The Stata Journal (2017)
17, Number 4, pp. 972–984

Testing for Granger causality in panel data

Luciano Lopez
University of Neuchâtel
Neuchâtel, Switzerland
luciano.lopez@unine.ch

Sylvain Weber
University of Neuchâtel
Neuchâtel, Switzerland
sylvain.weber@unine.ch

Abstract. With the development of large and long panel databases, the theory surrounding panel causality evolves quickly, and empirical researchers might find it difficult to run the most recent techniques developed in the literature. In this article, we present the community-contributed command `xtgcause`, which implements a procedure proposed by Dumitrescu and Hurlin (2012, *Economic Modelling* 29: 1450–1460) for detecting Granger causality in panel datasets. Thus, it constitutes an effort to help practitioners understand and apply the test. `xtgcause` offers the possibility of selecting the number of lags to include in the model by minimizing the Akaike information criterion, Bayesian information criterion, or Hannan–Quinn information criterion, and it offers the possibility to implement a bootstrap procedure to compute p-values and critical values.

Keywords: st0507, xtgcause, Granger causality, panel datasets, bootstrap

1 Introduction

Panel datasets composed of many individuals and many time periods are becoming widely available. A particularly salient case is the growing availability of cross-country data over time. As a consequence, the focus of panel-data econometrics is shifting from micropanels, with large N and small T, to macropanels, with large N and large T. In this setting, classical issues of time-series econometrics, such as (non)stationarity and (non)causality, also arise. In this article, we present the community-contributed command `xtgcause`, which implements a procedure developed by Dumitrescu and Hurlin (2012) (the DH test) to test for Granger causality in panel datasets.

Considering the fast evolution of the literature, practitioners may find it difficult to implement the latest econometric tests. Therefore, in this article, we summarize the test built by Dumitrescu and Hurlin (2012) and present `xtgcause` using examples based on simulated and real data. The objective of our contribution is to support the empirical literature using panel causality techniques. One recurrent concern is the selection of the number of lags to be included in the estimations; we have implemented an extension of the test based on Akaike information criteria (AIC), Bayesian information criteria (BIC, and Hannan–Quinn information criteria (HQIC) to facilitate this task. Finally, to deal with the empirical issue of cross-sectional dependence, we have implemented an option to compute p-values and critical values based on a bootstrap procedure.

 st0507

2 The Dumitrescu–Hurlin test

In a seminal article, Granger (1969) developed a methodology for analyzing the causal relationships between time series. Suppose x_t and y_t are two stationary series. The model

$$y_t = \alpha + \sum_{k=1}^{K} \gamma_k y_{t-k} + \sum_{k=1}^{K} \beta_k x_{t-k} + \varepsilon_t \qquad \text{with } t = 1, \ldots, T \tag{1}$$

can then be used to test whether x causes y. Essentially, if past values of x are significant predictors of the current value of y even when past values of y have been included in the model, then x exerts a causal influence on y. Using (1), one might easily investigate this causality based on an F test with the following null hypothesis:

$$H_0\colon \beta_1 = \cdots = \beta_K = 0$$

If H_0 is rejected, one can conclude that causality from x to y exists. The x and y variables can be interchanged to test for causality in the other direction, and it is possible to observe bidirectional causality (also called feedback).

Dumitrescu and Hurlin (2012) provide an extension designed to detect causality in panel data. The underlying regression is

$$y_{i,t} = \alpha_i + \sum_{k=1}^{K} \gamma_{ik} y_{i,t-k} + \sum_{k=1}^{K} \beta_{ik} x_{i,t-k} + \varepsilon_{i,t} \qquad \text{with } i = 1, \ldots, N \text{ and } t = 1, \ldots, T \tag{2}$$

where $x_{i,t}$ and $y_{i,t}$ are the observations of two stationary variables for individual i in period t. Coefficients are allowed to differ across individuals (note the i subscripts attached to coefficients) but are assumed to be time invariant. The lag order K is assumed to be identical for all individuals, and the panel must be balanced.

