Abstract

This article presents the user-written command `xtgcause`, which implements a procedure proposed by Dumitrescu and Hurlin (2012) for detecting Granger causality in panel datasets. 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 contribution constitutes an effort to help practitioners understand and apply Dumitrescu and Hurlin’s 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.
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:

$$ y_t = \alpha + \sum_{k=1}^{K} \beta_k y_{t-k} + \sum_{k=1}^{K} \gamma_k x_{t-k} + \varepsilon_t $$

(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:

$$ H_0 : \gamma_1 = \ldots = \gamma_K = 0 $$

(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:

$$y_{i,t} = \alpha_i + \sum_{k=1}^{K} \beta_{ik} y_{i,t-k} + \sum_{k=1}^{K} \gamma_{ik} x_{i,t-k} + \varepsilon_{i,t} \quad (3)$$

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:

$$H_0 : \gamma_{i1} = \ldots = \gamma_{iK} = 0 \quad \forall i = 1, \ldots, N \quad (4)$$

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:

$$H_1 : \gamma_{i1} = \ldots = \gamma_{iK} = 0 \quad \forall i = 1, \ldots, N_1$$

$$\gamma_{i1} \neq 0 \text{ or } \ldots \text{ or } \gamma_{iK} \neq 0 \quad \forall i = N_1 + 1, \ldots, N$$

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 $\gamma_{i1} = \ldots = \gamma_{iK} = 0$ to retrieve the individual Wald statistic $W_i$, and finally compute the average Wald statistic $\overline{W}$:

$$\overline{W} = \frac{1}{N} \sum_{i=1}^{N} W_i \quad (5)$$

---

1See 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.
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 $\bar{W}$ 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 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:

$$\bar{Z} = \sqrt{\frac{N}{2K}} \cdot (\bar{W} - K) \xrightarrow{d, \text{ as } T,N \to \infty} \mathcal{N}(0,1) \tag{6}$$

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

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

The testing procedure of the null hypothesis in (4) is finally based on $\bar{Z}$ and $\tilde{Z}$. 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, $\tilde{Z}$ 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 $K_{\text{max}}$ 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 $\tilde{Z}$ instead of asymptotic critical values. The procedure (adapted from DH) is composed of the following steps:

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

\(^2\)We thank Gareth Thomas (IHS Markit EViews) for bringing this point to our attention.

\(^3\) $K_{\text{max}}$ stands for the maximal number of lags to be considered.
2. Estimate the model under $H_0$: $y_{i,t} = \alpha_i + \sum_{k=1}^{K} \beta_{ik} y_{i,t-k} + \varepsilon_{i,t}$, and collect the residuals in matrix $\hat{\varepsilon}_{(T-K) \times N}$.

3. Build a matrix $\hat{\varepsilon}_b^{(T-K) \times N}$ by resampling (blocks of) rows (i.e., time periods) of matrix $\hat{\varepsilon}$ with replacement. Block bootstrap is useful in presence of autocorrelation.

4. Construct a resampled series $\tilde{y}_{i,t}^b = \hat{\alpha}_i + \sum_{k=1}^{K} \hat{\beta}_{ik} y_{i,t-k}^b + \tilde{\varepsilon}_{i,t}$.

5. Estimate the model: $\tilde{y}_{i,t}^b = \alpha_i + \sum_{k=1}^{K} \beta_{ik} \tilde{y}_{i,t-k}^b + \sum_{k=1}^{K} \gamma_{ik} x_{i,t-k} + \tilde{\varepsilon}_{i,t}$, and compute $\tilde{Z}^b$ and $\tilde{Z}^b$.

6. Run $B$ replications of steps 3 to 5.

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

3 The xtgcause command

The syntax of xtgcause is as follows:

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

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 $T > 5 + 3 \cdot #$. 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 $T-#$ 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.
\texttt{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.

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

\texttt{blevel} indicates the significance level (in %) for computing the bootstrapped critical values. By default, it is set to 95%.

\texttt{blength} indicates the size of the block length to be used in the bootstrap. By default, each time period is sampled independently with replacement (\texttt{blength(1)}). \texttt{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.

\texttt{seed} can be used to set the random-number seed. By default, the seed is not set.

\texttt{nodots} suppresses replication dots. By default, a dot is printed for each replication to provide an indication of the evolution of the bootstrap. \texttt{breps}, \texttt{blevel}, \texttt{blength}, \texttt{seed} and \texttt{nodots} are \texttt{bootstrap} suboptions. They can only be used if \texttt{bootstrap} is also specified.

