The Stata Journal (2022)
22, Number 2, pp. 355–378 DOI: 10.1177/1536867X221106403

Testing for time-varying Granger causality

Christopher F. Baum
Department of Economics
Boston College
Chestnut Hill, MA
baum@bc.edu
https://orcid.org/0000-0003-4766-3699

Stan Hurn
School of Economics and Finance
Queensland University of Technology
Brisbane, Australia
s.hurn@qut.edu.au
https://orcid.org/0000-0002-6134-7943

Jesús Otero
Facultad de Economía
Universidad del Rosario
Bogotá, Colombia
jesus.otero@urosario.edu.co
https://orcid.org/0000-0002-2551-0053

Abstract. The concept of Granger causality is an important tool in applied macroeconomics. Recently, recursive econometric methods have been developed to analyze the temporal stability of Granger-causal relationships. This article offers an implementation of these recursive procedures in Stata. An empirical example illustrates their use in analyzing the temporal stability of Granger causality among key U.S. macroeconomic series.

Keywords: st0675, tvgc, Granger causality, time variation, temporal stability, datestamping

1 Introduction

Causal relationships in the econometric analysis of time series are typically based on the concept of predictability and are established by testing for Granger causality (Granger 1969, 1988). A variable X causes a variable Y in Granger's sense if accounting for past values of X enables better predictions to be made for Y, other things being equal. The popularity of Granger causality stems partly from the fact that it is not specific to a particular structural model but depends solely on the stochastic nature of variables. Testing for Granger causality typically involves testing joint-zero restrictions on blocks of parameters in reduced-form vector autoregressive (VAR) models. Given that VAR models have proved to be a particularly convenient way of modeling the dynamic interactions between economic variables, it is not surprising that there is a voluminous literature on applications of Granger causality in economics. Examples include the money–income relationship (Friedman and Kuttner 1993; Swanson 1998; Shi, Hurn, and Phillips 2020), the relationship between gross domestic product and energy consumption (Lee 2006; Arora and Shi 2016), CO_2 emissions as related to economic growth (Grossman and Krueger 1995), economic growth and health progress (Tapia Granados and Ionides 2008), oil prices and output (Hamilton 1983), and price dependence among crude oil varieties (Wlazlowski, Hagströömer, and Giulietti 2011) and precious metals (Chan and Mountain 1988).

Standard software for the estimation and analysis of VAR models provides Granger causality tests. However, the results of these tests are often sensitive to the time period over which the VAR is estimated. Just as with other aspects of structural stability, the existence of Granger causality between a pair of variables may be supported over one time frame but fragile when alternative periods are considered (see Thoma [1994], Swanson [1998], and Psaradakis, Ravn, and Sola [2005]). Drawing on theoretical results by Phillips, Wu, and Yu (2011) and Phillips, Shi, and Yu (2014, 2015a,b) in the context of testing and datestamping episodes of asset price bubbles, Shi, Phillips, and Hurn (2018) and Shi, Hurn, and Phillips (2020) revisit the notion of time variation in testing for Granger causality. In a series of articles, they establish that it is possible to assess the stability of causal relationships over time. The context of the 2018 article is the stationary VAR model, while the 2020 article extends the analysis to the lag-augmented VAR model to allow for the possibility of nonstationary variables in the VAR model (see Toda and Yamamoto [1995] and Dolado and Lütkepohl [1996]).

Although conceptually straightforward, these change-detection algorithms offer significant challenges in terms of computational statistics and data analysis. First, there are large numbers of test statistics produced by these methods that must be efficiently stored and displayed for analysis. Second, the tests also require bootstrapping to ensure correct inference. We illustrate how the analysis can be accomplished using a new community-contributed command, `tvgc`, developed for the Stata environment.

The rest of the article is organized as follows. Section 2 lays out the Granger-causal framework, while section 3 addresses recursive techniques for assessing time variation in causal relationships. Section 4 is dedicated to how inference is conducted using the bootstrap methodology. Section 5 presents details of the `tvgc` command. Section 6 provides an empirical example focusing on key U.S. macroeconomic series, illustrating results in both tabular and graphical forms. Section 7 offers brief concluding comments.

2 Granger causality

Consider, without loss of generality, the bivariate VAR(m) model given by

$$\mathbf{y}_{1t} = \phi_0^{(1)} + \sum_{k=1}^{m} \phi_{1k}^{(1)} \mathbf{y}_{1\,t-k} + \sum_{k=1}^{m} \phi_{2k}^{(1)} \mathbf{y}_{2\,t-k} + \boldsymbol{\varepsilon}_{1t}$$
$$\mathbf{y}_{2t} = \phi_0^{(2)} + \sum_{k=1}^{m} \phi_{1k}^{(2)} \mathbf{y}_{1\,t-k} + \sum_{k=1}^{m} \phi_{2k}^{(2)} \mathbf{y}_{2\,t-k} + \boldsymbol{\varepsilon}_{2t}$$

where $\mathbf{y}_{1t}$ and $\mathbf{y}_{2t}$, respectively, represent economic time series of interest and $\boldsymbol{\varepsilon}_{1t}$ and $\boldsymbol{\varepsilon}_{2t}$ are serially uncorrelated but possibly heteroskedastic disturbance terms. Variable $\mathbf{y}_1$ is said to Granger-cause variable $\mathbf{y}_2$ if the past values of $\mathbf{y}_1$ have predictive power for the current value of $\mathbf{y}_2$, conditional on the past returns of $\mathbf{y}_2$. Formally, the null hypothesis of no Granger causality from $\mathbf{y}_1$ to $\mathbf{y}_2$ involves testing the joint significance of $\phi_{1k}^{(2)}$ $(k = 1, \ldots, m)$ with a Wald test.

It is useful to recast the system in matrix notation. Let $\mathbf{y}_t = [\ \mathbf{y}_{1t} \quad \mathbf{y}_{2t}\ ]'$, $\mathbf{x}_t = [\ 1 \quad \mathbf{y}'_{t-1} \quad \mathbf{y}'_{t-2} \quad \ldots \quad \mathbf{y}'_{t-m}\ ]'$, and $\boldsymbol{\Pi}_{2\times(2m+1)} = [\ \boldsymbol{\Phi}_0 \quad \boldsymbol{\Phi}_1 \quad \ldots \quad \boldsymbol{\Phi}_m\ ]$ with $\boldsymbol{\Phi}_0 = [\ \phi_0^{(1)} \quad \phi_0^{(2)}\ ]'$ and

$$\boldsymbol{\Phi}_k = \begin{bmatrix} \phi_{1k}^{(1)} & \phi_{2k}^{(1)} \\ \phi_{1k}^{(2)} & \phi_{2k}^{(2)} \end{bmatrix} \quad \text{for} \quad k = 1, \ldots, m$$

The bivariate VAR(m) can then be written very simply as

$$\mathbf{y}_t = \boldsymbol{\Pi}\mathbf{x}_t + \boldsymbol{\varepsilon}_t$$

The null hypothesis of no Granger causality from variable $\mathbf{y}_1$ to $\mathbf{y}_2$ is $\mathbf{R}_{1\to 2}\boldsymbol{\pi} = \mathbf{0}$, where $\mathbf{R}_{1\to 2}$ is the coefficient restriction matrix that selects all coefficients on lagged $\mathbf{y}_1$ in the $\mathbf{y}_2$ equation and $\boldsymbol{\pi} = \text{vec}(\boldsymbol{\Pi})$ using row vectorization.

