11-2018

Change detection and the causal impact of the yield curve

Shuping SHI

Peter C. B. PHILLIPS
Singapore Management University, peterphillips@smu.edu.sg

Stan HURN

Citation

SHI, Shuping; PHILLIPS, Peter C. B.; and HURN, Stan. Change detection and the causal impact of the yield curve. (2018). *Journal of Time Series Analysis*. 39, (6), 966-987.
Available at: https://ink.library.smu.edu.sg/soe_research/2349

Published Journal of Time Series Analysis, 2018, 39 (6), 966-987.
https://doi.org/10.1111/jtsa.12427

Submitted version

CHANGE DETECTION AND THE CAUSAL IMPACT OF THE YIELD CURVE

SHUPING SHI,[a] PETER C. B. PHILLIPS[b,c,d,e] AND STAN HURN[f]

[a] *Department of Economics, Macquarie University, Sydney, Australia*
[b] *Cowles Foundation, Yale University, New Haven, CT, USA*
[c] *Department of Economics, University of Auckland, Auckland, New Zealand*
[d] *Department of Economics, University of Southampton, Southampton, UK*
[e] *School of Economics, Singapore Management University Singapore, Singapore*
[f] *School of Economics and Finance, Queensland University of Technology, Brisbane, Australia*

Causal relationships in econometrics are typically based on the concept of predictability and are established by testing Granger causality. Such relationships are susceptible to change, especially during times of financial turbulence, making the real-time detection of instability an important practical issue. This article develops a test for detecting changes in causal relationships based on a recursive evolving window, which is analogous to a procedure used in recent work on financial bubble detection. The limiting distribution of the test takes a simple form under the null hypothesis and is easy to implement in conditions of homoskedasticity and conditional heteroskedasticity of an unknown form. Bootstrap methods are used to control family-wise size in implementation. Simulation experiments compare the efficacy of the proposed test with two other commonly used tests, the forward recursive and the rolling window tests. The results indicate that the recursive evolving approach offers the best finite sample performance, followed by the rolling window algorithm. The testing strategies are illustrated in an empirical application that explores the causal relationship between the slope of the yield curve and real economic activity in the United States over the period 1980–2015.

29 May 2017;

Keywords: Causality; forward recursion; hypothesis testing; recursive evolving test; rolling window; yield curve; real economic activity.

JEL. C12; C15; C32; G17.

MOS subject classification: 62M10; 62M20.

1. INTRODUCTION

Causality in econometrics typically relies on economic theory to justify the direction of causality between variables and to inform empirical testing of the causal hypotheses. In many situations, however, there is no relevant theoretical foundation for determining the causal ordering between variables that appear to be jointly determined. In these instances an empirical view of the concept of causality based on Granger (1969, 1988) has enjoyed widespread use. The popularity of Granger causality stems in part from the fact that it is not specific to a particular structural model but depends solely on the stochastic nature of variables, with no requirement to delimit some variables as dependent variables and others as independent variables.

In an early study of Granger causality in multiple time series models Newbold and Hotopp (1986) used a vector autoregressive moving average system with lag orders selected by information criteria to determine a fixed parameter system in which the tests were conducted. Since that early study it is has become well known that, among other things, testing for Granger causality is sensitive to the time period of estimation. Original contributions by

* Correspondence to: Peter C. B. Phillips, Yale University, POBox 208281, Cowles Foundation, 30 Hillhouse Avenue, New Haven, CT, USA. E-mail: peter.phillips@yale.edu

Thoma (1994) on the use of a forward expanding window for Granger causality testing method, and by Swanson (1998) on a rolling window version prompted interest in the problem of dealing with the time-varying nature of causal relationships in economics. See, for example, subsequent studies by Psaradakis *et al.* (2005), Balcilar *et al.* (2010) and Arora and Shi (2016). This article revisits the issue of time-varying Granger causality testing from the perspective of the recent literature for detecting and dating financial bubbles (Phillips *et al.*, 2011; Phillips *et al.*, 2015a,b, Phillips and Yu, 2011; Leybourne *et al.*, 2007. A new recursive test is proposed following the work of Phillips *et al.* (2015a,b) in the context of real-time detection of financial bubbles. The procedure involves intensive recursive calculations of the relevant test statistic, which in the current setting is a Wald test for Granger causality, in a backward expanding sample sequence in which the final observation of all samples is the (current) observation of interest. Inference regarding the presence of Granger causality for this observation relies on the supremum taken over the values of all the test statistics in the entire recursion. This procedure is therefore called a recursive evolving algorithm and its performance is thoroughly investigated and compared with the forward recursive and rolling window algorithms.

