University of Neuchatel

Institute of Economic Research

IRENE Working paper 17-03

Testing for Granger causality in panel data

*Luciano Lopez** *Sylvain Weber**

** University of Neuchatel (Institute of Economic Research)*

UNIVERSITÉ DE NEUCHÂTEL

Institut de recherches économiques

Testing for Granger causality in panel data

Luciano Lopez[*] Sylvain Weber[†]

July 18, 2017

Abstract

This article presents the Stata user-written command `xtgcause`, which implements a procedure proposed by Dumitrescu and Hurlin (2012) for testing Granger causality in panel datasets. With the development of large and long panel databases, theories surrounding panel causality evolve at a fast pace and empirical researchers may sometimes find it difficult to run the most recent tests developed in the literature. This contribution constitutes an effort to help practitioners understand and apply the test. In the same vein, the command offers the possibility to select the number of lags to include in the model by minimizing the Akaike, the Bayesian, or the Hannan-Quinn information criterion.

JEL Classification: C23, C87.

Keywords: Stata, Granger causality, panel datasets.

[*]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 raise in panels. 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 the `xtgcause` command using examples based on simulated and real data. The objective of our contribution is to support the empirical literature using panel causality techniques. In this line of reasoning, 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.

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:

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

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 test this causality based on an F-test with the following null hypothesis:

$$H_0: \gamma_1 = ... = \gamma_K = 0 \tag{2}$$

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 extended test designed to detect causality in panel data. The underlying regression writes as follows:

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

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 the 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:

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

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

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

$$\begin{aligned} H_1 : \ & \gamma_{i1} = ... = \gamma_{iK} = 0 && \forall\, i = 1, ..., N_1 \\ & \gamma_{i1} \neq 0 \text{ or } ... \text{ or } \gamma_{iK} \neq 0 && \forall\, i = N_1 + 1, ..., N \end{aligned}$$

where $N_1 \in [0, N-1]$ is unknown. If $N_1 = 0$, there is causality for all individuals in the panel. N_1 is 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 $\gamma_{i1} = ... = \gamma_{iK} = 0$ to retrieve W_i, and finally compute $\overline{W}$ as the average of the N individual Wald statistics:

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

where W_i is the standard adjusted Wald statistic for individual i observed during T periods.[1] We emphasize that the test is designed to detect

[1]See Dumitrescu and Hurlin (2012, p. 1453) for the mathematical definition of W_i used in `xtgcause`. Note however that T in DH's formula 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.

causality at the panel-level, and rejecting H_0 does not exclude that there is no 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.

Under the assumption that 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 corresponding normal critical values, then one should reject H_0 and conclude that there is Granger causality. 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.

3 The xtgcause command

The syntax of `xtgcause` is as follows:

`xtgcause` *varlist* [*if*] [*in*] [`, lags(# | aic [#] | bic [#] | hqic [#])` `regress`]

`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.[2] 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 the maximum authorized number of lags is used.

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

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(pvzbar)`	P-value of the Z-bar statistic
`r(zbart)`	Z-bar tilde statistic	`r(pvzbart)`	P-value of the Z-bar tilde statistic
Matrices			
`r(Wi)`	Individual Wald statistics	`r(PVi)`	P-values of the individual Wald statistics

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 a series of stationarity tests (Breitung, 2000; Harris and Tzavalis, 1999; Im et al., 2003; Levin and Lin, 1992; Levin et al., 2002; Pesaran, 2007).

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.[3] We start by importing the original Excel dataset directly from the abovementioned website. In the original csv file, the dataset is organized as a matrix, with all observations for the 1st individual in a single cell. Within this cell, the (10) values of variable x are separated by tabs, a comma

[2] We thank Gareth Thomas for bringing this point to our attention.

[3] The data is also available at http://www.runmycode.org/companion/view/42 in a zip file.