The heteroskedastic-consistent Wald statistic of the null hypothesis is denoted by $\mathcal{W}_{1\to 2}$ and is defined as

$$\mathcal{W}_{1\to 2} = T\left(\mathbf{R}_{1\to 2}\widehat{\boldsymbol{\pi}}\right)' \left\{\mathbf{R}_{1\to 2}\left(\widehat{\mathbf{V}}^{-1}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}^{-1}\right)\mathbf{R}'_{1\to 2}\right\}^{-1}\left(\mathbf{R}_{1\to 2}\widehat{\boldsymbol{\pi}}\right)$$

where $\widehat{\mathbf{V}} = \mathbf{I}_n \otimes \widehat{\mathbf{Q}}$, and $\widehat{\mathbf{Q}} = T^{-1}\sum_t \mathbf{x}_t\mathbf{x}'_t$, and $\widehat{\boldsymbol{\Sigma}} = T^{-1}\sum_t \widehat{\boldsymbol{\xi}}_t\widehat{\boldsymbol{\xi}}'_t$ with $\widehat{\boldsymbol{\xi}}_t = \widehat{\boldsymbol{\varepsilon}}_t \otimes \mathbf{x}_t$, and $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \widehat{\boldsymbol{\Pi}}\mathbf{x}_t$. Generalizing the formulation of a test for Granger causality beyond the bivariate VAR(m) model described here is straightforward.

The recursive algorithms for dealing with testing for time-varying Granger causality developed by Shi, Phillips, and Hurn (2018) and Shi, Hurn, and Phillips (2020) are now described.

3 Recursive testing algorithms

To allow for time variation in Granger causal orderings and to datestamp the timing of the changes, recursive estimation methods are required. A sequence of test statistics of Granger causality—one for each time period of interest—must be computed, and this information must then be used for inference. There are three algorithms that generate a sequence of test statistics: the forward expanding (FE) window, the rolling (RO) window, and the recursive evolving (RE) window algorithms. A schematic representation of the different algorithms is given in figure 1, in which each of the arrows is representative of a possible subsample over which the relevant test statistic is computed.

Consider a sample of $T+1$ observations $\{y_0, y_1, \ldots, y_T\}$ and a number r such that $0 < r < 1$. Also consider $[Tr]$ to denote the integer part of the product. Then $\mathcal{T}_{r_1,r}$ will be taken to denote a Wald test statistic computed over a subsample starting at $y_{[Tr_1]}$ and ending at $y_{[Tr]}$.

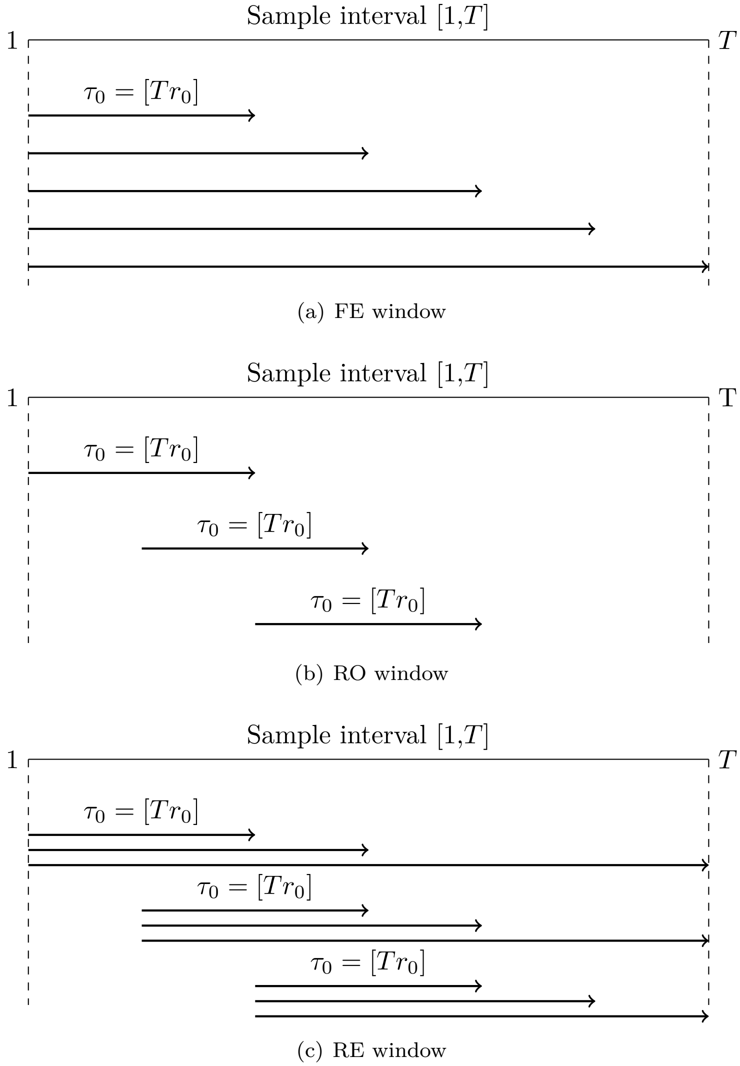


Figure 1. Sample sequences and window widths. Adapted from Phillips, Shi, and Yu (2015b).

3.1 The FE algorithm

The FE algorithm (Thoma 1994) is a standard forward recursion and is illustrated in panel (a) of figure 1. The Wald test statistic is computed first for a minimum window length, $\tau_0 = [Tr_0] > 0$, and the sample size then expands sequentially by one observation until the final test statistic is computed using the entire sample. Note that the starting point of every subsample in this recursion is the first data point. At the conclusion of the FE algorithm, a sequence of Wald test statistics, $\mathcal{T}_{r_1,r}$ with $r_1 = 0$ and $r \in [r_0, 1]$, is obtained.

3.2 The RO algorithm

The RO algorithm (Swanson 1998; Arora and Shi 2016) is illustrated in panel (b) of figure 1. A window of size $[Tw]$ is rolled through the sample, advancing one observation at a time, and a Wald test statistic is computed for each window. The output from the RO algorithm is a sequence of test statistics $\mathcal{T}_{r_1,r}$ with $r_1 = r - w$ and $r \in [r_0, 1]$, where each test statistic is computed from a sample of the same size, $[Tw]$, with $0 < w < 1$.

3.3 The RE algorithm

The RE algorithm (Phillips, Shi, and Yu 2015a; Shi, Hurn, and Phillips 2020) is illustrated in panel (c) of figure 1. For a given observation of interest, this algorithm runs a test regression for every possible subsample of size r_0 or larger, with the observation of interest providing the common endpoint of all the subsamples. This procedure is repeated, taking the observation of interest to be every point in the sample, subject only to the minimum window size. The result is that every observation in the sample, apart from the first subsample that defines the minimum window size, will have a set of Wald test statistics associated with it. Phillips, Shi, and Yu (2015b) propose that inference be based on the sequence of supremum norms of these statistics. The RE algorithm, therefore, produces a sequence of test statistics, denoted $\overline{\mathcal{T}}_{r_1,r}$ with $r_1 \in [0, r - r_0]$ and $r \in [r_0, 1]$, where every statistic in the sequence represents the supremum norm of the set of Wald statistics associated with each observation.