### 3.1 Saved results

\texttt{xtgcause} saves the following results in \texttt{r()}:

**Scalars**

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

**Bootstrap scalars**

- \texttt{r(zbar_cv)}: critical value for the Z-bar statistic
- \texttt{r(zbart_cv)}: critical value for the Z-bar tilde statistic
- \texttt{r(breps)}: number of bootstrap replications
- \texttt{r(blevel)}: significance level for bootstrap critical values
- \texttt{r(blength)}: size of the block length
- \texttt{r(Wi)}: individual Wald statistics
- \texttt{r(PVi)}: p-values of the individual Wald statistics

### 4 Examples

Before presenting a couple of examples, we recall that the test implemented in \texttt{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; 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. 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 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.

```stata
. import delimited using "http://www.execandshare.org/execandshare/htdocs/data/MetaSite/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
<table>
<thead>
<tr>
<th></th>
<th>wide</th>
<th>long</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of obs.</td>
<td>20</td>
<td>200</td>
</tr>
<tr>
<td>Number of variables</td>
<td>21</td>
<td>4</td>
</tr>
<tr>
<td>j variable (10 values)</td>
<td>-&gt; id</td>
<td></td>
</tr>
<tr>
<td>xij variables:</td>
<td>x1 x2 ... x10 -&gt; x</td>
<td>y1 y2 ... y10 -&gt; y</td>
</tr>
</tbody>
</table>
. xtset id t
  panel variable: id (strongly balanced)
  time variable: t, 1 to 20
  delta: 1 unit
```

*Data and MATLAB code are also available at http://www.runmycode.org/companion/view/42 in a zip file.*
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 (\( x_1, \ldots, x_{10} \), and \( y_1, \ldots, y_{10} \)) were created. These variables were immediately converted from string to numeric thanks to `split`’s `destring` option. 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)
-----------------------------------------------------------------

\( \text{H}_0: x \text{ does not Granger-cause } y. \)
\( \text{H}_1: x \text{ does Granger-cause } y \text{ for at least one panelvar (id)}. \)

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 $W$ (W-bar), $Z$ (Z-bar), and $\bar{Z}$ (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 $r(W_i)$ 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  .5665945  .46256089
id2  .11648998 .73731411
id3  .09081952 .76701924
id4  .1263612  .01156476
id5  .18687517 .67129995
id6  .0060395  .38417583
id7  .53073859 .47681675
id8  .00158371 .96748225
id9  .4365413  .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, lag(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. The test can also be conducted using a bootstrap procedure to compute p-values and critical values:

```
. xtgcause y x, bootstrap l(1) breps(100) seed(20170905)
```

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* = 0.5600, 95% critical value = 2.8926)
Z-bar tilde = 0.2590 (p-value* = 0.7500, 95% critical value = 2.0138)
--------------------------------------------------------------
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.\footnote{See http://www.sciencedirect.com/science/article/pii/S0140988316300214.} In their Table 8, they report a series of pairwise panel causality tests between variables such as economic output, CO\textsubscript{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:

```plaintext
.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 \texttt{xtgcause} to test for the causality from \texttt{output} to \texttt{co2} and from \texttt{output} to \texttt{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, \texttt{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 \textit{output} to \textit{fdi} specifying the option \texttt{lags(bic)}:

\begin{verbatim}
.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, \texttt{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 less 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, \texttt{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.\footnote{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.} 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 \texttt{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

\begin{verbatim}
. 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, \texttt{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 less 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, \texttt{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.\footnote{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.} 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 \texttt{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 to make their choice in a transparent way.

Finally, \texttt{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 \textit{output} to \textit{fdi} by adding the \texttt{bootstrap} option (we also use \texttt{seed} for replicability reasons and \texttt{nodots} for the sake of space):

\begin{verbatim}
. xtgcause fdi output, l(bic) bootstrap seed(20170905) nodots

Bootstrap replications (1000)

Dumitrescu & Hurlin (2012) Granger non-causality test results:

Optimal number of lags (BIC): 1 (lags tested: 1 to 5).

\begin{tabular}{l}
W-bar = 1.3027 \\
Z-bar = 0.9572 (p-value* = 0.4840, 95\% critical value = 3.6304) \\
Z-bar tilde = 0.4260 (p-value* = 0.6990, 95\% critical value = 2.5521) \\
\end{tabular}

\end{verbatim}

H0: output does not Granger-cause \textit{fdi}.
H1: output does Granger-cause \textit{fdi} for at least one panelvar (id).

*p-values computed using 1000 bootstrap replications.

What \texttt{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 the previous estimation, and it then computes the bootstrapped p-values and critical values. By default, 1,000 bootstrap replications are performed. Comparing with the asymptotic p-values displayed in the former estimation, we observe that the p-value for the Z-bar increases substantially (from 0.34 to 0.48), while that for the Z-bar tilde remains very close. 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 \texttt{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