The time-varying Granger causality tests can be translated into a test for the joint significance of a subset of the model parameters against the alternative of these parameters being significant over either the whole or a fraction of the sample period. Two articles of significant relevance are Rossi (2005, 2013), which propose several tests for parameter restrictions taking potential parameter instability into consideration in a more general context. Importantly, the forward, rolling and recursive evolving procedures utilize only historical information and hence could serve as real-time monitoring devices for causality. The methods proposed by Rossi (2005), however, are ex-post testing procedures, which require estimating model parameters for each segment of the sample. Due to this particular requirement in constructing the tests, they cannot identify breaks occurring towards the end of the sample period – the primary interest of real-time identification. This method is, therefore, not used for comparative purposes in this article.

Asymptotic distributions under the null hypothesis of no Granger causality in a stationary system[1] are derived for the subsample Wald statistic for forward and rolling window versions of the test and the subsample sup Wald statistic for the recursive evolving window procedure. Limit theory under the assumption of homoskedasticity is provided first. To take potential influences of conditional heteroskedasticity[2] into account, heteroskedastic consistent versions of the Wald and sup Wald statistics are proposed. The asymptotic distributions of these test statistics are then derived under the assumption of conditional heteroskedasticity of unknown form. The major result for practical work that emerges from this limit theory is that the robust test statistics have the same pivotal asymptotics under homoskedasticity and conditional heteroskedasticity.

From an empirical perspective the many extensively studied problems in this area include: (i) the money–income relationship (Stock and Watson, 1989; Thoma, 1994; Swanson, 1998; Psaradakis *et al.*, 2005); (ii) the energy consumption and economic output relationship (Stern, 2000; Balcilar *et al.*, 2010; Arora and Shi, 2016); and (iii) the detection of changes in patterns of systemic risk (Billio *et al.*, 2012; Chen *et al.*, 2014). The present article employs change detection algorithms to investigate the causal relationship between the yield curve spread and real economic activity in the United States over the period 1980–2015. Of particular importance to the current research is the finding in the empirical literature that the predictive relationship between the slope of the yield curve and macroeconomic activity has not been constant over time (Haubrich and Dombrosky, 1996; Dotsey, 1998; Stock and Watson, 1999; Estrella *et al.*, 2003; Chauvet and Potter, 2005; Giacomini and Rossi, 2006; Benati and Goodhart, 2008; Chauvet and Senyuz, 2016; Chinn and Kucko, 2015; Hamilton, 2011). The test procedures developed here provide a natural mechanism for causal detection that allows such temporal fragilities in causal relationships to be captured through intensive subsample data analysis.

The article is organized as follows. Section 2 reviews the concept of Granger causality and describes the forward expanding window, rolling window, and the new recursive evolving Granger causality tests. Section 3 gives the

[1] For the analysis of a possibly integrated system, see Shi *et al.* (2018a). Simulation results suggest that deviations from the stationarity assumption (or the presence of integrated variables in the system) may result in a loss of power in the testing procedures.

[2] The presence of conditional heteroskedasticity in financial and macroeconomic data series has been well documented in the literature. See, for example, Engle and Bollerslev (1986), Bollerslev (1987), Nelson (1991), and Elder (2004), and Yogo (2004).

limit distributions of these test statistics under the null hypothesis of no causality and assumptions of homoskedasticity and conditional heteroskedasticity. The focus of the present article is on developing the new real-time causal identification procedure, and proofs are given in the online supplement (Shi *et al.*, 2018b) accompanying this paper. Section 4 reports the results of simulations investigating performance characteristics of the various tests. In Section 5, we use the three procedures to investigate the causal relationship between the yield curve spread and real economic activity in the United States over the last three decades. Section 6 concludes. Robustness checks are provided in the supplement.[3]