It is clear that the RE algorithm encompasses both the FE and RO recursions as special cases. For each observation in turn, a sequence of test statistics is defined that can be arranged in an upper triangular square matrix with column and row dimensions equal to the largest number of usable observations. Each column of this matrix corresponds to a particular observation of interest, as outlined previously. The FE Wald statistic is the leading entry in each column. The RO Wald statistic is located on the main diagonal. Finally, the largest elements of each column of the matrix are the relevant RE statistics.

The information derived from these test statistics can be used over the full sample or analyzed through the period to focus on the timing of these time-varying phenomena via datestamping.

3.4 Full-sample analysis

If the null hypothesis of interest is whether a particular variable does not Granger-cause another variable at any time during the sample, with the alternative that there is evidence of Granger causality at some time in the sample, then a single test statistic is required. The maximal FE statistic is taken to be the largest element of the first row of the matrix. The maximal RO statistic is the largest element of the main diagonal of the matrix. The maximal RE statistic is the largest element of the entire upper triangular matrix.

3.5 Datestamping

Beyond these summary measures for the full sample, the sequence of FE, RO, and RE statistics can be graphed and compared with the bootstrap percentiles. These estimates can then be used to identify periods in which the potential Granger-causal relationships vary significantly. The estimated origination date of a change is determined as the first instance at which the test statistic exceeds its critical value. Subsequent changes are then identified in a similar fashion.

4 Inference

The empirical distribution of the test statistics under the null hypothesis is computed by bootstrapping and controlling for size using the methodology described in section 3 of Shi, Hurn, and Phillips (2020); see also section 4.1 of Shi, Phillips, and Hurn (2018). Following these authors, with a slight change of notation, we next describe the five steps of the bootstrap procedure in the context of a simple bivariate VAR(1) for the null hypothesis of no Granger causality from y_{2t} to y_{1t}:

1. Fit the bivariate VAR(1) model over the full-sample period under the null hypothesis of no Granger causality from y_{2t} to y_{1t}:

$$\begin{bmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11}^{(1)} & 0 \\ \phi_{11}^{(2)} & \phi_{21}^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{1,t-1} \\ \mathbf{y}_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

   The estimates of the coefficients are denoted by $\widehat{\phi}_{11}^{(1)}$, $\widehat{\phi}_{11}^{(2)}$, and $\widehat{\phi}_{21}^{(2)}$, and the residuals are denoted by $\mathbf{e}_{1t}$ and $\mathbf{e}_{2t}$.

2. Denote the sample size of the bootstrapped data series by $T_b = \tau_0 + \tau_b - 1$, where $\tau_0 = [T r_0]$ and $\tau_b = [T r_b]$. The bootstrap sample is generated by

$$\begin{bmatrix} \mathbf{y}_{1t}^b \\ \mathbf{y}_{2t}^b \end{bmatrix} = \begin{bmatrix} \widehat{\phi}_{11}^{(1)} & 0 \\ \widehat{\phi}_{11}^{(2)} & \widehat{\phi}_{21}^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{1,t-1}^b \\ \mathbf{y}_{2,t-1}^b \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{1t}^b \\ \mathbf{e}_{2t}^b \end{bmatrix}$$

   The residuals $\mathbf{e}_{1t}^b$ and $\mathbf{e}_{2t}^b$ are randomly drawn with replacement from the estimated residuals $\mathbf{e}_{1t}$ and $\mathbf{e}_{2t}$, respectively. The initial values are obtained using $y_{11}^b = y_{11}$ and $y_{21}^b = y_{21}$.

3. Use the bootstrap series $\mathbf{y}_{1t}^b$ and $\mathbf{y}_{2t}^b$ to compute the test statistic sequences for the forward ($\{\mathcal{T}_{1,t}^b\}_{t=\tau_0}^{\tau_0+\tau_b-1}$), RO ($\{\mathcal{T}_{t-\tau_0+1,t}^b\}_{t=\tau_0}^{\tau_0+\tau_b-1}$), and RE ($\left\{\overline{\mathcal{T}}_t^b(\tau_0)\right\}_{t=\tau_0}^{\tau_0+\tau_b-1}$) estimation windows, respectively. Recall that, for the first two algorithms, the sequence comprises Wald statistics, but in the case of the RE algorithm, the sequence is made up of the supremum norm of the set of Wald statistics for each observation of interest. Calculate the maximum values of each bootstrapped test statistic sequence as

$$\text{Forward} : \mathcal{M}_{1,t}^b = \max_{t\in[\tau_0,\tau_0+\tau_b-1]} \mathcal{T}_{1,t}^b$$

$$\text{RO} : \mathcal{M}_{t-\tau_0+1,t}^b = \max_{t\in[\tau_0,\tau_0+\tau_b-1]} \mathcal{T}_{t-\tau_0+1,t}^b$$

$$\text{RE} : \overline{\mathcal{M}}_t^b(\tau_0) = \max_{t\in[\tau_0,\tau_0+\tau_b-1]} \left\{\overline{\mathcal{T}}_t^b(\tau_0)\right\}$$

   where the notation $\overline{\mathcal{M}}_t^b(\tau_0)$ is used to denote the supremum norm of a sequence of supremum norm Wald tests.

4. Repeat steps 2 and 3 a total of $b = 1, \ldots, B$ times.

5. Estimate the 90%, 95%, and 99% critical values of the forward, RO, and RE Wald statistics as the corresponding 90th, 95th, and 99th percentiles of the resulting B bootstrapped statistics computed in step 4.

At this point, it is worth describing the Stata implementation of some of the steps in the bootstrap described above. In step 1, noting that a VAR model can be viewed as a seemingly unrelated regression model where the explanatory variables are the same in each regression (see, for example, Judge et al. [1988]), estimation of the VAR model under the null hypothesis of Granger causality is simplified with the use of the `sureg` command, along with the `constraints()` option suitably defining the required exclusion restrictions. Next the `predict` postestimation command is used to create variables that contain the linear predictions of the model (using the option `xb`) and the associated residuals (using the option `residuals`). The residuals are randomly drawn with replacement, and the resulting bootstrapped residuals are then added to the linear predictions of the model to produce the required bootstrapped version of the original variables ($\mathbf{y}_{1t}^b$ and $\mathbf{y}_{2t}^b$ in our example above).

As for step 2, in their study of the money–income relationship, Shi, Hurn, and Phillips (2020) set $\tau_b = 12$ and 60 monthly observations, which is equivalent to controlling the size of the tests over periods of one and five years, respectively. Hence, there is no need to generate the bootstrapped version of the variables over all T observations (unless one chooses to do so) but only over the first T_b observations, which saves a lot of computational effort when applying the different windows of estimation.

Finally, the same bootstrapped statistics that are generated for the recursions in step 5 are used with the datestamping part of the method and are also used to conduct inference in the full-sample case when there is a single test statistic.

5 The tvgc command

The `tvgc` command calculates the time-varying Granger causality test statistics proposed by Shi, Phillips, and Hurn (2018) and Shi, Hurn, and Phillips (2020).

5.1 Syntax

Before using the `tvgc` command, and similar to many other Stata time-series commands, it is necessary to `tsset` the data. The syntax of `tvgc` is as follows:

`tvgc` *varlist* [*if*] [*in*] [`,` `prefix(`*prefix*`)` `p(`*integer*`)` `d(`*integer*`)` `robust` `trend` `matrix` `window(`*integer*`)` `boot(`*integer*`)` `seed(`*integer*`)` `sizecontrol(`*integer*`)` `noprint` `graph` `eps` `pdf` `notitle` `restab`]

The `tvgc` command tests whether the first variable in the *varlist* is Granger-caused by the remaining variables.

