max order is given to find optimal minimal lag order of the autoregressive process.

The usage of the adfcs function is as follows:

```
adfcs(t, max = floor(12 * (length(t)/100)^(1/4)), type = c("c"))
```

The default value of max is Schwert's approximation (Schwert, 1989). When using adfcs, one should care to specify the column of the variable for which the ADF test is wanted when data frame is supplied for the parameter t. The arguments of the adfcs function are as follows:

t: a numeric vector or time series object whose stationarity is examined
max: selected max lag order to find optimal minimal lag order for the autoregressive process in ADF test. The default is Schwert’s approximation.
type: a character string describing the type of the unit root regression. Valid choices are “nc” for a regression with no intercept (constant) nor time trend, and "c" for a regression with an intercept (constant) but no time trend, “ct” for a regression with an intercept (constant) and a time trend. The default is “c”.

2.4 adfcstable function

We bouqoueted adfcs functions in adfcstable function whose functional value is a table that reveals all of the needed stationarity analysis of all the variables in a given system in a couple of seconds. In the returned ADF table, the results of all of the three standard cases (“both drift and time trend”, “drift without time trend” and “no drift, no time trend”) are presented for all of the variables in question. Whether the drift and time trend coefficients in the ADF regressions is significant is specified. adfcstable reveals the inconclusivities of ADF tests (the coefficient of the 1st lag of the dependent variable in the right of ADF regression is not “<0”; in the left, the dependent variable appear with the differenced form) whenever there appears such cases. It also presents optimal minimum lag order for the ADF regressions.

The usage of the adfcstable function is as follows:

```
adfcstable(d, max = 5)
```

The arguments of the adfcstable function are as follows:

d: The system of variables in the form of matrix, data frame, etc. It is suggested that observation labels are coded as the row names of data frame, not as a separate column for the simplicity of the analysis.
max: The maximum order of integration that will be checked for variables.

As an example, consider the following command:

```
R> adfcstable(((1:100)+ runif(100, min=0, max=0.1))^3)
```

2.5 Datasets

We used three datasets from various fields: a dataset of functional integration of brain, a dataset for the determinants of foreign direct investment in Turkey, and a dataset for the determinants of current account deficit of Turkey.

2.5.1 The dataset of functional integration of brain

We used the data "grangerdata" in (Roelstraete and Rosseel, 2011) that was used there for Granger causality computations. Note that the grangerdata that is used in the article is from FIAR v0.3, which we entitled as "granger.df" in our paper (Roelstraete and Rosseel used different dataset with the same name, i.e. grangerdata in documentation of FIAR v0.5).

The granger.df dateframe includes 5 variables (x, y, z, q, w) with 2000 observations. These 5 variables are numeric vectors in R. Let's see the content of the dataset (notice the command lines are indicated by "R>" to distinguish them from the output of the commands; sometimes outputs are shortened in a way their messages are understood for the length considerations):

```
R> head(granger.df); tail(granger.df)
                  x          y          z          q           w
     1  3.7509252  0.2692284  1.293244  0.3135299 -0.28840307
     2  0.7023175 -1.8135786 -2.798717 -5.2041356 -3.85116801
     3  5.6832003  9.0572561  7.152931  3.8353702  9.18229397
     ……………………………………………………………………..
  1998 11.848116  -2.411037  0.8868141   2.822734  -3.63950265
  1999  9.988895   7.306975  2.9845350  -2.545514   0.07715012
  2000  0.922414   3.289325 -6.5278651 -11.523919  -0.63251466
```

Now, apply the adfcs function for one of the variables in granger.df, say x (See Figure 2).