2. IDENTIFYING CHANGES IN CAUSAL RELATIONSHIPS

The unrestricted VAR(p) may be written as

$$\mathbf{y}_t = \Phi_0 + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \varepsilon_t, \tag{1}$$

or in multi-variate regression format simply as

$$\mathbf{y}_t = \Pi \mathbf{x}_t + \varepsilon_t, \quad t = 1, \ldots, T \tag{2}$$

where $\mathbf{y}_t = \left(y_{1t}, y_{2t}\right)'$, $\mathbf{x}_t = \left(1, \mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_{t-p}\right)'$, and $\Pi_{2\times(2p+1)} = \left[\Phi_0, \Phi_1, \ldots, \Phi_p\right]$. Let $\widehat{\Pi}$ be the ordinary least squares estimator of Π, $\hat{\Omega} = T^{-1}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}'_t$ with $\hat{\varepsilon}_t = \mathbf{y}_t - \widehat{\Pi}\mathbf{x}_t$, and $\mathbf{X}' = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$ be the observation matrix of the regressors in (2). The Wald test of the restrictions imposed by the null hypothesis that y_{2t} does not cause y_{1t} in Granger's sense, or $H_0 : y_{2t} \nrightarrow y_{1t}$, has the simple form

$$\mathcal{W} = \left[\mathbf{R}\,\mathrm{vec}(\widehat{\Pi})\right]' \left[\mathbf{R}\left(\hat{\Omega} \otimes \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)\mathbf{R}'\right]^{-1} \left[\mathbf{R}\,\mathrm{vec}(\widehat{\Pi})\right], \tag{3}$$

where $\mathrm{vec}(\widehat{\Pi})$ denotes the (row vectorized) $2\,(2p+1) \times 1$ coefficients of $\widehat{\Pi}$ and $\mathbf{R}$ is the $p \times 2(2p+1)$ selection matrix. Each row of $\mathbf{R}$ picks one of the coefficients to set to zero under the non-causal null hypothesis. In the present case these are the p coefficients on the lagged values of y_{2t} in (2). Under the null hypothesis and assumption of conditional homoskedasticity, the Wald test statistic is asymptotically χ^2_p, with degrees of freedom corresponding to the number of zero restrictions being tested. See, for example, Granger and Newbold (1986), for more details.

As indicated in the introductory remarks, there is ample reason to expect causal relationships to change over time, because of changes in economic policy, regulatory structure, governing institutions, or operating environments that impinge on the variables y_{1t} and y_{2t}. In these circumstances, testing that is based on the entire sample using a statistic like (3), averages the sample information and inevitably destroys potentially valuable economic intelligence concerning the impact of changes in policy or structures. Testing for Granger casualty in exogenously defined subsamples of the data does provide useful information but does not enable the data to reveal the changes or change points. Consequently, the ultimate objective is to conduct tests that allow the change points to be determined (and hence identified) endogenously in the sample data.

Thoma (1994) and Swanson (1998) respectively, suggest using forward expanding and rolling window Wald tests to detect changes in causal relationships. It is convenient to work in terms of sample fractions in the following exposition. Let f be the (fractional) observation of interest and f_0 be the minimum (fractional) window size required to estimate the model. For the Thoma procedure, the starting point of the regression (f_1) is fixed on the first available observation. As the observation of interest f moves forward from f_0 (the minimum required for the regression) to 1, the regression window size expands (fractionally) from f_0 to 1 and hence the test arising from this approach is referred to as a forward expanding window Wald test. The Swanson rolling procedure, by contrast, keeps the size

[3] Data and Matlab code for the empirical application are available for download from https://sites.google.com/site/shupingshi/GC_DataProgram.zip?attredirects=0&d=1attredirects=0&d=1.

of the regression window (f_w) constant in the sequence of regressions. As the observation of interest (f and hence the terminal point of the regression f_2) rolls forward from f_0 to 1, the starting point follows accordingly, maintaining a fixed distance from f_2. A significant change in causality is detected when an element of the Wald statistic sequence exceeds its corresponding critical value, so that the origination (termination) date of a change in causality is identified as the first observation whose test statistic value exceeds (goes below) its corresponding critical value.