Note that *varlist* may not contain gaps but may contain time-series operators. `tvgc` does not support the `by:` prefix.

The community-contributed `moremata` package (Jann 2005) is required; for the latest version, type `ssc install moremata`.

5.2 Options

`prefix(`*prefix*`)` provides a "stub" with which variables created in `tvgc` will be named. By default, three variables (*prefix*`forward_`*varname*, *prefix*`rolling_`*varname*, and *prefix*`recursive_`*varname*) will be created for the appropriate date range. These variables must not already exist in memory. These variables record the Wald statistics that result from fitting the VAR or lag-augmented VAR model using forward recursive, RO, and RE windows. The `prefix()` option must be specified to enable the `graph` option, which includes 90th and 95th percentile bootstrap critical values in the plots.

`p(`*integer*`)` sets the number of lags to be included in the VAR model. The default is `p(2)`. This can be determined using the Stata command `varsoc` (see [TS] **varsoc**).

`d(`*integer*`)` sets the number of lags to be included in the lag-augmented part of the VAR model. The default is `d(1)`. This option must be used when there are integrated variables in the *varlist*. Set `d(0)` if no augmented lags are needed.

`robust` specifies that heteroskedasticity-robust test statistics are to be computed.

`trend` specifies the modeling of intercepts and trends. By default, `tvgc` assumes *varlist* is a nonzero mean stochastic process, so a constant is included in the VAR model. If the `trend` option is specified, a constant and a linear trend are included in the VAR model.

`matrix` specifies that the $T \times T$ matrices of test statistics be returned. They are named `r(m_`*rhsvar*`)`, where *rhsvar* is one of the test variables.

`window(`*integer*`)` specifies the number of observations to be included in the RO window. By default, 20% of the sample is used.

`boot(`*integer*`)` computes right-tail Monte Carlo critical values for the 90th, 95th, and 99th percentiles based on the bootstrap advocated by Shi, Phillips, and Hurn (2018) and Shi, Hurn, and Phillips (2020), using the specified number of replications. The default is `boot(199)`; at least 20 must be specified. The bootstrap critical values can be replicated if the option `seed()` is used.

`seed(`*integer*`)` sets the initial seed for random-number generation.

`sizecontrol(`*integer*`)` specifies the number of observations to be included in the bootstrap computations to control the empirical size. The default is `sizecontrol(12)`.

`noprint` specifies that detailed results not be printed.

`graph` specifies that the time series of the three test statistics be graphed along with their 90% and 95% critical values. The graphs will be saved with names specified by the `prefix()` option as *prefix*`forward_`*varname*, *prefix*`rolling_`*varname*, and *prefix*`recursive_`*varname*.

`eps` specifies that graphs be saved as `.eps` files and be displayed in the Graph Window.

`pdf` specifies that graphs be saved as `.pdf` files and be displayed in the Graph Window.

`notitle` specifies that graph titles are to be suppressed.

`restab` specifies that a LaTeX table containing the test statistics and their 95th and 99th percentile values be written to `restab.tex`. The file will be replaced if it exists. When including this fragment in a LaTeX document, the LaTeX `booktabs` package is required.

5.3 Stored results

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

Scalars

`r(T)`	number of observations
`r(p)`	lag order of VAR
`r(d)`	lag augmentation order
`r(bootrepl)`	number of bootstrap replications
`r(window)`	window width
`r(sizecontrol)`	window width for bootstrapping

Macros

`r(cmd)`	`tvgc`
`r(depvar)`	target variable
`r(rhsvars)`	test variables
`r(tstart)`	first period in analysis
`r(tend)`	last period in analysis

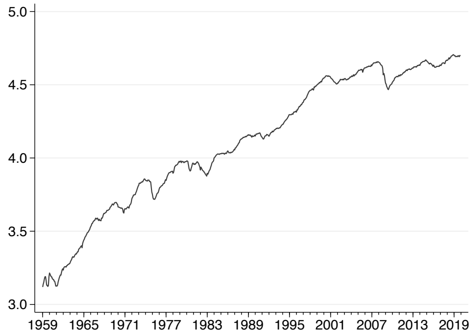
Matrices

`r(gcres)`	matrix of test statistics
`r(gccv90)`	matrix of critical values, 90th percentile
`r(gccv95)`	matrix of critical values, 95th percentile
`r(gccv99)`	matrix of critical values, 99th percentile
`r(m_`*rhsvar*`)`	complete matrix of test statistics for test variable *rhsvar*

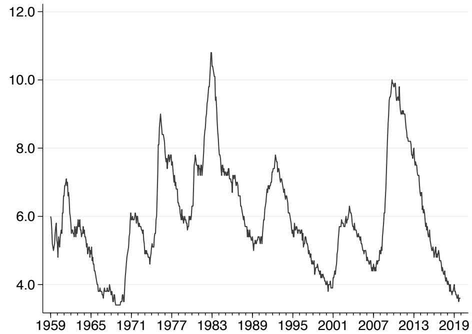
6 An empirical illustration

In this section, we illustrate the use of the command `tvgc`, using a three-variable VAR specification for monthly U.S. data. The VAR includes the logarithm of industrial production ($\ln i$), unemployment (u), and the logarithm of the price of crude oil ($\ln o$).[1] All three series are considered as the target variables for the Granger causality tests. The variables in the VAR constitute a subset of those used by Hamilton (1983) to analyze the relationship between oil and the macroeconomy using Granger causality tests. The sample period runs from January 1959 to December 2019, which yields 732 observations. The source of the data is FRED, the Federal Reserve Economic Data of the Federal Reserve Bank of St. Louis.[2] Figure 2 provides time-series plots of the variables in levels. It is apparent that the industrial production and oil price variables are trending.

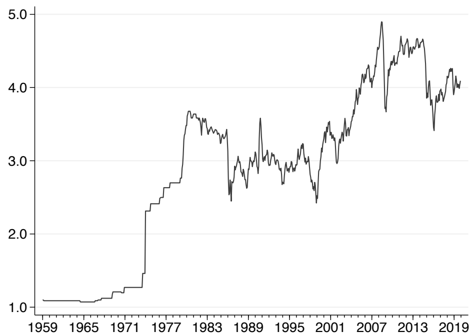
1. The command `tvgc` was applied to the dataset used by Shi, Hurn, and Phillips (2020) in their analysis of the money–income relationship in the United States. Our results match the MATLAB statistics obtained by these authors to three or four decimals. In terms of speed, our command compares favorably with their MATLAB routine.
2. The task of downloading the individual series from FRED is simplified with the command `freduse`; see Drukker (2006). The dataset used in this illustration can be downloaded from the Boston College Economics datasets by using the command `bcuse us_outoil` (Baum 2012).



(a) $\ln i$



(b) u



(c) $\ln o$

Figure 2. Variables in levels

We commence our empirical application by examining the time-series properties of the variables in the VAR. To this end, we apply the unit-root tests of Leybourne (1995) and Elliott, Rothenberg, and Stock (1996). The former test achieves power gains over the standard Dickey and Fuller (1979) testing procedure by applying the augmented Dickey–Fuller (ADF) regression to the forward and reverse realizations of the time series of interest and testing for the presence of a unit root based on the maximum ADF t statistic that results from the two regressions. Hence, the test is commonly referred to as ADF_{max}. In turn, Elliott, Rothenberg, and Stock (1996) are able to increase power over the standard Dickey–Fuller approach through generalized least-squares (DFGLS) removal of the underlying mean (or trend) in the variable of interest. In this case, the test is often referred to as the DFGLS test. Both tests are implemented with the commands `adfmaxur` and `ersur` developed by Otero and Baum (2018, 2017), respectively, as detailed in the note to table 1.