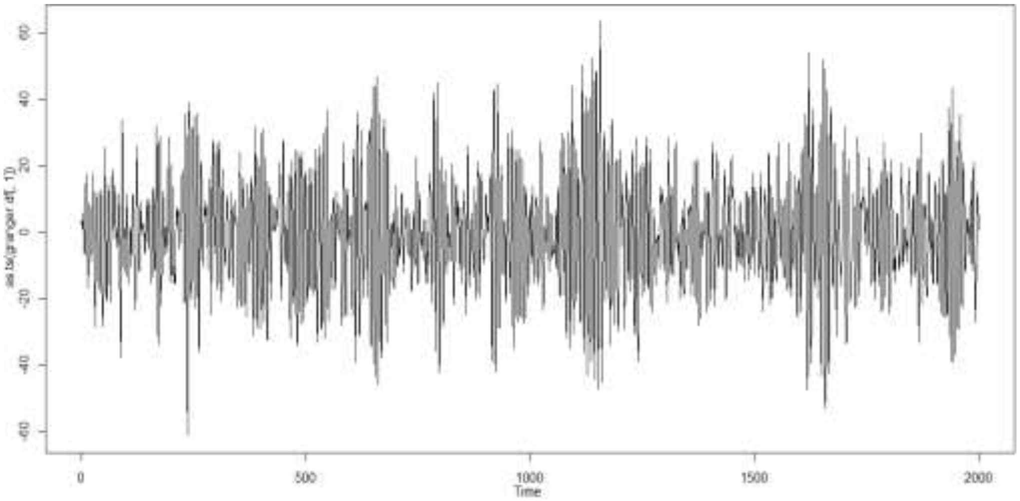

Fig. 2. The first variable in granger.df, i.e., x obtained via plot(ts(granger.df[,1])).

```
R> adfcs(granger.df[,1]) # granger.df[,1] is the 1st column in the dataframe, i.e., x.
# optimal minimum lag orders for Schwarz (Bayesian) Information Criterion and
# Akaike Infirmation Criterion
$optmins
```

```
$optmins[[1]]
[1] 1
$optmins[[2]]
[1] 1
$SBCAIC
  SBCvalues AICvalues
1    12016.6  11994.25
2   12023.82  11995.88
………………………
25  12187.67  12031.21
26  15375.29  15358.53

$adfcst
Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 1
  STATISTIC:
    DF: -75.6024
  P VALUE:
    t: 1e-04
    n: 8.626e-08
```

From the output of the adfcs function, the p value of the test statistics of the ADF test is 10^{-4}, and hence less than 0.05. Hence, one may conclude the x is a stationary variable (See Fig. 1).

Before proceeding to adfcstable function, let's show how the same values (-75.6024 and 1e-04) are obtained via Eviews.

First take the granger.df dataframe from R to EViews; the following creates the .csv file in R's working directory (which can be found via getwd() command in R prompt):

```
R> write.csv(granger.df,file="granger.df.EViews.csv")
```

Import granger.df.EViews.csv to EViews: in Eviews platform, click "File – Import – Import from file…"; browse to R's working directory, choose granger.df.EViews, Open. When one follows the course, the variables x, y, z, q, w appears in EViews panel. Double click x. In order to perform ADF test for x in EViews, click "View – Unit Root Test…", select ADF as test type. The Schwert's approximation (Schwert, 1989)

$$\left[\!\left[12\sqrt[4]{\frac{T}{100}} \right]\!\right]$$

specifies that the chosen maximal lag order is 25 for x (The Schwertz's approximation is the default maximal lag order to be tried in adfcs as well; check in R that floor(12*(2000/100)^(1/4)) is equal to 25). Since the default type for the ADF regression equation is the one with only the constant (drift) case, let's keep the ADF test regression type in EViews as well. Notice that in EViews 8, the default ADF regression type is the same: only the inclusion of constant in ADF test equation. Leave the "Test for unit root in" as "Level".

Lag Length is “automatic selection (Schwarz Info Criterion)”. Now, click OK. The resulting pane reveals ADF test statistics and probability to be (-75.82448 and 0.0001) respectively. Now, the difference in the ADF test result in EViews and R (resp. -75.82448 and -75.6024) is due to the fact that though EViews also focuses on the usage of the same sub-sample for all lags in the lag selection procedure, it prints the ADF test results by focusing on the sub-sample that suits optimal minimal lag order. Keeping this in mind, we can obtain the same result found in R from EViews as well:

The resulting pane in EViews also states “Included observations: 1998 after adjustments”. Since the same common sub-sample must be used in all of the lag selection procedures in Econometrics, this 1998 is erroneous: one must trim 25+1 lags from the beginning to reach that the same 2000-(25+1)=1974 observations are used in all of the 25 lag selection course in the loop (when no lag present in ADF regression, the difference of the dependent variable on the left of the regression equation by design causes the losing of one observation, hence, “+1”). To satisfy this, let’s fix sub-sample in EViews and try ADF test again:

- Double click on Range, change “Workfile structure type” to “Undated with ID Series”, choose series01 as Identifier series (note that series01 holds values from 1 to 2000), and click OK.
- Click Sample, and change “@all” to “27 2000”, and click OK to see “Sample: 27 2000 – 1974 obs”. Now, applying the same operations above for the ADF test of x in EViews results in the pair (-75.60244 0.0001) which is the equivalent of (-75.6024 and 1e-04) in R.

Now, let’s apply adfcstable function that performs stationarity analysis for all of the variables in a given dataframe. In our example, adfcstable will investigate x, y, z, q, w in granger.df.

```
R> adfcstable(granger.df) # The below output is obtained nearly within 2 minutes
```

The output of the adfcstable function is the Table 1&2. The users of causfinder package may desire to get the output of the function in R to various platforms (like MS Word). This is done as follows:

Table 1. The result of the adfcstable function for the variables in the granger.df dataframe (see Table 2).