As in Granger (1969), the procedure to determine the existence of causality is to test for significant effects of past values of x on the present value of y. The null hypothesis is therefore defined as

$$H_0\colon \beta_{i1} = \cdots = \beta_{iK} = 0 \qquad \forall\, i = 1, \ldots, N \tag{3}$$

which corresponds to the absence of causality for all individuals in the panel.

The DH test assumes there can be causality for some individuals but not necessarily for all. Thus, the alternative hypothesis is

$$\begin{aligned} H_1\colon & \beta_{i1} = \cdots = \beta_{iK} = 0 && \forall\, i = 1, \ldots, N_1 \\ & \beta_{i1} \neq 0 \text{ or } \ldots \text{ or } \beta_{iK} \neq 0 && \forall\, i = N_1 + 1, \ldots, N \end{aligned}$$

where $N_1 \in [0, N-1]$ is unknown. If $N_1 = 0$, there is causality for all individuals in the panel. N_1 must be strictly smaller than N; otherwise, there is no causality for all individuals, and H_1 reduces to H_0.

Against this backdrop, Dumitrescu and Hurlin (2012) propose the following procedure: run the N individual regressions implicitly enclosed in (2), perform F tests of the

K linear hypotheses $\beta_{i1} = \cdots = \beta_{iK} = 0$ to retrieve the individual Wald statistic W_i, and finally compute the average Wald statistic $\overline{W}$:[1]

$$\overline{W} = \frac{1}{N} \sum_{i=1}^{N} W_i$$

We emphasize that the test is designed to detect causality at the panel level, and rejecting H_0 does not exclude noncausality for some individuals. Using Monte Carlo simulations, Dumitrescu and Hurlin (2012) show that $\overline{W}$ is asymptotically well behaved and can genuinely be used to investigate panel causality.

Under the assumption that the Wald statistics W_i are independently and identically distributed across individuals, it can be shown that the standardized statistic $\overline{Z}$ when $T \to \infty$ first and then $N \to \infty$ (sometimes interpreted as "T should be large relative to N") follows a standard normal distribution:

$$\overline{Z} = \sqrt{\frac{N}{2K}} \times \left(\overline{W} - K\right) \quad \xrightarrow[T,N\to\infty]{d} \quad \mathcal{N}(0,1) \tag{4}$$

Also, for a fixed T dimension with $T > 5 + 3K$, the approximated standardized statistic $\widetilde{Z}$ follows a standard normal distribution:

$$\widetilde{Z} = \sqrt{\frac{N}{2K} \times \frac{T-3K-5}{T-2K-3}} \times \left(\frac{T-3K-3}{T-3K-1} \times \overline{W} - K\right) \quad \xrightarrow[N\to\infty]{d} \quad \mathcal{N}(0,1) \tag{5}$$

The testing procedure of the null hypothesis in (3) is finally based on $\overline{Z}$ and $\widetilde{Z}$. If these are larger than the standard critical values, then one should reject H_0 and conclude that Granger causality exists. For large N and T panel datasets, $\overline{Z}$ can be reasonably considered. For large N but relatively small T datasets, $\widetilde{Z}$ should be favored. Using Monte Carlo simulations, Dumitrescu and Hurlin (2012) have shown that the test exhibits good finite sample properties, even when both T and N are small.

The lag order (K) selection is an empirical issue for which Dumitrescu and Hurlin (2012) provide no guidance. One way to tackle this issue is to select the number of lags based on an information criterion (AIC/BIC/HQIC). In this process, all estimations have to be conducted on a common sample to be nested and therefore comparable.[2] Practically, this implies that the first $K_{\max}$[3] time periods must be omitted during the entire lag-selection process.