Table 1. Time-series properties of the data

Variable		ADF_{max}						DFGLS					
		Levels			First difference			Levels			First difference		
		p	Statistic		p	Statistic		p	Statistic		p	Statistic	
$\ln i$	A	12	−1.464	[0.714]	11	−7.415	[0.000]	12	−0.823	[0.895]	8	−3.271	[0.016]
	S	3	−1.407	[0.737]	2	−10.343	[0.000]	3	−0.829	[0.891]	3	−4.972	[0.000]
	GS5	12	−1.464	[0.714]	11	−7.415	[0.000]	12	−0.823	[0.898]	8	−3.271	[0.016]
u	A	12	−2.917	[0.016]	11	−6.577	[0.000]	12	−2.920	[0.004]	11	−4.530	[0.000]
	S	4	−2.950	[0.013]	3	−8.113	[0.000]	4	−2.952	[0.003]	4	−5.250	[0.000]
	GS5	12	−2.917	[0.016]	11	−6.577	[0.000]	12	−2.920	[0.004]	11	−4.530	[0.000]
$\ln o$	A	1	−2.463	[0.221]	0	−20.960	[0.000]	1	−2.488	[0.119]	0	−20.911	[0.000]
	S	1	−2.463	[0.209]	0	−20.960	[0.000]	1	−2.488	[0.114]	0	−20.911	[0.000]
	GS5	6	−2.072	[0.396]	5	−11.818	[0.000]	6	−2.111	[0.255]	5	−11.659	[0.000]

NOTE: ADF_{max} and DFGLS are the unit-root tests of Leybourne (1995) and Elliott, Rothenberg, and Stock (1996), respectively. These tests are implemented with the commands `adfmaxur` and `ersur`; see Otero and Baum (2018, 2017). The test regression for u includes a constant as the deterministic component, while those for $\ln i$ and $\ln o$ include a constant and a trend. p is the number of lags of the dependent variable that are included in the test regression to account for residual serial correlation. A and S indicate that p was determined using the Akaike (1974) and Schwarz (1978) information criteria, respectively; GS5 indicates that p was determined using the general-to-specific algorithm advocated by Campbell and Perron (1991) and Hall (1994) using a 5% significance level. When determining the optimal number of lags, we set $p_{\text{max}} = 12$.

To assess the robustness of our findings, the number of lags in the test regression is determined using the information criteria put forward by Akaike (1974) and Schwarz (1978), as well as the general-to-specific algorithm advocated by Campbell and Perron (1991) and Hall (1994). According to our results, both the ADF_{max} and the DFGLS tests support the presence of a unit root in $\ln i$ and $\ln o$ when considered in levels. In the case of u, both tests support the view that the rate of unemployment is stationary. Because there are variables integrated of order 1 in the VAR model under consideration, our analysis proceeds in the context of a lag-augmented VAR model, where $d = 1$.

The next task is to ascertain the optimum lag order of the VAR model by using the Stata command `varsoc`. The command is applied to a model that includes a linear trend that enters the model as an exogenous variable. The maximum number of lags is set to 12 because the data are monthly. The results, which are not reported here but are available upon request, indicate that the Schwartz lag-order selection statistic recommends the use of $p = 2$ lags in the VAR model, while Akaike favors $p = 6$ lags. The more parsimonious choice of $p = 2$ is adopted here.

Letting $x \xrightarrow{\text{GC?}} y$ to denote that the direction of Granger causality being tested runs from x to y, the following relationships are tested:

- $u \xrightarrow{\text{GC?}} \ln i$ and $\ln o \xrightarrow{\text{GC?}} \ln i$;
- $\ln i \xrightarrow{\text{GC?}} u$ and $\ln o \xrightarrow{\text{GC?}} u$; and
- $\ln i \xrightarrow{\text{GC?}} \ln o$ and $u \xrightarrow{\text{GC?}} \ln o$.

The required commands are, respectively, as follows:

```
. tvgc li u lo, trend window(72) sizecontrol(12) p(2) d(1) seed(123) boot(499)
> robust
. tvgc u li lo, trend window(72) sizecontrol(12) p(2) d(1) seed(123) boot(499)
> robust
. tvgc lo li u, trend window(72) sizecontrol(12) p(2) d(1) seed(123) boot(499)
> robust
```

In all three cases, the chosen options indicate the presence of a linear trend as an exogenous variable, `trend`; two lags in the VAR, `p(2)`; one lag in the lag-augmented part of the VAR, `d(1)`; and an initial estimation window of 72 observations, `window(72)`. The bootstrapped critical values are tabulated using 499 replications, `boot(499)`, where the size of the tests is controlled over a one-year period, `sizecontrol(12)`. The seed of the random-number generator is `seed(123)`. The tests are robust to heteroskedasticity if the option `robust` is applied. To produce the output reported in this article, the options `prefix()`, `graph`, and `eps` are required. Note, however, that these options are not included in the command lines above for reasons of space. The first command produces the following output. The graphs produced by this command are included in the first column of figure 3.

```
. tvgc li u lo, trend window(72) sizecontrol(12) p(2) d(1) seed(123) boot(499)
> robust prefix(LI_) graph eps notitle
Time-varying LA-VAR Granger causality test including trend, 1959m1 - 2019m12
TVGC robust test statistics for H0: li is GC
             Max_Wald_forward    Max_Wald_rolling  Max_Wald_recursive

          u            20.524              31.073              38.806
         lo            12.037              28.322              31.689
90th percentile of test statistics [499 replications]
             Max_Wald_forward    Max_Wald_rolling  Max_Wald_recursive

          u             7.763               8.051               8.637
         lo             6.763               7.169               7.552
95th percentile of test statistics [499 replications]
             Max_Wald_forward    Max_Wald_rolling  Max_Wald_recursive

          u            10.283              10.355              10.775
         lo             8.709               8.970               9.324
99th percentile of test statistics [499 replications]
             Max_Wald_forward    Max_Wald_rolling  Max_Wald_recursive

          u            15.751              15.110              16.131
         lo            12.459              13.526              14.389
```

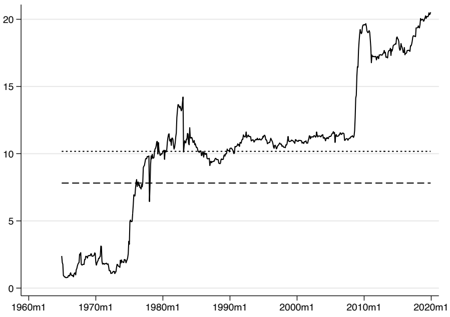
The results for the full sample, summarized in table 2, show that we fail to reject the null hypothesis of no Granger causality from income and unemployment to the price of oil when applying the FE window. In all other cases, the joint-zero restrictions on the relevant coefficients are rejected at the 5% level, as indicated by the computed statistic exceeding the 95th percentile of the empirical distribution of the bootstrap test statistics. This strong rejection of the null hypothesis is evidence of Granger causality between all the variables of the system.