	var	type	inc	levelt	Pc	c	Pt	t	prob	omlo
1	x	dt	-0.553	-75.583	0.427		0.944		0	1
2	x	d	-0.553	-75.602	0.087		X	X	0	1
3	x	-	-0.553	-75.546	X	X	X	X	0	1
4	y	dt	-0.709	-14.369	0.493		0.898		0	9
5	y	d	-0.709	-14.372	0.11		X	X	0	9
6	y	-	-0.701	-14.278	X	X	X	X	0	9
7	z	dt	-1.822	-42.07	0.807		0.987		0	4
8	z	d	-1.822	-42.081	0.604		X	X	0	4
9	z	-	-1.821	-42.085	X	X	X	X	0	4
10	q	dt	-1.125	-13.322	0.586		0.834		0	10
11	q	d	-1.125	-13.324	0.145		X	X	0	10
12	q	-	-1.111	-13.24	X	X	X	X	0	10
13	w	dt	-0.928	-13.602	0.499		0.948		0	8
14	w	d	-0.928	-13.606	0.141		X	X	0	8
15	w	-	-0.918	-13.522	X	X	X	X	0	8

Table 2. The continuation of Table 1 from the right.

	var	type	inc	1stDifft	Pc	c	Pt	t	prob	omlo	interorder
1	x	dt	-3.924	-21.599	0.991		0.959		0	13	
2	x	d	-3.924	-21.605	0.945		X	X	0	13	I(0)
3	x	-	-3.924	-21.611	X	X	X	X	0	13	
4	y	dt	-5.66	-20.896	0.949		0.911		0	13	
5	y	d	-5.66	-20.901	0.948		X	X	0	13	I(0)
6	y	-	-5.66	-20.906	X	X	X	X	0	13	
7	z	dt	-21.378	-18.506	0.949		0.951		0	22	
8	z	d	-21.377	-18.511	0.983		X	X	0	22	I(0)
9	z	-	-21.377	-18.515	X	X	X	X	0	22	
10	q	dt	-8.817	-25.288	0.972		0.973		0	12	
11	q	d	-8.817	-25.295	0.991		X	X	0	12	I(0)
12	q	-	-8.817	-25.301	X	X	X	X	0	12	
13	w	dt	-9.519	-19.301	0.941		0.918		0	16	
14	w	d	-9.519	-19.306	0.975		X	X	0	16	I(0)
15	w	-	-9.519	-19.311	X	X	X	X	0	16	

Note: Variables: x, y, z, q, w. In ADF regression, the results were given in the order of "both drift and time trend (dt)", "drift without time trend (d)" and "no drift, no time trend (-)". Next, whether ADF tests are inconclusive (the case when the coefficient of the 1st lag of the dependent variable in the right of ADF regression is not "<0"; in the left, the dependent variable appear with the differenced form)) or not is reported in "inc" column. levelt is ADF (Tau) statistics of the level of variable; 1stDifft is ADF (Tau) statistics of the 1st difference of variable. Level and 1st difference of variables are analyzed one after the other; hence, the relevant ADF analysis of 1st difference of variables are performed just after when all Pc, c, Pt, t data of the levels are supplied. 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. In such cases where ADF test is inconclusive, one may disregard all the statistics and data (ADF statistics, the significance of drift and time trend term, probability value, optimal minimum lag order) since they do not make sense in the inconclusive situation. prob columns gives the p values of the relevant ADF test statistics. "omlo" is optimal minimum lag order. Eventually, integration orders of the levels are presented.

Step 1: write the output as .csv file with (you may choose any name for the .csv file; R automatically creates the file, hence, one does not need to create it in the outer environment like Windows Explorer) the following command:

```
R> write.csv(adfcstable(granger.df),file="output.csv")
```

Note that output.csv is created in R's working directory (can be found via getwd()).

Step 2: open the .csv file created in Step 1 in your spreadsheet program (libreOffice Calc, MS Excel etc.). Both of the MS Excel and libreOffice Calc ways will be shown below:

R's adfcstable output to MS Word via MS Excel:

Step 3: Create an empty .xlsx file. Double click it to open.

Step 4: Click the Data tab. Then, "Dış Veri Al, Metinden". Select the output.csv in R's working directory. Import.
Step 5: The file type that best describe the data: limited; the data includes metadata. Click to Next. As the separator, choose both Tab and Comma. Click to Next and Finish.

Step 6: Choose any part of the data sheet you desire to get MS Word (copy; control-c).

Step 7: Go to MS Word. Paste special as "RTF".

R's adfcstable output to MS Word via libreOffice Calc: libreOffice Calc can automatically tabulate the content of the .csv file. See below. Open the file output.csv.

Step 3: Change all of the fields labeled as "Standard" to "Text" (change the decimal separator (comma or point) from control panel if necessary). Press OK to get the data values appear in LibreOffice Calc.

Step 4: Choose any part of the data sheet you desire to get MS Word (copy; control-c).

Step 5: Go to MS Word. Paste special as "RTF".

Now, as is seen from Table 1 and Table 2, the variables in granger.df (x, y, z, q, w) are all I(0), i.e., are all stationary.