Another empirical issue to consider in panel data is cross-sectional dependence. To this end, a block bootstrap procedure is proposed in section 6.2 of Dumitrescu and

1. See Dumitrescu and Hurlin (2012, 1453) for the mathematical definition of W_i. Note, however, that T in their formulas must be understood as the number of observations remaining in the estimations, that is, the number of periods minus the number of lags included. To be consistent with our notation, we therefore replaced Dumitrescu and Hurlin's (2012) T by $T-K$ in the following formulas of the present article.
2. We thank Gareth Thomas (IHS Markit EViews) for bringing this point to our attention.
3. $K_{\max}$ stands for the maximum possible number of lags to be considered in the entire procedure.

Hurlin (2012) to compute bootstrapped critical values for $\overline{Z}$ and $\widetilde{Z}$ instead of asymptotic critical values. The procedure has the following steps:[4]

1. Fit (2) and obtain $\overline{Z}$ and $\widetilde{Z}$ as defined in (4) and (5).

2. Fit the model under $H_0: y_{i,t} = \alpha_i^0 + \sum_{k=1}^K \gamma_{ik}^0 y_{i,t-k} + \varepsilon_{i,t}$, and collect the residuals in matrix $\widehat{\boldsymbol{\varepsilon}}_{(T-K)\times N}$.

3. Build a matrix $\boldsymbol{\varepsilon}^\star_{(T-K)\times N}$ by resampling (overlapping blocks of) rows (that is, time periods) of matrix $\widehat{\boldsymbol{\varepsilon}}$. Block bootstrap is useful when there is autocorrelation.

4. Generate a random draw $(\mathbf{y_1^\star}, \dots, \mathbf{y_K^\star})'$, with $\mathbf{y_t^\star} = \left(y_{1,t}^\star, y_{2,t}^\star, \cdots, y_{N,t}^\star\right)$, by randomly selecting a block of K consecutive time periods with replacement (see Stine [1987] and Berkowitz and Kilian [2000]).

5. Construct the resampled series $y_{i,t}^\star = \widehat{\alpha}_i^0 + \sum_{k=1}^K \widehat{\beta}_{ik}^0 y_{i,t-k}^\star + \varepsilon_{i,t}^\star$ conditional on the random draw for the first K periods.

6. Fit the model $y_{i,t}^\star = \alpha_i^b + \sum_{k=1}^K \gamma_{ik}^b y_{i,t-k}^\star + \sum_{k=1}^K \beta_{ik}^b x_{i,t-k} + \varepsilon_{i,t}$ and compute $\overline{Z}^b$ and $\widetilde{Z}^b$.

7. Run B replications of steps 3 to 6.

8. Compute p-values and critical values for $\overline{Z}$ and $\widetilde{Z}$ based on the distributions of $\overline{Z}^b$ and $\widetilde{Z}^b$, $b = 1, \dots, B$.

3 The xtgcause command

3.1 Syntax

The syntax of `xtgcause` is

`xtgcause` *depvar indepvar* [*if*] [*in*] [, `lags(#|aic [#]|bic [#]|hqic [#])` `regress bootstrap breps(#) blevel(#) blength(#) seed(#) nodots`]

3.2 Options

`lags(#|aic [#]|bic [#]|hqic [#])` specifies the lag structure to use for the regressions performed in computing the test statistic. The default is `lags(1)`.

4. The procedure we present here differs slightly from that proposed by Dumitrescu and Hurlin (2012) in the numbering of the steps but, more importantly, in the definition of the initial conditions (our step 4), which is not addressed by Dumitrescu and Hurlin (2012), and in the construction of the resampled series (our step 5). We are indebted to David Ardia (University of Neuchâtel) for his valuable advice on the bootstrap procedure.

Specifying `lags(#)` requests that # lags of the series be used in the regressions. The maximum authorized number of lags is such that $T > 5 + 3 \times \#$.