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    |
     +-------------------------------------+

. 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 `stset`ted data, we can now run `xtgcause`. The simplest possible test in order to determine 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.[4]

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
```

[4] In a first version of this paper, some results differed from what is presented here, but coincided with those provided by DH at http://www.execandshare.org (let alone a mismatch in the Z-bar statistic due to a typo in DH code). After an exchange between Elena-Ivona Dumitrescu, Christophe Hurlin, Gareth Thomas (IHS Markit, Eviews), and us, it turns out the differences arose from the issue detailed in footnote 1. We have modified our `xtgcause` code, and DH have modified theirs (including the typo plaguing the Z-bar). The issue is thus cleared and the outputs obtained with `www.execandshare.org`, Eviews, and `xtgcause` coincide.

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.

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 and codes available. We found several such papers (e.g., Paramati et al., 2016, 2017; Salahuddin et al., 2016), and all of these use EViews to run the tests.

In particular, Paramati et al. (2016) (hereafter PUA) investigate the effect of foreign direct investment and stock market growth on clean energy use.[5] 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. Moreover, based on the detailed elements provided in sheet "Panel-Causality" of Results.xlsx, we can replicate their results:

```
. import excel using "./Recent empirical papers using DH test/Data-WDI.xlsx", cl
> ear first case(lower) cellrange(A1:I441)  sheet(Final-Raw-Data)
. d id year output co2 cec
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------------
id              byte    %10.0g                ID
year            int     %10.0g                Year
output          double  %10.0g                Output
co2             double  %10.0g                CO2
cec             double  %10.0g                CEC
. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 1991 to 2012
                delta:  1 unit
```

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

```
. foreach var of varlist cec-stock {
  2. qui: gen delta_l`var´ = ln(`var´) - ln(l.`var´)
  3. }
. xtgcause delta_lcec delta_loutput, l(3)
Dumitrescu & Hurlin (2012) Granger non-causality test results:
--------------------------------------------------------------
Lag order: 3
W-bar =          5.0300
Z-bar =          3.7062   (p-value = 0.0002)
Z-bar tilde =    1.5554   (p-value = 0.1199)
--------------------------------------------------------------
H0: delta_loutput does not Granger-cause delta_lcec.
H1: delta_loutput does Granger-cause delta_lcec for at least one panelvar (id).
. xtgcause delta_lco2 delta_loutput, 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: delta_loutput does not Granger-cause delta_lco2.
H1: delta_loutput does Granger-cause delta_lco2 for at least one panelvar (id).
```

The first lines of the above code import the raw data provided by PUA (file Data-WDI.xlsx, sheet “Final-Raw-Data”) and construct the variables (first differences of the variables in logarithms) on which the tests are conducted. The constructed variables correspond to what is provided in sheet “FirstDif-Data” of the excel file.

We then use `xtgcause` to test for the causality from *output* to *cec* and from *output* to *co2*, which correspond to the tests reported in the first two lines of Table 8 in PUA. We use 3 and 2 lags, respectively, to match the numbers indicated by PUA in their accompanying appendix file. In our output, the Z-bar statistics are 3.7062 and 0.9442 and the Z-bar tilde statistics are 1.5554 and 0.1441, respectively. Therefore, it turns out that the Z-bar tilde coincide with the “Zbar-Stat” reported by PUA. We note in passing that the denomination “Zbar-Stat.” used in EViews corresponds to Z-bar tilde while Z-bar is not displayed.

Finally, note that `xtgcause` allows to request that the lag order 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. Extending our example above, we can for instance test the causality from *output* to *cec* specifying the option `lags(aic)`:

```
. xtgcause delta_lcec delta_loutput, l(aic)
Dumitrescu & Hurlin (2012) Granger non-causality test results:
--------------------------------------------------------------
Optimal number of lags (AIC): 2 (lags tested: 1 to 5).
W-bar =          2.5393
Z-bar =          1.2058   (p-value = 0.2279)
Z-bar tilde =    0.3336   (p-value = 0.7387)
--------------------------------------------------------------
H0: delta_loutput does not Granger-cause delta_lcec.
H1: delta_loutput does Granger-cause delta_lcec for at least one panelvar (id).
```

In practice, what `xtgcause` does in this situation is run all sets of regressions containing 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), keeping the number of observations in all estimations constant. Said otherwise, if at most 5 lags are to be considered, the first 4 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. We do so 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 AIC appears to be 2, which is different from the lag order PUA used for this test. The number of lags selected would also be 2 using HQIC, but it would be 1 using BIC. Worryingly, this difference is not without consequences, since the conclusion of the test is in this case reversed. More precisely, the null hypothesis is not rejected with the optimally-selected 1 or 2 lags, but PUA use 3 lags and therefore reject the null hypothesis. Considering that empirical research in economics is used to formulate policy recommendations, 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 to make their choice in a transparent way.

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 conclusions. With this command and this article, we hope to bring some useful clarifications and help practitioners.

References

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.

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. and Lin, C.-F. (1992). Unit root tests in panel data: asymptotic and finite-sample properties. *UC San Diego, Working Paper 92-23.*

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.