2.5.2 The dataset for the determinants of foreign direct investment in Turkey

causfinder package includes the dataframe V6Nonstationary43Obs.df (6 variables; some are nonstationary; 43 observations) which is used to find the determinants of foreign direct investment in Turkey during 1970-2012. The 6 variables in this dataframe are trade openness, political stability, exchange rate, the natural logarithm of foreign direct investment (lnFDI), the natural logarithm of gross domestic product (lnGDP), and the number of flights per capita annually.

```
R> colnames(V6Nonstationary43Obs.df) <- c("openness", "stability", "exchangerate",
"lnFDI", "lnGDP","flight")
R> head(V6Nonstationary43Obs.df); tail(V6Nonstationary43Obs.df)
          openness stability exchangerate   lnFDI    lnGDP  flight
     1970  0.06690     1.00      0.00001  4.06044 10.04151 0.07525
     1971  0.06902     0.07      0.00001  3.80666 10.19493 0.09031
     1972  0.07601     0.09      0.00001  3.76120 10.37972 0.11146
     ……………………………………………………………………….....
     2010  0.26000     0.94      1.50285  9.10897 13.95671 1.39441
     2011  0.29830     0.97      1.67495  9.68328 14.04634 1.57406
     2012  0.28651     0.97      1.79600  9.42698 14.12133 1.72360
```

Let's perform adfcstable on V6Nonstationary43Obs.df:

```
R> adfcstable(V6Nonstationary43Obs.df)
```

Take the output in R to MS Word as described in the above step-by-step procedure:

```
R> write.csv(adfcstable(V6Nonstationary43Obs.df),file="V6Nonstationary43Obs.df.csv")
```

Table 3. The result of the adfcstable function for the variables in the V6Nonstationary43Obs.df dataframe (see Table 4).

	var	type	inc	levelt	Pc	c	Pt	t	prob	omlo
1	open	dt	-0.308	-2.302	0.048	s	0.043	s	0.422	0
2	open	d	-0.067	-0.905	0.197		X	X	0.774	0
3	open	-	0.025	0.988	X	X	X	X	0.911	0
4	stab	dt	-0.389	-2.751	0.114		0.12		0.224	0
5	stab	d	-0.303	-2.259	0.037	s	X	X	0.191	0
6	stab	-	-0.031	-0.595	X	X	X	X	0.452	0
7	exc	dt	-0.133	-2.051	0.158		0.029	s	0.554	1
8	exc	d	-0.003	-0.094	0.238		X	X	0.942	1
9	exc	-	0.02	0.69	X	X	X	X	0.86	1
10	lnFDI	dt	-0.576	-3.516	0.001	s	0.002	s	0.054	0
11	lnFDI	d	-0.034	-0.59	0.273		X	X	0.86	1
12	lnFDI	-	0.028	1.697	X	X	X	X	0.976	1
13	lnGDP	dt	-0.766	-4.136	0	s	0	s	0.013	2
14	lnGDP	d	-0.009	-0.643	0.153		X	X	0.848	4
15	lnGDP	-	0.011	4.568	X	X	X	X	1	4
16	flight	dt	0.088	1.61	0.376		0.539		1	0
17	flight	d	0.119	5.249	0.517		X	X	1	0
18	flight	-	0.107	7.609	X	X	X	X	1	0

Table 4. The continuation of Table 3 from the right.

	var	type	inc	1stDifft	Pc	c	Pt	t	prob	omlo	interorder
1	open	dt	-1.049	-5.71	0.947		0.547		0	0	
2	open	d	-1.038	-5.738	0.244		X	X	0	0	I(1)
3	open	-	-0.983	-5.584	X	X	X	X	0	0	
4	stab	dt	-1.018	-6.238	0.54		0.77		0	0	
5	stab	d	-1.023	-6.399	0.457		X	X	0	0	I(1)
6	stab	-	-1.019	-6.42	X	X	X	X	0	0	
7	exc	dt	-0.605	-3.546	0.944		0.349		0.051	0	
8	exc	d	-0.563	-3.421	0.168		X	X	0.018	0	I(1)
9	exc	-	-0.478	-3.072	X	X	X	X	0.003	0	
10	lnFDI	dt	-1.291	-8.159	0.085		0.504		0	0	
11	lnFDI	d	-1.296	-8.278	0.022	s	X	X	0	0	I(1)
12	lnFDI	-	-1.202	-7.376	X	X	X	X	0	0	
13	lnGDP	dt	-2.519	-5.069	0	s	0.508		0.001	4	
14	lnGDP	d	-2.462	-5.082	0	s	X	X	0	4	I(1)
15	lnGDP	-	-0.179	-1.105	X	X	X	X	0.239	4	
16	flight	dt	-0.756	-4.204	0.446		0.014	s	0.012	0	
17	flight	d	-0.46	-3.008	0.034	s	X	X	0.045	0	I(1)
18	flight	-	-0.269	-2.003	X	X	X	X	0.045	0	