Specifying `lags(aic|bic|hqic` [#]`)` requests that the number of lags of the series be chosen such that the average AIC/BIC/HQIC for the set of regressions is minimized. Regressions with 1 to # lags will be conducted, restricting the number of observations to $T - \#$ for all estimations to make the models nested and therefore comparable. Displayed statistics come from the set of regressions for which the average AIC/BIC/HQIC is minimized (reestimated using the total number of observations available). If # is not specified in `lags(aic|bic|hqic` [#]`)`, then it is set to the maximum number of lags authorized.

`regress` can be used to display the results of the N individual regressions on which the test is based. This option is useful to have a look at the coefficients of individual regressions. When the number of individuals in the panel is large, this option will result in a very long output.

`bootstrap` requests p-values and critical values be computed using a bootstrap procedure as proposed in Dumitrescu and Hurlin (2012, sec. 6.2). The bootstrap procedure is useful when there is cross-sectional dependence.

`breps(#)` indicates the number of bootstrap replications to perform. The default is `breps(1000)`.

`blevel(#)` indicates the significance level (in %) for computing the bootstrapped critical values. The default is `blevel(95)`.

`blength(#)` indicates the size of the block length to be used in the bootstrap. By default, each time period is sampled independently with replacement `blength(1)`. `blength()` allows the user to implement the bootstrap by dividing the sample into overlapping blocks of # time periods and sampling the blocks independently with the replacement. Using blocks of more than one time period is useful if autocorrelation is suspected.

`seed(#)` can be used to set the random-number seed. By default, the seed is not set.

`nodots` suppresses replication dots. By default, a dot is printed for each replication to provide an indication of the evolution of the bootstrap.

`breps()`, `blevel()`, `blength()`, `seed()`, and `nodots` are `bootstrap` suboptions. They can be used only if `bootstrap` is also specified.

3.3 Stored results

`xtgcause` stores the following in `r()`:

Scalars

`r(wbar)`	average Wald statistic	`r(zbart)`	`Z-bar tilde` statistic
`r(lags)`	number of lags used for the test	`r(zbart_pv)`	p-value of the `Z-bar tilde` statistic
`r(zbar)`	`Z-bar` statistic		
`r(zbar_pv)`	p-value of the `Z-bar` statistic		

Matrices

`r(Wi)`	individual Wald statistics	`r(PVi)`	p-values of the individual Wald statistics

`xtgcause` with the `bootstrap` option also stores the following additional results in `r()`:

Scalars

`r(zbarb_cv)`	critical value for the `Z-bar` statistic	`r(blevel)`	significance level for bootstrap critical values
`r(zbartb_cv)`	critical value for the `Z-bar tilde` statistic	`r(blength)`	size of the block length
`r(breps)`	number of bootstrap replications		

Matrices

`r(ZBARb)`	`Z-bar` statistics from the bootstrap procedure	`r(ZBARTb)`	`Z-bar tilde` statistics from the bootstrap procedure

4 Examples

Before presenting some examples, we recall that the test implemented in `xtgcause` assumes that the variables are stationary. We will not go through this first step here, but it is the user's responsibility to verify that the data satisfy this condition. To this end, the user might consider `xtunitroot`, which provides various panel stationarity tests with alternative null hypotheses (in particular, Breitung [2000]; Hadri [2000]; Harris and Tzavalis [1999]; Im, Pesaran, and Shin [2003]; Levin, Lin, and Chu [2002]). The user may also want to perform second-generation panel unit-root tests such as the one proposed by Pesaran (2007) to control for cross-sectional dependence.

4.1 Example based on simulated data

To illustrate `xtgcause`, we first use simulated data provided by Dumitrescu and Hurlin (2012) at http://www.execandshare.org/execandshare/htdocs/data/MetaSite/upload/companionSite51/data/ in the file `data-demo.csv`.[5] Then, we import the original Excel dataset directly from the website. In the original CSV file, the dataset is organized as a matrix, with all observations for each individual in a single cell. Within this cell, the 10 values of variable x are separated by tabs, a comma separates the last value of x and the first value of y, and the 10 values of variable y are then separated by tabs. Hence, the following lines of code allow shaping the data to be understood as a panel by Stata.

