Testing for Granger causality in panel data

Luciano Lopez

Sylvain Weber

uni**n**e

UNIVERSITÉ DE
NEUCHÂTEL

Institut de
recherches économiques

Testing for Granger causality in panel data

Luciano Lopez[*]   Sylvain Weber[†]

November 3, 2017

Abstract

With the development of large and long panel databases, the theory surrounding panel causality evolves at a fast pace and empirical researchers may sometimes find it difficult to run the most recent techniques developed in the literature. This article presents the Stata user-written command `xtgcause`, which implements a procedure proposed by Dumitrescu and Hurlin (2012) for detecting Granger causality in panel datasets, and thus constitutes an effort to help practitioners understand and apply the test. The command offers the possibility to select the number of lags to include in the model by minimizing the AIC, BIC, or HQIC, and to implement a bootstrap procedure to compute p-values and critical values.

JEL Classification: C23, C87.

Keywords: Stata, Granger causality, panel datasets, bootstrap.

---

[*]University of Neuchâtel, Institute of Economic Research, Rue Abram-Louis Breguet 2, 2000 Neuchâtel, Switzerland. luciano.lopez@unine.ch.

[†]University of Neuchâtel, Institute of Economic Research, Rue Abram-Louis Breguet 2, 2000 Neuchâtel, Switzerland. sylvain.weber@unine.ch.

1 Introduction

Panel datasets comprised 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 micro panel, with large N and small T, to macro panels, where both N and T are large. In this setting, classical issues of time-series econometrics, such as (non-)stationarity and (non-)causality, also arise. This paper discusses the user-written command `xtgcause`, which implements a procedure recently developed by Dumitrescu and Hurlin (2012) (hereafter DH) in order 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. In this paper, we therefore summarize the test built by DH 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 being related to the selection of the number of lags to be included in the estimations, we have implemented an extension of the test based on Akaike, Bayesian, and Hannan-Quinn information criteria to facilitate this task. Finally, and 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.

2 The Dumitrescu-Hurlin test

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

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

can be used to test whether x causes y. The basic idea is that 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:

$$H_0: \ \beta_1 = ... = \beta_K = 0 \qquad (2)$$

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

DH provide an extension designed to detect causality in panel data. The underlying regression writes:

$$y_{i,t} = \alpha_i + \sum_{k=1}^{K} \gamma_{ik} y_{i,t-k} + \sum_{k=1}^{K} \beta_{ik} x_{i,t-k} + \varepsilon_{i,t} \qquad \text{with } i = 1, ..., N \text{ and } t = 1, ..., T \tag{3}$$

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 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:

$$H_0: \ \beta_{i1} = ... = \beta_{iK} = 0 \qquad \forall\, i = 1, ..., N \tag{4}$$

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

DH test assumes there can be causality for some individuals but not necessarily for all. The alternative hypothesis thus writes:

$$H_1: \ \beta_{i1} = ... = \beta_{iK} = 0 \qquad \forall\, i = 1, ..., N_1$$
$$\beta_{i1} \neq 0 \text{ or ... or } \beta_{iK} \neq 0 \quad \forall\, i = N_1 + 1, ..., N$$

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, DH propose the following procedure: run the N individual regressions implicitly enclosed in (3), perform F-tests of the K linear hypotheses $\beta_{i1} = ... = \beta_{iK} = 0$ to retrieve the individual Wald statistic W_i, and finally compute the average Wald statistic $\overline{W}$:[1]

$$\overline{W} = \frac{1}{N} \sum_{i=1}^{N} W_i \tag{5}$$

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

---

[1]See Dumitrescu and Hurlin (2012, p. 1453) for the mathematical definition of W_i. Note however that T in DH's 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. In order to be consistent with our notation, we therefore replaced DH's T by $T-K$ in the following formulas of the present paper.

Under the assumption that the Wald statistics W_i are independently and identically distributed across individuals, it can be showed that the standardized statistic $\bar{Z}$ 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:

$$\bar{Z} = \sqrt{\frac{N}{2K}} \cdot \left(\overline{W} - K\right) \quad \xrightarrow[T,N\to\infty]{d} \quad \mathcal{N}(0,1) \tag{6}$$

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

$$\widetilde{Z} = \sqrt{\frac{N}{2K} \cdot \frac{T-3K-5}{T-2K-3}} \cdot \left[\frac{T-3K-3}{T-3K-1} \cdot \overline{W} - K\right] \quad \xrightarrow[N\to\infty]{d} \quad \mathcal{N}(0,1) \tag{7}$$

The testing procedure of the null hypothesis in (4) is finally based on $\bar{Z}$ 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, $\bar{Z}$ can be reasonably considered. For large N but relatively small T datasets, $\widetilde{Z}$ should be favored. Using Monte Carlo simulations, DH have shown that the test exhibits very good finite sample properties, even with both T and N small.