Granger causality analysis requires the variables to be stationary (when time-varying and windowing techniques are not used) and since the order of integration of variables in

V6Nonstationary43Obs.df is all 1 (can be found via adfcs), taking the first differences for the 6 variables in V6Nonstationary43Obs.df results in V6Stationary42Obs1D.df (6 variables; all are stationary; 42 observations). Now, we will have causfinder confirmed this as well.

```
R> V6Nonstationary43Obs.df

R> V6Stationary42Obs1D.df <- data.frame(matrix(NA, nrow = 42, ncol = 6))
   V6Stationary42Obs1D.df <- data.frame(diff(V6Nonstationary43Obs.df[,1],
differences=1),diff(V6Nonstationary43Obs.df[,2],
differences=1),diff(V6Nonstationary43Obs.df[,3],
differences=1),diff(V6Nonstationary43Obs.df[,4],
differences=1),diff(V6Nonstationary43Obs.df[,5],
differences=1),diff(V6Nonstationary43Obs.df[,6], differences=1))
R> colnames(V6Stationary42Obs1D.df) <- c("openness1d", "stability1d", "exchangerate1d",
"lnFDI1d", "lnGDP1d","flight1d")

R> row.names(V6Stationary42Obs1D.df) <- 1971:2012

R> V6Stationary42Obs1D.df

R> head(V6Stationary42Obs1D.df); tail(V6Stationary42Obs1D.df)
```

```
     openness1d stability1d exchangerate1d lnFDI1d lnGDP1d flight1d
     1971    0.00212      -0.93          0e+00 -0.25378 0.15342  0.01506
     1972    0.00699       0.02          0e+00 -0.04546 0.18479  0.02115
     1973    0.01198       0.56          0e+00  0.60825 0.18331  0.02355
     ……………………………………………………………………….
    2010    0.02596       0.00      -0.04711  0.04215  0.10335  0.21598
    2011    0.03830       0.03       0.17210  0.57431  0.08963  0.17965
    2012   -0.01179       0.00       0.12105 -0.25630  0.07499  0.14954
```

Let's perform adfcstable on V6Stationary42Obs1D.df:

```
R> adfcstable(V6Stationary42Obs1D.df)
```

Take the output in R to MS Word as described in the above step-by-step procedure:

```
R> write.csv(adfcstable(V6Stationary42Obs1D.df),file="V6Stationary42Obs1D.df.csv")
```

Table 5. The result of the adfcstable function for the variables in the V6Stationary42Obs1D.df dataframe (see Table 6).

	var	type	inc	levelt	Pc	c	Pt	t	prob	omlo
1	openness1d	dt	-1.049	-5.71	0.947		0.547		0	0
2	openness1d	d	-1.038	-5.738	0.244		X	X	0	0
3	openness1d	-	-0.983	-5.584	X	X	X	X	0	0
4	stability1d	dt	-1.018	-6.238	0.54		0.77		0	0
5	stability1d	d	-1.023	-6.399	0.457		X	X	0	0
6	stability1d	-	-1.019	-6.42	X	X	X	X	0	0
7	exchangerate1d	dt	-0.605	-3.546	0.944		0.349		0.051	0
8	exchangerate1d	d	-0.563	-3.421	0.168		X	X	0.018	0
9	exchangerate1d	-	-0.478	-3.072	X	X	X	X	0.003	0
10	lnFDI1d	dt	-1.291	-8.159	0.085		0.504		0	0
11	lnFDI1d	d	-1.296	-8.278	0.022	s	X	X	0	0
12	lnFDI1d	-	-1.202	-7.376	X	X	X	X	0	0
13	lnGDP1d	dt	-2.519	-5.069	0	s	0.508		0.001	4
14	lnGDP1d	d	-2.462	-5.082	0	s	X	X	0	4
15	lnGDP1d	-	-0.179	-1.105	X	X	X	X	0.239	4
16	flight1d	dt	-0.756	-4.204	0.446		0.014	s	0.012	0
17	flight1d	d	-0.46	-3.008	0.034	s	X	X	0.045	0
18	flight1d	-	-0.269	-2.003	X	X	X	X	0.045	0

Table 6. The continuation of Table 5 from the right.