5. Data and MATLAB code are also available at http://www.runmycode.org/companion/view/42 in a ZIP file.

```
. import delimited using "http://www.execandshare.org/execandshare/htdocs/data/
> MetaSite/upload/companionSite51/data/data-demo.csv", delimiter(",")
> colrange(1:2) varnames(1)
(2 vars, 20 obs)

. quietly: split x, parse(`=char(9)´) destring

. quietly: split y, parse(`=char(9)´) destring

. drop x y

. generate t = _n

. reshape long x y, i(t) j(id)
(note: j = 1 2 3 4 5 6 7 8 9 10)
Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                       20   ->     200
Number of variables                  21   ->       4
j variable (10 values)                    ->   id
xij variables:
                          x1 x2 ... x10   ->   x
                          y1 y2 ... y10   ->   y
-----------------------------------------------------------------------------

. xtset id t
       panel variable:  id (strongly balanced)
        time variable:  t, 1 to 20
                delta:  1 unit

. list id t x y in 1/5

     +------------------------------------+
     | id   t            x             y  |
     |------------------------------------|
  1. |  1   1    .55149203     .81872837  |
  2. |  1   2    .64373514    -.42077179  |
  3. |  1   3   -.58843258    -.40312278  |
  4. |  1   4   -.55873336     .14674849  |
  5. |  1   5   -.32486386     .42924677  |
     +------------------------------------+

. list id t x y in 21/25

     +------------------------------------+
     | id   t            x             y  |
     |------------------------------------|
 21. |  2   1   -1.4703536     1.2586422  |
 22. |  2   2    1.3356281    -.71173904  |
 23. |  2   3   -.21564623    -.73264199  |
 24. |  2   4    .08435614    -.67841901  |
 25. |  2   5    1.5766581     -.2562083  |
     +------------------------------------+
```

Some sections of the above code are quite involved and require explanations. We started by importing the data as if values were separated by commas, which is only partly true. This created two string variables, named `x` and `y`, each containing 10 values (separated by tabs) in each observation. We then invoked `split`, using `char(9)` (which indeed corresponds to a tab) as the parse string. We used the prefix `quietly` to avoid a long output indicating that 2 sets of 10 variables ($x_1, \ldots, x_{10}$, and $y_1, \ldots, y_{10}$) were created. These variables were immediately converted from string to numeric thanks to `split`'s `destring` option. To have a well-shaped panel that Stata can correctly interpret, we combined these 2 sets of 10 variables into only 2 variables using `reshape`.

A few observations (the first five for individuals 1 and 2) are displayed to show how the data are finally organized.

After we format and `xtset` the data, we can now run `xtgcause`. The simplest possible test to investigate whether x causes y would be

```
. xtgcause y x

Dumitrescu & Hurlin (2012) Granger non-causality test results:
--------------------------------------------------------------
Lag order: 1
W-bar =          1.2909
Z-bar =          0.6504   (p-value = 0.5155)
Z-bar tilde =    0.2590   (p-value = 0.7956)
--------------------------------------------------------------
H0: x does not Granger-cause y.
H1: x does Granger-cause y for at least one panelvar (id).
```

Because we did not specify any lag order, `xtgcause` introduced a single lag by default. In this case, the outcome of the test does not reject the null hypothesis. The output reports the values obtained for $\overline{W}$ (`W-bar`), $\overline{Z}$ (`Z-bar`), and $\widetilde{Z}$ (`Z-bar tilde`). For the latter two statistics, p-values are provided based on the standard normal distribution.