The lag order (K) selection is an empirical issue for which DH 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 in order 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 that of cross-sectional dependence. To this end, a block bootstrap procedure is proposed in section 6.2 of DH to compute bootstrapped critical values for $\bar{Z}$ and $\widetilde{Z}$ instead of asymptotic critical values. The procedure is composed of the following steps:[4]

1. Estimate (3) and obtain $\bar{Z}$ and $\widetilde{Z}$ as defined in (6) and (7).

2. Estimate 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 $\widehat{\boldsymbol{\varepsilon}}_{(T-K)\times N}$.

---

[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.

[4]The procedure we present here differs slightly from that proposed by DH, in the numbering of the steps, but more importantly also in the definition of the initial conditions (our step 4), which is not addressed in DH, and 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.

3. Build a matrix $\boldsymbol{\varepsilon}^{\star}_{(T-K)\times N}$ by resampling (overlapping blocks of) rows (i.e., time periods) of matrix $\widehat{\boldsymbol{\varepsilon}}$. Block bootstrap is useful in presence of autocorrelation.

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

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

6. Estimate the model: $y^{\star}_{i,t} = \alpha^b_i + \sum_{k=1}^{K} \gamma^b_{ik} y^{\star}_{i,t-k} + \sum_{k=1}^{K} \beta^b_{ik} x_{i,t-k} + \varepsilon_{i,t}$, and compute $\bar{Z}^b$ and $\widetilde{Z}^b$.

7. Run B replications of steps 3 to 6.

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

3 The xtgcause command

The syntax of `xtgcause` is as follows:

`xtgcause` *depvar indepvar* $\left[\,if\,\right]\left[\,in\,\right]\left[\,,\,\underline{\texttt{lags}}(\texttt{\# | aic [\#] | bic [\#] | hqic [\#]})\right.$
$\left.\underline{\texttt{regress}}\,\underline{\texttt{boot}}\texttt{strap}\,\underline{\texttt{breps}}(\texttt{\#})\,\underline{\texttt{blevel}}(\texttt{\#})\,\underline{\texttt{blength}}(\texttt{\#})\,\texttt{seed}(\texttt{\#})\,\texttt{nodots}\,\right]$

`lags` specifies the lag structure to use for the regressions performed in computing the test statistic. By default, 1 lag is included. Specifying `lags(#)` requests that `#` lags of the series be used in the regressions. The maximum authorized number of lags is such that $T > 5 + 3 \cdot \texttt{\#}$. Specifying `lags(aic|bic|hqic [#])` requests that the number of lags of the series be chosen such that the average Akaike/Bayesian/Hannan-Quinn information criterion (AIC/BIC/HQIC) for the set of regressions is minimized. Regressions with 1 to `#` lags will be conducted, restricting the number of observations to $T-\texttt{\#}$ 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 (re-estimated 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 to be computed using a bootstrap procedure as proposed in section 6.2 of Dumitrescu and Hurlin (2012). Bootstrap is useful in presence of cross-sectional dependence.

breps indicates the number of bootstrap replications to perform. By default, it is set to 1000.

blevel indicates the significance level (in %) for computing the bootstrapped critical values. By default, it is set to 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 to implement the bootstrap by dividing the sample into blocks of `#` time periods and sampling the blocks independently with 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 only be used if **bootstrap** is also specified.

3.1 Saved results

`xtgcause` saves the following results in `r()`:

Scalars

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

Bootstrap scalars

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

Matrices

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

Bootstrap matrices

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

4 Examples

Before presenting a couple of 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 check his 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 et al., 2003; Levin et al., 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 the functioning of `xtgcause`, we first use simulated data provided by DH at http://www.execandshare.org in the file data-demo.csv.[5] We start by importing the original Excel dataset directly from the website. In the original CSV file, the dataset is organized as a matrix, with all observations for the 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 so as to be understood as a panel by Stata.

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

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

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

. drop x y

. gen 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

. l 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   │
      └───────────────────────────────────┘
```

---

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

```
. l 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 piece of code are quite involved, and a few explanations are in order. 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` in order to avoid a long output indicating that 2 sets of 10 variables ($x_1, ..., x_{10}$, and $y_1, ..., y_{10}$) were created. These variables were immediately converted from string to numeric thanks to `split`'s `destring` option. In order to have a well-shaped panel that Stata can correctly interpret, we combined these 2 sets of 10 variables into only 2 variables, which we did using `reshape`. A few observations (the first five for individuals 1 and 2) are displayed to show how the data is finally organized.