Drawing from the recent literature on dating multiple financial bubbles (Phillips *et al.*, 2015a, 2015b), this article suggests an additional test that is based on the supremum norm (sup) of a series of recursively evolving Wald statistics that are calculated as follows. For each (fractional) observation of interest ($f \in [f_0, 1]$), where f_0 is again the minimum sample size to accommodate the regression, the Wald statistics are computed for a sequence of backward expanding samples. As above, the end point of the sample sequences is temporarily fixed at the latest observation under study $f = f_2$ and evolves forward from this point. However, the starting points of the sample sequences used in these regressions extend backwards from $f_1 = (f_2 - f_0)$ all the way to the first observation (represented by the sample fraction 0). The Wald statistic obtained for each subsample regression (using observations over $[f_1, f_2]$ with a sample size fraction of $f_w = f_2 - f_1 \geq f_0$) is denoted by $\mathcal{W}_{f_2}(f_1)$ and the sup Wald statistic (up to the latest observation corresponding to $f = f_2$) is defined as

$$\mathcal{SW}_f(f_0) = \sup_{(f_1, f_2) \in \Lambda_0, f_2 = f} \left\{ \mathcal{W}_{f_2}(f_1) \right\},$$

where $\Lambda_0 = \left\{ (f_1, f_2) : 0 < f_0 + f_1 \leq f_2 \leq 1, \text{ and } 0 \leq f_1 \leq 1 - f_0 \right\}$ for some (given) minimal sample size $f_0 \in (0, 1)$ in the regressions. We call this procedure the recursive evolving procedure. Unlike the rolling window approach, this procedure allows variation in the window widths $f_w = f_2 - f_1 \geq f_0$ used in the regressions.

Both the forward expanding and rolling window procedures are special cases of the new procedure: the forward expanding window has $f_1 = 0$ fixed and sets $f = f_2$; the rolling window has fixed window width $f_w = f_2 - f_1 = f_0$ (assuming the rolling window width is fixed to f_0) and window initialization $f_1 = f_2 - f_0$. Importantly, all three procedures rely only on past information and can therefore be used for real-time monitoring at the present observation f. The added flexibility obtained by relaxing f_1 allows the procedure to search for the optimum starting point of the regression for each observation (in the sense of finding the largest Wald statistic). This flexibility accommodates re-initialization in the subsample to accord with (and thereby help to detect) any changes in structure or causal direction that may occur within the full sample.

Let f_e and f_f denote the origination and termination points in the causal relationship. These are estimated as the first chronological observation whose test statistic respectively exceeds or falls below the critical value. In a single switch case, the dating rules are giving by the following crossing times:

$$\text{Forward} : \hat{f}_e = \inf_{f \in [f_0, 1]} \left\{ f : \mathcal{W}_f(0) > cv \right\} \text{ and } \hat{f}_f = \inf_{f \in [\hat{f}_e, 1]} \left\{ f : \mathcal{W}_f(0) < cv \right\}, \tag{4}$$

$$\text{Rolling} : \hat{f}_e = \inf_{f \in [f_0, 1]} \left\{ f : \mathcal{W}_f(f - f_0) > cv \right\} \text{ and } \hat{f}_f = \inf_{f \in [\hat{f}_e, 1]} \left\{ f : \mathcal{W}_f(f - f_0) < cv \right\}, \tag{5}$$

$$\text{Recursive Evolving} : \hat{f}_e = \inf_{f \in [f_0, 1]} \left\{ f : \mathcal{SW}_f(f_0) > scv \right\} \text{ and } \hat{f}_f = \inf_{f \in [\hat{f}_e, 1]} \left\{ f : \mathcal{SW}_f(f_0) < scv \right\}, \tag{6}$$

where cv and scv are the corresponding critical values of the $\mathcal{W}_f$ and $\mathcal{SW}_f$ statistics. The origination and termination dates are estimated in a similar fashion when there are multiple switches in the sample period. We search the origination and termination dates of episode i with $i \geq 2$ in the sample ranges of $[\hat{f}_{i-1f}, 1]$ and $[\hat{f}_{ie}, 1]$ respectively.

3. ASYMPTOTIC DISTRIBUTIONS

The notation introduced in the previous section is now used for the general n dimensional multi-variate case of (2), which allows both for changing coefficients in subsamples of the data and for changing (fractional) sample sizes in the asymptotic theory.

Let $\|\cdot\|$ denote the Euclidean norm, $\|.\|_p$ be the L_p-norm so that $\|\mathbf{x}\|_p = \left(\mathbb{E}\|\mathbf{x}\|^p\right)^{1/p}$, and $\mathcal{F}_t = \sigma\left\{\varepsilon_t, \varepsilon_{t-1} \ldots\right\}$ be the natural filtration. Consider a $n \times 1$ vector of dependent variables $\mathbf{y}_t$ whose dynamics follow a VAR(p) given by

$$\mathbf{y}_t = \Phi_0 + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \varepsilon_t, \tag{7}$$

with constant coefficients over the subsample $t = \lfloor Tf_1 \rfloor, \ldots, \lfloor Tf_2 \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. The sample size in this regression is $T_w = \lfloor Tf_w \rfloor$ where $f_w \in \left[f_0, 1\right]$ for some fixed $f_0 \in (0,1)$.