One could additionally display the individual Wald statistics and their corresponding values by displaying the stored matrices `r(Wi)` and `r(PVi)` (which we first combine into a single matrix for the sake of space):

```
. matrix Wi_PVi = r(Wi), r(PVi)

. matrix list Wi_PVi

Wi_PVi[10,2]
             Wi        PVi
 id1  .56655945  .46256089
 id2  .11648998  .73731411
 id3  .09081952  .76701924
 id4  8.1263612  .01156476
 id5  .18687517  .67129995
 id6  .80060395  .38417583
 id7  .53075859  .47681675
 id8  .00158371  .96874825
 id9  .43635413   .5182858
id10  2.0521113  .17124367
```

Using the `lags()` option, we run a similar test introducing two lags of the variables `x` and `y`:

```
. xtgcause y x, lags(2)

Dumitrescu & Hurlin (2012) Granger non-causality test results:
--------------------------------------------------------------
Lag order: 2
W-bar =          1.7302
Z-bar =         -0.4266   (p-value = 0.6696)
Z-bar tilde =   -0.7052   (p-value = 0.4807)
--------------------------------------------------------------
H0: x does not Granger-cause y.
H1: x does Granger-cause y for at least one panelvar (id).
```

The conclusion of the test is similar to before.

Alternatively, the test could also be conducted using a bootstrap procedure to compute p-values and critical values:

```
. xtgcause y x, bootstrap lags(1) breps(100) seed(20171020)
----------------------------
Bootstrap replications (100)
----------------------------
..................................................    50
..................................................   100

Dumitrescu & Hurlin (2012) Granger non-causality test results:
--------------------------------------------------------------
Lag order: 1
W-bar =          1.2909
Z-bar =          0.6504   (p-value* = 0.4700, 95% critical value = 1.7316)
Z-bar tilde =    0.2590   (p-value* = 0.7100, 95% critical value = 1.3967)
--------------------------------------------------------------
H0: x does not Granger-cause y.
H1: x does Granger-cause y for at least one panelvar (id).
*p-values computed using 100 bootstrap replications.
```

In this case, the bootstrapped p-values are relatively close to the asymptotic ones displayed in the first test above.

4.2 Example based on real data

To provide an example based on real data, we searched for articles reporting Dumitrescu and Hurlin's (2012) tests and published in journals that make authors' datasets available. We found several such articles (for example, Paramati, Ummalla, and Apergis [2016]; Paramati, Apergis, and Ummalla [2017]; Salahuddin, Alam, and Ozturk [2016]). In particular, Paramati, Ummalla, and Apergis (2016) investigate the effect of foreign direct investment and stock market growth on clean energy use. In their table 8, they report a series of pairwise panel causality tests between variables such as economic output, CO_2 emissions, or foreign direct investment. As indicated in their online supplementary data (file `Results.xlsx`), they conduct the tests using EViews 8. We replicate some of their results:

```
. import excel using data-wdi.xlsx, clear first case(lower) cellrange(A1:I421)
> sheet(FirstDif-Data)
. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 1992 to 2012
                delta:  1 unit
. xtgcause co2 output, lags(2)
Dumitrescu & Hurlin (2012) Granger non-causality test results:
--------------------------------------------------------------
Lag order: 2
W-bar =          2.4223
Z-bar =          0.9442   (p-value = 0.3451)
Z-bar tilde =    0.1441   (p-value = 0.8855)
--------------------------------------------------------------
H0: output does not Granger-cause co2.
H1: output does Granger-cause co2 for at least one panelvar (id).
. xtgcause fdi output, lags(2)
Dumitrescu & Hurlin (2012) Granger non-causality test results:
--------------------------------------------------------------
Lag order: 2
W-bar =          4.6432
Z-bar =          5.9103   (p-value = 0.0000)
Z-bar tilde =    3.7416   (p-value = 0.0002)
--------------------------------------------------------------
H0: output does not Granger-cause fdi.
H1: output does Granger-cause fdi for at least one panelvar (id).
```

The above code imports the dataset constructed by Paramati, Ummalla, and Apergis (2016) (file `Data-WDI.xlsx`, sheet `FirstDif-Data`) as a first step. We then use `xtgcause` to test for the causality from `output` to `co2` and from `output` to `fdi`, which correspond to some tests reported in their table 8. We use two lags in both cases to match the numbers indicated by Paramati, Ummalla, and Apergis (2016) in their accompanying appendix file. Compared with their output, it turns out that the denomination "Zbar-Stat" used in EViews corresponds to the `Z-bar tilde` statistic (while the `Z-bar` statistic is not provided in EViews).