Table 2. Wald tests of Granger causality

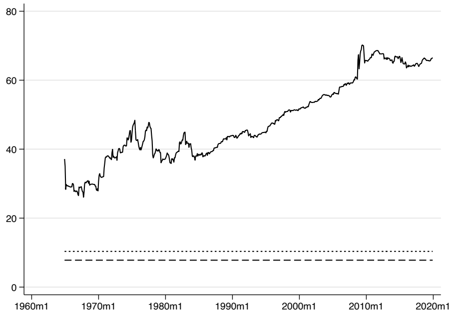
Direction of causality	Max Wald FE	Max Wald RO	Max Wald RE
$u \xrightarrow{\text{GC?}} \ln i$	20.524	31.073	38.806
	(10.283)	(10.355)	(10.775)
	[15.751]	[15.110]	[16.131]
$\ln o \xrightarrow{\text{GC?}} \ln i$	12.037	28.322	31.689
	(8.709)	(8.970)	(9.324)
	[12.459]	[13.526]	[14.389]
$\ln i \xrightarrow{\text{GC?}} u$	70.205	68.762	75.290
	(10.360)	(10.469)	(10.544)
	[15.850]	[17.025]	[17.892]
$\ln o \xrightarrow{\text{GC?}} u$	46.355	42.252	64.877
	(9.673)	(9.807)	(10.118)
	[13.607]	[13.607]	[13.962]
$\ln i \xrightarrow{\text{GC?}} \ln o$	4.349	25.639	30.328
	(7.913)	(8.565)	(9.344)
	[14.964]	[14.956]	[14.964]
$u \xrightarrow{\text{GC?}} \ln o$	3.440	17.229	17.253
	(9.333)	(9.333)	(10.121)
	[15.417]	[14.775]	[15.417]

NOTE: $x \xrightarrow{\text{GC?}} y$ indicates that the direction of Granger causality being tested runs from x to y. The underlying VAR model is fit with $p = 2$ lags and with $d = 1$ lag in the lag-augmented part, and it includes a trend. The study period is January 1959–December 2019. The minimum window size is set at 72 observations. The 95th and 99th percentiles of the empirical distribution of the bootstrap test statistics are shown in parentheses and brackets, respectively, and are based on 499 replications with a one-year period to control size. Wald tests are robust to heteroskedasticity.

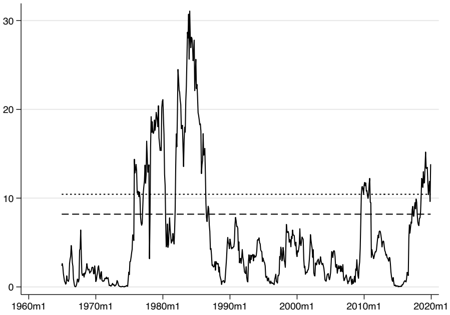
The time-varying Granger causality test results between income and unemployment, income and the oil price, and unemployment and the oil price are presented in figures 3, 4, and 5, respectively. In general, these plots all support the conclusion that Granger-causal relationships are extremely dynamic and that the patterns of causation found in the data depend on the type of recursive algorithm used.



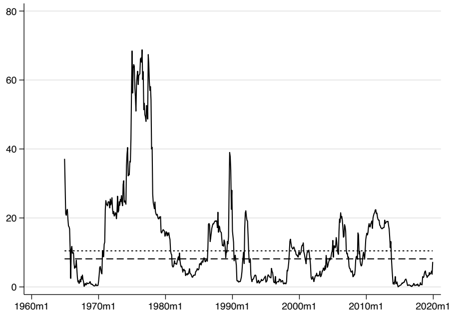
(a) Forward window: $u \xrightarrow{\text{GC?}} \ln i$



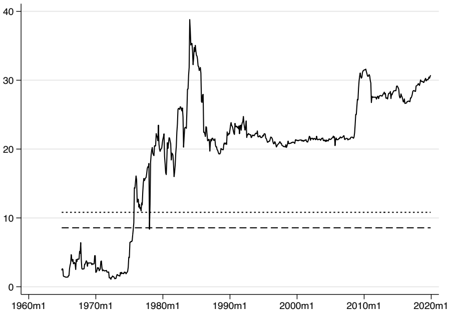
(b) Forward window: $\ln i \xrightarrow{\text{GC?}} u$



(c) Rolling window: $u \xrightarrow{\text{GC?}} \ln i$



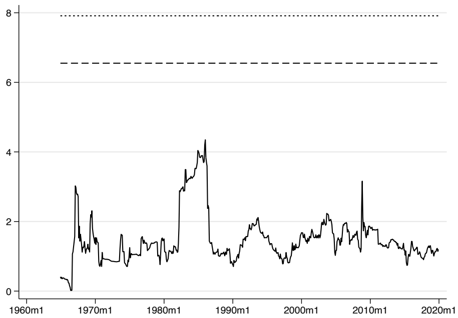
(d) RO window: $\ln i \xrightarrow{\text{GC?}} u$



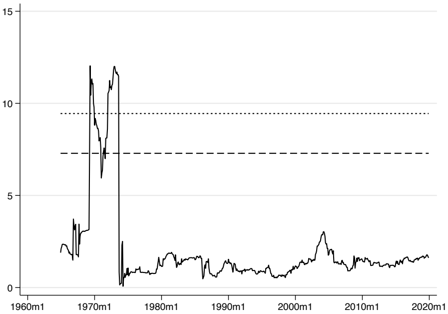
(e) Recursive window: $u \xrightarrow{\text{GC?}} \ln i$

(f) Recursive window: $\ln i \xrightarrow{\text{GC?}} u$

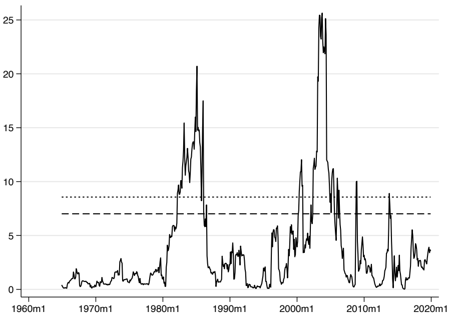
Figure 3. Time-varying Granger causality tests between $\ln i$ and u. $x \xrightarrow{\text{GC?}} y$ indicates that the direction of Granger causality being tested runs from x to y. The underlying VAR model is fit with $p = 2$ lags and with $d = 1$ lag in the lag-augmented part, and it includes a trend. The study period is January 1959–December 2019. The 10% and 5% bootstrapped critical values (lower and upper horizontal dashed lines, respectively) are based on 499 replications with a one-year period to control size. The minimum window size is set at 72 observations. Wald statistics are heteroskedasticity robust.