	var	type	inc	1stDifft	Pc	c	Pt	t	prob	omlo	interder
1	open1d	dt	-2.156	-7.795	0.843		0.813		0	1	
2	open1d	d	-2.151	-7.937	0.981		X	X	0	1	I(0)
3	open1d	-	-2.15	-8.082	X	X	X	X	0	1	
4	stab1d	dt	-3.943	-5.341	0.729		0.782		0.001	4	
5	stab1d	d	-3.949	-5.452	0.812		X	X	0	4	I(0)
6	stab1d	-	-3.947	-5.551	X	X	X	X	0	4	
7	exc1d	dt	-1.267	-6.937	0.961		0.953		0	0	
8	exc1d	d	-1.267	-7.06	0.836		X	X	0	0	I(0)
9	exc1d	-	-1.265	-7.173	X	X	X	X	0	0	
10	lnFDI1d	dt	-2.114	-7.556	0.843		0.896		0	1	
11	lnFDI1d	d	-2.116	-7.709	0.856		X	X	0	1	I(0)
12	lnFDI1d	-	-2.118	-7.86	X	X	X	X	0	1	
13	lnGDP1d	dt	-3.746	-5.704	0.887		0.865		0	4	
14	lnGDP1d	d	-3.739	-5.818	0.972		X	X	0	4	I(0)
15	lnGDP1d	-	-3.738	-5.935	X	X	X	X	0	4	
16	flight1d	dt	-2.005	-6.492	0.845		0.421		0	1	
17	flight1d	d	-1.967	-6.48	0.279		X	X	0	1	I(0)
18	flight1d	-	-1.894	-6.37	X	X	X	X	0	1	

As is seen from the tables, adfcstable correctly determines that all of the variables are stationary, i.e., integrated of order 0 (I(0)).

2.5.3 The dataset for the determinants of current account deficit in Turkey

We only take a few determinants of the current account deficit of Turkey since our aim here is not to determine those determinants explicitly in an extensive study, but to show how causfinder performs quick and seamless stationarity analysis. The dataframe V3CAD35obs.df has 3 variables (the ratio of current account deficit (minus sign indicates deficit) to gross domestic product (GDP), exchange rate and growth rate). These variables are shown as CA_GDP, exchangerate and growthrate:

```
R> V3CAD35obs.df
                                 CA_GDP exchangerate growthrate
                                  1980  -3.28    0.00007    -0.779
                                  1981  -1.99    0.00012     4.365
                                  1982  -1.10    0.00016     3.429
                                  1983  -2.32    0.00023     4.758
                                  1984  -1.78    0.00037     6.823
                                  1985  -1.12    0.00052     4.258
                                  1986  -1.46    0.00069     6.941
                                  1987  -0.70    0.00086    10.027
                                  1988   1.31    0.00143     2.121
                                  1989   0.67    0.00212     0.253
                                  1990  -1.30    0.00261     9.255
                                  1991  -0.04    0.00416     0.926
                                  1992  -0.46    0.00688     5.984
                                  1993  -3.20    0.01099     8.042
                                  1994   0.28    0.02972    -5.456
                                  1995  -2.37    0.04581     7.190
                                  1996  -1.00    0.08125     7.007
                                  1997  -1.03    0.15163     7.528
                                  1998   0.80    0.26047     3.092
                                  1999  -0.37    0.41783    -3.365
                                  2000  -3.72    0.62496     6.774
                                  2001   1.92    1.22528    -5.697
                                  2002  -0.27    1.50775     6.164
                                  2003  -2.49    1.50032     5.265
                                  2004  -3.62    1.42526     9.363
                                  2005  -4.44    1.34428     8.402
                                  2006  -6.01    1.43288     6.893
                                  2007  -5.84    1.30437     4.669
                                  2008  -5.53    1.30098     0.659
                                  2009  -1.97    1.55042    -4.826
                                  2010  -6.21    1.50204     9.157
                                  2011  -9.69    1.67505     8.773
                                  2012  -6.15    1.79659     2.127
                                  2013  -7.93    1.90430     4.050
                                  2014  -5.98    2.17137     3.026
```

Let’s perform adfcstable on V3CAD35obs.df:

```
R> adfcstable(V3CAD35obs.df)
```

Take the output in R to MS Word as described in the above step-by-step procedure:

```
R> write.csv(adfcstable(V3CAD35obs.df),file="V3CAD35obs.df.csv")
```

Table 7. The result of the adfcstable function for the variables in the V3CAD35obs.df dataframe (see Table 8).

	var	**type**	**inc**	**levelt**	**Pc**	**c**	**Pt**	**t**	**prob**	**omlo**
1	CA_GDP	dt	-0.857	-4.049	0.387		0.006	s	0.02	0
2	CA_GDP	d	-0.365	-2.298	0.059		X	X	0.18	0
3	CA_GDP	-	-0.148	-1.211	X	X	X	X	0.201	0
4	exchangerate	dt	-0.28	-2.528	0.169		0.016	s	0.313	1
5	exchangerate	d	-0.006	-0.144	0.206		X	X	0.934	1
6	exchangerate	-	0.033	1.066	X	X	X	X	0.92	1
7	growthrate	dt	-1.184	-5.72	0.026	s	0.917		0	0
8	growthrate	d	-1.185	-5.857	0.001	s	X	X	0	0
9	growthrate	-	-0.664	-3.434	X	X	X	X	0.001	0

Table 8. The continuation of Table 7 from the right.