Optionally, `xtgcause` allows the user to request the lag order to be chosen so that the AIC, BIC, or HQIC be minimized. Given that Dumitrescu and Hurlin (2012) offer no guidance regarding the choice of the lag order, this feature might be appealing to practitioners. We can, for example, test the causality from `output` to `fdi` specifying the option `lags(bic)`:

```
. xtgcause fdi output, lags(bic)
Dumitrescu & Hurlin (2012) Granger non-causality test results:
--------------------------------------------------------------
Optimal number of lags (BIC): 1 (lags tested: 1 to 5).
W-bar =          1.3027
Z-bar =          0.9572   (p-value = 0.3385)
Z-bar tilde =    0.4260   (p-value = 0.6701)
--------------------------------------------------------------
H0: output does not Granger-cause fdi.
H1: output does Granger-cause fdi for at least one panelvar (id).
```

In practice, `xtgcause` runs all sets of regressions with a lag order from 1 to the highest possible number (that is, such that $T > 5 + 3K$ or optionally specified by the user below this limit), maintaining a common sample. Said otherwise, if at most five lags are to be considered, the first five observations of the panel will never be considered in the estimations, even if it would be possible to do so with fewer than five lags. This ensures nested models, which can then be appropriately compared using AIC, BIC, or HQIC. After this series of estimations, `xtgcause` selects the optimal outcome (that is, such that the average AIC/BIC/HQIC of the N individual estimations is the lowest) and reruns all estimations with the optimal number of lags and using all observations available. Statistics based on the latter are reported as output.

In the above example, the optimal lag order using BIC appears to be 1, which is different from the lag order selected by Paramati, Ummalla, and Apergis (2016) for this test.[6] This difference is not without consequences, because the conclusion of the test is then reversed. More precisely, the null hypothesis is not rejected with the optimally selected single lag, but Paramati, Ummalla, and Apergis (2016) use two lags and therefore reject the null hypothesis. Considering that empirical research in economics is used to formulate policy recommendations, such inaccurate conclusions may potentially be harmful. We therefore consider `xtgcause`'s option allowing the user to select the number of lags based on AIC/BIC/HQIC as an important improvement. It will allow researchers to rely on these widely accepted criteria and make the selection in a transparent way.

Finally, `xtgcause` makes it possible to compute the p-values and critical values associated with the `Z-bar` and `Z-bar tilde` via a bootstrap procedure. Computing bootstrapped critical values (rather than asymptotic ones) may be useful when there is cross-sectional dependence. Based on the Paramati, Ummalla, and Apergis (2016) data, we test the causality from `output` to `fdi` by adding the `bootstrap` option (we also use `seed` for replicability reasons and `nodots` for the sake of space):

```
. xtgcause fdi output, lags(bic) bootstrap seed(20171020) nodots
----------------------------
Bootstrap replications (1000)
----------------------------

Dumitrescu & Hurlin (2012) Granger non-causality test results:
--------------------------------------------------------------
Optimal number of lags (BIC): 1 (lags tested: 1 to 5).
W-bar =          1.3027
Z-bar =          0.9572   (p-value* = 0.4530, 95% critical value = 3.0746)
Z-bar tilde =    0.4260   (p-value* = 0.7080, 95% critical value = 2.1234)
--------------------------------------------------------------
H0: output does not Granger-cause fdi.
H1: output does Granger-cause fdi for at least one panelvar (id).
*p-values computed using 1000 bootstrap replications.
```