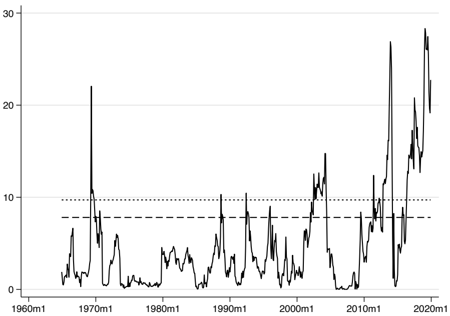
(a) Forward window: $\ln i \xrightarrow{\text{GC?}} \ln o$



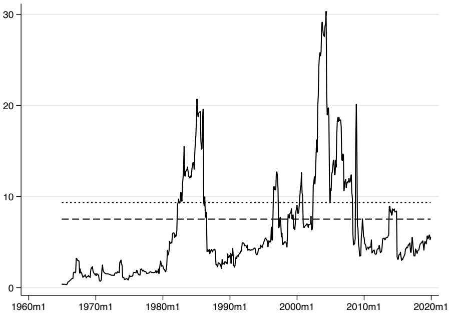
(b) Forward window: $\ln o \xrightarrow{\text{GC?}} \ln i$



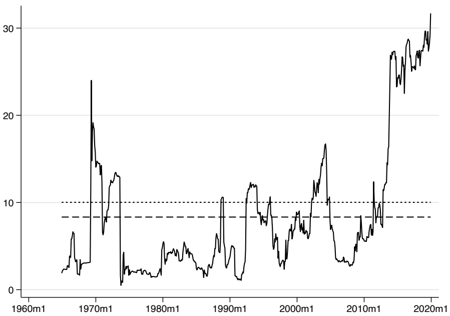
(c) Rolling window: $\ln i \xrightarrow{\text{GC?}} \ln o$



(d) RO window: $\ln o \xrightarrow{\text{GC?}} \ln i$



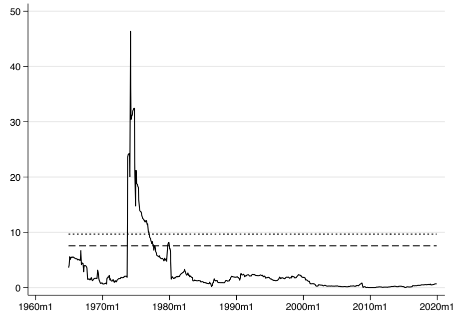
(e) Recursive window: $\ln i \xrightarrow{\text{GC?}} \ln o$



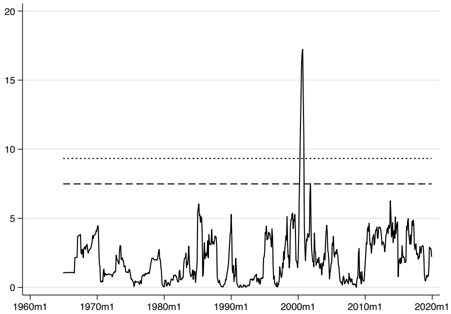
(f) Recursive window: $\ln o \xrightarrow{\text{GC?}} \ln i$

Figure 4. Time-varying Granger causality tests between $\ln i$ and $\ln o$. $x \xrightarrow{\text{GC?}} y$ indicates that the direction of Granger causality being tested runs from x to y. The underlying VAR model is fit with $p = 2$ lags and with $d = 1$ lag in the lag-augmented part, and it includes a trend. The study period is January 1959–December 2019. The 10% and 5% bootstrapped critical values (lower and upper horizontal dashed lines, respectively) are based on 499 replications with a one-year period to control size. The minimum window size is set at 72 observations. Wald statistics are heteroskedasticity robust.

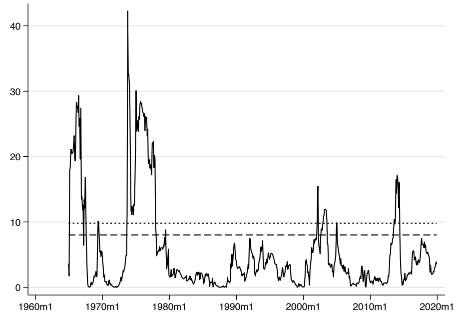
(a) Forward window: $u \xrightarrow{\text{GC?}} \ln o$



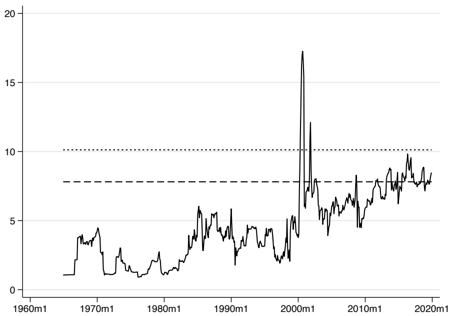
(b) Forward window: $\ln o \xrightarrow{\text{GC?}} u$



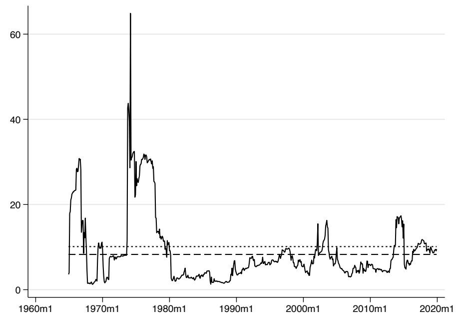
(c) Rolling window: $u \xrightarrow{\text{GC?}} \ln o$



(d) RO window: $\ln o \xrightarrow{\text{GC?}} u$



(e) Recursive window: $u \xrightarrow{\text{GC?}} \ln o$



(f) Recursive window: $\ln o \xrightarrow{\text{GC?}} u$

Figure 5. Time-varying Granger causality tests between u and $\ln o$. $x \xrightarrow{\text{GC?}} y$ indicates that the direction of Granger causality being tested runs from x to y. The underlying VAR model is fit with $p = 2$ lags and with $d = 1$ lag in the lag-augmented part, and it includes a trend. The study period is January 1959–December 2019. The 10% and 5% bootstrapped critical values (lower and upper horizontal dashed lines, respectively) are based on 499 replications with a one-year period to control size. The minimum window size is set at 72 observations. Wald statistics are heteroskedasticity robust.

There are several findings worth mentioning. For example, estimation using the FE and RE windows indicates that, during most of the study period, there is evidence of Granger causality from unemployment to income and vice versa. The presence of Granger causality in figures 3(a) and 3(e) and figures 3(b) and 3(f) is indicated whenever the value of the test statistic (solid line) exceeds the empirical critical values (dashed lines) obtained from the bootstrap distribution. Whenever this situation arises, the null hypothesis of no Granger causality is rejected. These results strongly support the intuition that these two measures of economic activity are closely related.

Additionally, FE estimation shows that the price of oil Granger-causes income in the late 1960s and early 1970s; see figure 4(b). By contrast, strong evidence of Granger causality from income to the price of oil is apparent in the 1980s and 2000s with the RO and RE windows; see figures 4(c) and 4(e), respectively. The fact that the FE window fails to pick up the opening of this causal channel late in the sample period confirms a well-known problem with the FE algorithm: namely, that it is not sensitive to changes late in the sample period.

A particularly strong illustration of the effects of the first oil shock of October 1973 is evident in figures 5(b), 5(d), and 5(f), showing Granger causality from the oil price to unemployment. All the algorithms identify a period of strong causality that starts at the time of the first oil shock and lasts until the second oil shock in 1979. Interestingly, although the causal channel from the oil price to unemployment is active at times in the latter half of the sample period, the channel is not open during the great recession of 2008–2009. As expected, there is little evidence of causality from unemployment to the oil price except for a short burst in the early 2000s. The reason for this anomalous result is unclear.

7 Conclusions

Evaluation of Granger-causal relationships among macroeconomic aggregates is an important component of macroeconometric modeling. It is crucial that the temporal stability of these relationships can be assessed formally. This article describes the implementation of a command, `tvgc`, to compute these test procedures, which can produce full-sample test statistics as well as datestamping of periods during which there are significant findings of Granger-causal relationships. Use of the `tvgc` command is illustrated with an example from monthly U.S. macroeconomic data. The results obtained from these data support the conclusion that causal relationships can change dramatically over any given sample period. It follows that arbitrarily choosing the sample period over which to conduct causality tests is bound to be an inferior strategy to one that allows data-driven identification of change points.