Using the formatted and `xtset`ted data, we can now run `xtgcause`. The simplest possible test in order 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).
```

Since 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), $\bar{Z}$ (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):

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

. mat li 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 2 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 as before.

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

```
. xtgcause y x, bootstrap l(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

In order to provide an example based on real data, we searched for papers reporting Dumitrescu and Hurlin's tests and published in journals that make authors' datasets available. We found several such papers (e.g., Paramati et al., 2016, 2017; Salahuddin et al., 2016). In particular, Paramati et al.

(2016) (hereafter PUA) investigate the effect of foreign direct investment and stock market growth on clean energy use.[6] In their Table 8, they report a series of pairwise panel causality tests between variables such as economic output, CO_2 emissions, or clean energy consumption. 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, l(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, l(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 first line of the above code imports the dataset constructed by PUA (file Data-WDI.xlsx, sheet "FirstDif-Data"). 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 PUA's Table 8. We use 2 lags in both cases to match the numbers indicated by PUA in their accompanying appendix file. Comparing with PUA's 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).

Optionally, `xtgcause` allows the user to request the lag order to be chosen so that the Akaike, Bayesian, or Hannan-Quinn information criteria be minimized. Given that DH offer no guidance regarding the choice of the lag order, this feature might be appealing to practitioners. We can for instance test the causality from *output* to *fdi* specifying the option `lags(bic)`:

---

[6]See http://www.sciencedirect.com/science/article/pii/S0140988316300214.

```
. xtgcause fdi output, l(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 (i.e., such that $T > 5 + 3K$ or optionally specified by the user below this limit), maintaining a common sample. Said otherwise, if at most 5 lags are to be considered, the first 5 observations of the panel will never be considered in the estimations, even if it would be possible to do so with fewer than 5 lags. This is required in order to have nested models, which can then be appropriately compared using AIC, BIC, or HQIC. After this series of estimations, `xtgcause` selects the optimal outcome (i.e., such that the average AIC/BIC/HQIC of the N individual estimations is the lowest) and re-runs all estimations with the optimal number of lags and using the maximal number of 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 PUA for this test.[7] Worryingly, this difference is not without consequences, since the conclusion of the test in this case is reversed. More precisely, the null hypothesis is not rejected with the optimally-selected single lag, but PUA use 2 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 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 in presence of cross-sectional dependence. Extending our example based on PUA 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):

---

[7]The number of lags would be 3 using HQIC and 4 using AIC. Therefore, while PUA 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.

```
. xtgcause fdi output, l(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.
```

What `xtgcause` does in this case is first to compute the Z-bar and Z-bar tilde statistics using the optimal number of lags as in previous series of estimations, and it then computes the bootstrapped 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 to 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 that asymptotic p-values are under-estimated. Bootstrapped p-values indicate that the null hypothesis is far from being rejected, strengthening our above concerns about PUA conclusions based on the asymptotic p-values and obtained with 2 lags.

5 Conclusion

This paper has presented the user-written command `xtgcause`, which automates a procedure introduced by Dumitrescu and Hurlin (2012) in order to detect Granger causality in panel datasets. In this branch of econometrics, the empirical literature appears to be lagging, with the latest theoretical developments being not always available in statistical packages. One important contribution of our command is to allow the user to select the number of lags based on the Akaike, the Bayesian, or the Hannan-Quinn information criterion. This choice may have an impact on the conclusion of the test, but some researchers may have overlooked it. As a consequence, several empirical papers might have reached erroneous conclusions. Another useful contribution of `xtgcause` is that it allows to calculate bootstrapped critical values, a very useful option in presence of cross-sectional dependence. With this command and this article, we therefore hope to bring some useful clarifications and help practitioners conduct sound research.

References

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

Breitung, J. (2000). The local power of some unit root tests for panel data. *Advances in Econometrics*, 15:161–177.

Dumitrescu, E.-I. and Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. *Economic Modelling*, 29(4):1450–1460.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.

Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *Econometrics Journal*, 3(2):148–161.

Harris, R. D. and Tzavalis, E. (1999). Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics*, 91(2):201–226.

Im, K. S., Pesaran, M., and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115(1):53–74.

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

Paramati, S. R., Apergis, N., and Ummalla, M. (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., Ummalla, M., and Apergis, N. (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(2):265–312.

Salahuddin, M., Alam, K., and Ozturk, I. (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(400):1072–1078.