	var	**type**	**inc**	**1stDifft**	**Pc**	**c**	**Pt**	**t**	**prob**	**omlo**	**intorder**
1	CA_GDP	dt	-2	-5.758	0.834		0.732		0	1	
2	CA_GDP	d	-1.995	-5.872	0.27		X	X	0	1	I(1)
3	CA_GDP	-	-1.902	-5.733	X	X	X	X	0	1	
4	exchangerate	dt	-0.617	-2.998	0.654		0.581		0.153	0	
5	exchangerate	d	-0.605	-3.001	0.097		X	X	0.049	0	I(1)
6	exchangerate	-	-0.44	-2.375	X	X	X	X	0.02	0	
7	growthrate	dt	-1.536	-8.732	0.877		0.917		0	0	
8	growthrate	d	-1.537	-8.947	0.891		X	X	0	0	I(0)
9	growthrate	-	-1.537	-9.149	X	X	X	X	0	0	

adfcstable correctly found the stationarities of the variables. CA_GDP is nonstationary (I(1)), exchange rate is nonstationary (I(1)), and growth rate is stationary (I(0)). One can check these stationarities via adfcs function or R's famous plot function etc.:

```
R> plot(as.ts(V3CAD35obs.df[,1]))
```

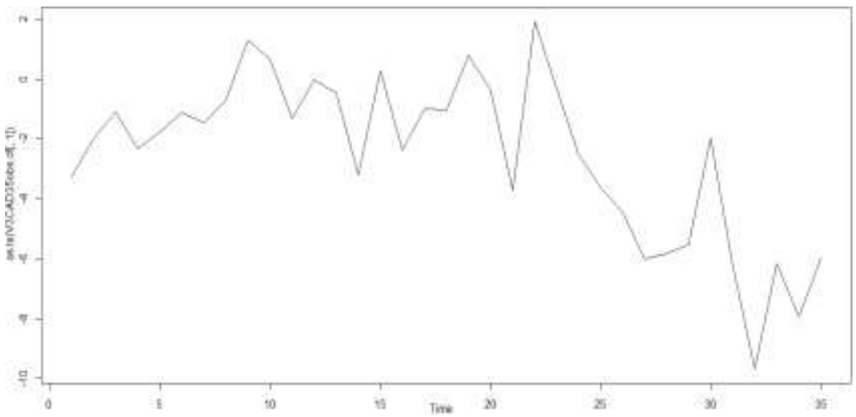

Fig. 3. V3CAD35obs.df[,1], i.e., CA/GDP.

```
R> plot(diff(as.ts(V3CAD35obs.df[,1]), differences=1))
```

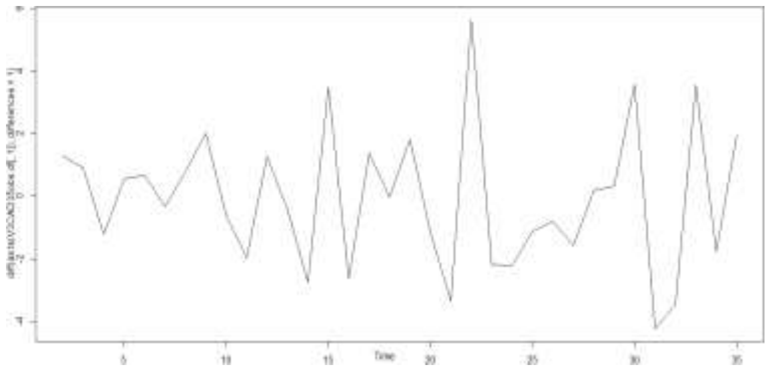

Fig. 4. Δ(CA/GDP).

```
R> plot(as.ts(V3CAD35obs.df[,2]))
```

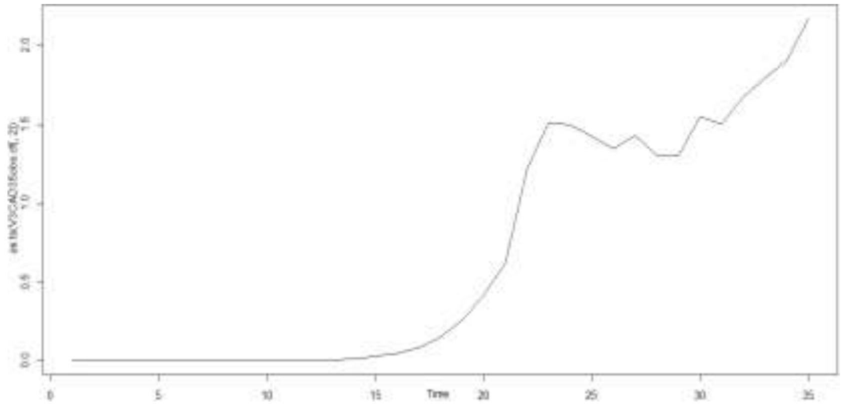

Fig. 5. Exchange rate.

```
R> plot(diff(as.ts(V3CAD35obs.df[,2]), differences=1))
```