Here `xtgcause` first computes the `Z-bar` and `Z-bar tilde` statistics using the optimal number of lags as in previous series of estimations; then, it computes the boot-

6. The number of lags would be three using HQIC and four using AIC. Therefore, while Paramati, Ummalla, and Apergis (2016) state in their table 8 that "the appropriate lag length is chosen based on SIC", we do not find the same number with any of the information criterion considered.

strapped p-values and critical values. By default, 1,000 bootstrap replications are performed. We observe that the bootstrapped p-value for the `Z-bar` increases substantially compared with the asymptotic p-value obtained before (from 0.34 to 0.45), while that for the `Z-bar tilde` remains closer. This should be interpreted as a signal that the estimations suffer from small sample biases, so asymptotic p-values are underestimated. Bootstrapped p-values indicate that the null hypothesis is far from being rejected, which strengthens the concerns about Paramati, Ummalla, and Apergis's (2016) conclusions based on the asymptotic p-values and obtained with two lags.

5 Conclusion

In this article, we presented the community-contributed command `xtgcause`, which automates a procedure introduced by Dumitrescu and Hurlin (2012) to detect Granger causality in panel datasets. In this branch of econometrics, the empirical literature appears to be lagging, with the latest theoretical developments not always being available in statistical packages. One important contribution of our command is to allow the user to select the number of lags based on the AIC, the BIC, or the HQIC. This choice may impact the conclusion of the test, but some researchers may have overlooked it. Thus, several empirical articles might have reached erroneous conclusions. `xtgcause` also allows the user to calculate bootstrapped critical values, which is a useful option when there is cross-sectional dependence. With this command and this article, we hope to bring some useful clarifications and help practitioners conduct sound research.

6 References

Berkowitz, J., and L. Kilian. 2000. Recent developments in bootstrapping time series. *Econometric Reviews* 19: 1–48.

Breitung, J. 2000. The local power of some unit root tests for panel data. In *Advances in Econometrics: Vol. 15—Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, ed. B. H. Baltagi, 161–178. New York: Elsevier.

Dumitrescu, E.-I., and C. Hurlin. 2012. Testing for Granger non-causality in heterogeneous panels. *Economic Modelling* 29: 1450–1460.

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424–438.

Hadri, K. 2000. Testing for stationarity in heterogeneous panel data. *Econometrics Journal* 3: 148–161.

Harris, R. D. F., and E. Tzavalis. 1999. Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 91: 201–226.

Im, K. S., M. H. Pesaran, and Y. Shin. 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115: 53–74.

Levin, A., C.-F. Lin, and C.-S. J. Chu. 2002. Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* 108: 1–24.

Paramati, S. R., N. Apergis, and M. Ummalla. 2017. Financing clean energy projects through domestic and foreign capital: The role of political cooperation among the EU, the G20 and OECD countries. *Energy Economics* 61: 62–71.

Paramati, S. R., M. Ummalla, and N. Apergis. 2016. The effect of foreign direct investment and stock market growth on clean energy use across a panel of emerging market economies. *Energy Economics* 56: 29–41.

Pesaran, M. H. 2007. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics* 22: 265–312.

Salahuddin, M., K. Alam, and I. Ozturk. 2016. The effects of Internet usage and economic growth on CO_2 emissions in OECD countries: A panel investigation. *Renewable and Sustainable Energy Reviews* 62: 1226–1235.

Stine, R. A. 1987. Estimating properties of autoregressive forecasts. *Journal of the American Statistical Association* 82: 1072–1078.

About the authors

Luciano Lopez is a PhD student at the University of Neuchâtel (Switzerland). He is currently writing a PhD thesis in development economics.

Sylvain Weber is a postdoctoral researcher at the Institute of Economic Research of the University of Neuchâtel (Switzerland). His main fields of research are labor and energy economics, and he is interested in microeconometrics in general.