8 Acknowledgments

We are grateful to Stephen P. Jenkins (Managing Editor) and an anonymous referee for useful comments and suggestions. Jesús Otero thanks the Universidad del Rosario and the Alianza EFI-Colombia Científica grant with code 60185 and contract number FP44842-220-2018. The usual disclaimer applies.

9 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 22-2
. net install st0675     (to install program files, if available)
. net get st0675         (to install ancillary files, if available)
```

10 References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Arora, V., and S. Shi. 2016. Energy consumption and economic growth in the United States. *Applied Economics* 48: 3763–3773. https://doi.org/10.1080/00036846.2016.1145347.

Baum, C. F. 2012. bcuse: Stata module to access instructional datasets on Boston College server. Statistical Software Components S457508, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s457508.html.

Campbell, J. Y., and P. Perron. 1991. Pitfalls and opportunities: What macroeconomists should know about unit roots. *NBER Macroeconomics Annual 1991* 6: 141–201. https://doi.org/10.1086/654163.

Chan, M. W. L., and C. Mountain. 1988. The interactive and causal relationships involving precious metal price movements: An analysis of the gold and silver markets. *Journal of Business and Economic Statistics* 6: 69–77. https://doi.org/10.2307/1391419.

Dickey, D. A., and W. A. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74: 427–431. https://doi.org/10.2307/2286348.

Dolado, J. J., and H. Lütkepohl. 1996. Making Wald tests work for cointegrated VAR systems. *Econometric Reviews* 15: 369–386. https://doi.org/10.1080/07474939608800362.

Drukker, D. M. 2006. Importing Federal Reserve economic data. *Stata Journal* 6: 384–386. https://doi.org/10.1177/1536867X0600600307.

Elliott, G., T. J. Rothenberg, and J. H. Stock. 1996. Efficient tests for an autoregressive unit root. *Econometrica* 64: 813–836. https://doi.org/10.2307/2171846.

Friedman, B. M., and K. N. Kuttner. 1993. Another look at the evidence on money-income causality. *Journal of Econometrics* 57: 189–203. https://doi.org/10.1016/0304-4076(93)90064-C.

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424–438. https://doi.org/10.2307/1912791.

———. 1988. Some recent development in a concept of causality. *Journal of Econometrics* 39: 199–211. https://doi.org/10.1016/0304-4076(88)90045-0.

Grossman, G. M., and A. B. Krueger. 1995. Economic growth and the environment. *Quarterly Journal of Economics* 110: 353–377. https://doi.org/10.2307/2118443.

Hall, A. 1994. Testing for a unit root in time series with pretest data-based model selection. *Journal of Business and Economic Statistics* 12: 461–470. https://doi.org/10.1080/07350015.1994.10524568.

Hamilton, J. D. 1983. Oil and the macroeconomy since World War II. *Journal of Political Economy* 91: 228–248. https://doi.org/10.1086/261140.

Jann, B. 2005. moremata: Stata module (Mata) to provide various functions. Statistical Software Components S455001, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s455001.html.

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lütkepohl, and T. C. Lee. 1988. *Introduction to the Theory and Practice of Econometrics*. 2nd ed. New York: Wiley.

Lee, C.-C. 2006. The causality relationship between energy consumption and GDP in G-11 countries revisited. *Energy Policy* 34: 1086–1093. https://doi.org/10.1016/j.enpol.2005.04.023.

Leybourne, S. J. 1995. Testing for unit roots using forward and reverse Dickey–Fuller regressions. *Oxford Bulletin of Economics and Statistics* 57: 559–571. https://doi.org/10.1111/j.1468-0084.1995.tb00040.x.

Otero, J., and C. F. Baum. 2017. Response surface models for the Elliott, Rothenberg, and Stock unit-root test. *Stata Journal* 17: 985–1002. https://doi.org/10.1177/1536867X1801700413.

———. 2018. Unit-root tests based on forward and reverse Dickey–Fuller regressions. *Stata Journal* 18: 22–28. https://doi.org/10.1177/1536867X1801800103.

Phillips, P. C. B., S. Shi, and J. Yu. 2014. Specification sensitivity in right-tailed unit root testing for explosive behaviour. *Oxford Bulletin of Economics and Statistics* 76: 315–333. https://doi.org/10.1111/obes.12026.

———. 2015a. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review* 56: 1043–1077. https://doi.org/10.1111/iere.12132.

———. 2015b. Testing for multiple bubbles: Limit theory of real time detectors. *International Economic Review* 56: 1079–1134. https://doi.org/10.1111/iere.12131.

Phillips, P. C. B., Y. Wu, and J. Yu. 2011. Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review* 52: 201–226. https://doi.org/10.1111/j.1468-2354.2010.00625.x.

Psaradakis, Z., M. O. Ravn, and M. Sola. 2005. Markov switching causality and the money-output relationship. *Journal of Applied Econometrics* 20: 665–683. https://doi.org/10.1002/jae.819.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461–464. https://doi.org/10.1214/aos/1176344136.

Shi, S., S. Hurn, and P. C. B. Phillips. 2020. Causal change detection in possibly integrated systems: Revisiting the money–income relationship. *Journal of Financial Econometrics* 18: 158–180. https://doi.org/10.1093/jjfinec/nbz004.

Shi, S., P. C. B. Phillips, and S. Hurn. 2018. Change detection and the causal impact of the yield curve. *Journal of Time Series Analysis* 39: 966–987. https://doi.org/10.1111/jtsa.12427.

Swanson, N. R. 1998. Money and output viewed through a rolling window. *Journal of Monetary Economics* 41: 455–474. https://doi.org/10.1016/S0304-3932(98)00005-1.

Tapia Granados, J. A., and E. L. Ionides. 2008. The reversal of the relation between economic growth and health progress: Sweden in the 19th and 20th centuries. *Journal of Health Economics* 27: 544–563. https://doi.org/10.1016/j.jhealeco.2007.09.006.

Thoma, M. A. 1994. Subsample instability and asymmetries in money-income causality. *Journal of Econometrics* 64: 279–306. https://doi.org/10.1016/0304-4076(94)90066-3.

Toda, H. Y., and T. Yamamoto. 1995. Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* 66: 225–250. https://doi.org/10.1016/0304-4076(94)01616-8.

Wlazlowski, S., B. Hagströömer, and M. Giulietti. 2011. Causality in crude oil prices. *Applied Economics* 43: 3337–3347. https://doi.org/10.1080/00036841003636250.

About the authors

Christopher F. Baum is a professor of economics and social work at Boston College, where he chairs the department of economics. He is an associate editor of the *Stata Journal*. Baum founded and manages the Boston College Statistical Software Components Archive at RePEc (http://repec.org). His recent research has addressed issues in social epidemiology and the progress of refugee immigrants in developed economies.

Stan Hurn is a professor of econometrics in the School of Economics and Finance at Queensland University of Technology and the director of the National Centre for Econometric Research in Australia.

Jesús Otero is a professor of economics at the Universidad del Rosario in Colombia. He received his PhD in economics from Warwick University in the United Kingdom. His field of study is applied time-series econometrics.