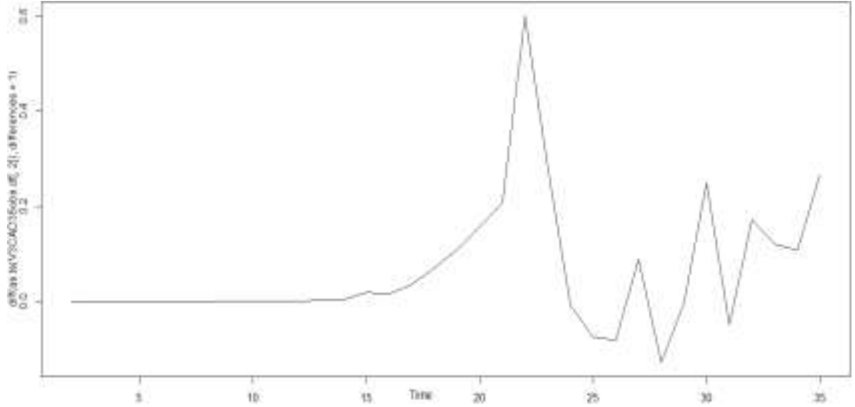

Fig. 6. Δ(Exchange rate).

```
R> plot(as.ts(V3CAD35obs.df[,3]))
```

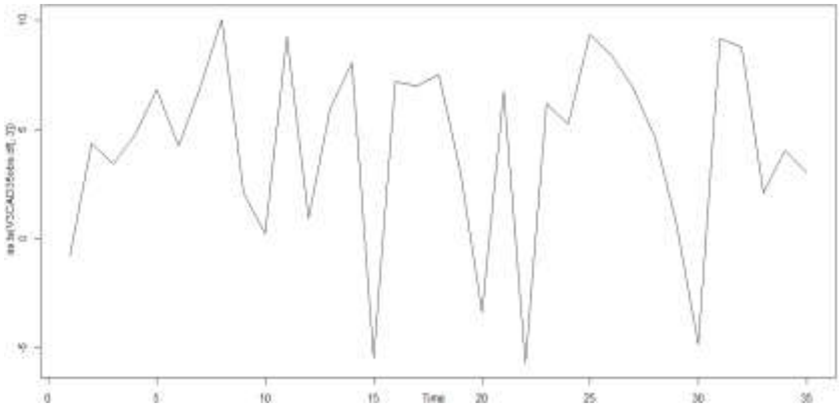

Fig. 7. Growth rate.

Since growth rate is already found to be stationary, we will not draw its difference.

3 CONCLUSION

We demonstrated how a very detailed stationarity analysis can be performed with only a single command. adfcstable function in causfinder package analyzes all the variables in a given system taking into account all the three ADF regression equations (none, only drift, both drift and time trend), and performs all the stationarity calculations of all the variables at once for the three mentioned cases. The operation time changed from 5 seconds to about 2 minutes in our datasets; the higher the number of observations the longer it took. We constructed adfcstable as one of the building blocks to reach the final aim of causfinder, i.e., to realize all the conditional and partial Granger causality analysis with only 1 click when a system is given. Since stationarity is a must for Granger causality when time-varying and windowing techniques are not employed, we think that we removed an obstacle in our way.

4 FUTURE SCOPE

Since R has the ability to perform parallelization, we believe that the calculations above will be even shorter, even less than 1 second in small samples when parallelization is employed. Our aim in next versions of causfinder is to reach that 1 second threshold.

Acknowledgments

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 causfinder package. Besides this, I want to thank a lot to those who helped me in theoretical sides of the related studies I conducted.

NOTE: Any work performed belongs to the author, and any view or opinion expressed in this paper does not necessarily represent those of the Ministry of Science, Industry and Technology of Turkey, or bind it in any way.

References

Cevher, Erdogan. (2014), causfinder: An R package for systemwise analysis of conditional and partial Granger causalities. International Journal of Science and Advanced Technology, 4(10), 6-15. http://www.ijsat.com/pdf.php?pdf_id=696.pdf

Roelstraete, B., & Rosseel, Y. (2011). FIAR: An R package for analyzing functional integration in the brain. *Journal of Statistical Software (JSS)*, 44(13). http://www.jstatsoft.org/v44/i13/paper

Schwert, G. W. 1989. Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics*, 2: 147–159. http://schwerts.com/jbes89.pdf

Erdogan Cevher is an industry and technology expert at the Ministry of Science, Industry and Technology, Ankara, Turkey since 2004. He received his BS degree in Mathematics from Bogazici University in 1999. After graduation, he worked in the Mathematics Dept. of Dogus University in 1999-2002. Then, he worked at Flashdesign Comp. Inc., Istanbul in 2002-2003. He received the expert title with his thesis “An Interoperable Minimal System Infrastructure among Public Institutions in the Scope of eTransformation and a Sample Model Proposal”. His current research interests include causality, especially conditional and partial Granger causality, transfer entropy, network topologies, and parallelization in R.