Assumption A0. The roots of $\left|\mathbf{I}_n - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p\right| = 0$ lie outside the unit circle.

Under assumption **A0**, $\mathbf{y}_t$ has a simple moving average (linear process) representation in terms of the past history of ε_t

$$\mathbf{y}_t = \tilde{\Phi}_0 + \Psi(L)\,\varepsilon_t,$$

where $\Psi(L) = \left(\mathbf{I}_n - \Phi_1 L - \Phi_2 L^2 \ldots - \Phi_p L^p\right)^{-1} = \sum_{i=0}^{\infty} \Psi_i L^i$ with $\left\|\Psi_i\right\| < C\theta^i$ for some $\theta \in (0,1)$ and $\tilde{\Phi}_0 = \Psi(1)\,\Phi_0$. The model can be written in regression format as

$$\mathbf{y}_t = \Pi_{f_1,f_2} \mathbf{x}_t + \varepsilon_t, \tag{8}$$

in which $\mathbf{x}_t = \left(1, \mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_{t-p}\right)'$ and $\Pi_{f_1,f_2} = \left[\Phi_0, \Phi_1, \ldots, \Phi_p\right]$.

The ordinary least squares (or Gaussian maximum likelihood with fixed initial conditions) estimator of the autoregressive coefficients is denoted by $\hat{\Pi}_{f_1,f_2}$ and defined as

$$\underset{n\times(np+1)}{\hat{\Pi}_{f_1,f_2}} = \left[\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{y}_t \mathbf{x}'_t\right] \left[\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t \mathbf{x}'_t\right]^{-1}.$$

Let $k = np+1$ and let $\hat{\pi}_{f_1,f_2} \equiv \operatorname{vec}\left(\hat{\Pi}_{f_1,f_2}\right)$ denote the (row vectorized) $nk \times 1$ coefficients resulting from an ordinary least squares regression of each of the elements of $\mathbf{y}_t$ on $\mathbf{x}_t$ for a sample running from $\lfloor Tf_1 \rfloor$ to $\lfloor Tf_2 \rfloor$ given by

$$\hat{\pi}_{f_1,f_2} = \left[\begin{array}{cccc} \hat{\pi}_{1,f_1,f_2} & \hat{\pi}_{2,f_1,f_2} & \ldots & \hat{\pi}_{n,f_1,f_2} \end{array}\right]',$$

in which $\hat{\pi}_{i,f_1,f_2} = \left[\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} y_{it} \mathbf{x}'_t\right] \left[\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t \mathbf{x}'_t\right]^{-1}$. It follows that

$$\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} = \left[\mathbf{I}_n \otimes \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t \mathbf{x}'_t\right]^{-1} \left[\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \xi_t\right], \tag{9}$$

where π_{f_1,f_2} denotes the population coefficient of $\hat{\pi}_{f_1,f_2}$ and $\xi_t \equiv \varepsilon_t \otimes \mathbf{x}_t$. The estimator of the residual variance matrix is $\hat{\Omega}_{f_1,f_2} = T_w^{-1} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \hat{\varepsilon}_t \hat{\varepsilon}'_t$, where $\hat{\varepsilon}'_t = \left[\hat{\varepsilon}_{1t}, \hat{\varepsilon}_{2t}, \ldots, \hat{\varepsilon}_{mt}\right]$ and $\hat{\varepsilon}_{it} = y_{it} - \mathbf{x}'_t \hat{\pi}_{i,f_1,f_2}$. The final factor $\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \xi_t$ on the right side of (9) may be interpreted as the simple composition of functionals $\sum_{t=1}^{\lfloor Tf_2 \rfloor} \xi_t - \sum_{t=1}^{\lfloor Tf_1 \rfloor - 1} \xi_t$ of the partial sum process $\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_t$ defined on the product space $D[0,1]^{nk}$. This interpretation is useful in developing limit theory for statistics based on the recursively evolving regression coefficients $\hat{\pi}_{f_1,f_2}$.

The primary concern is the distribution of the Wald statistic for testing causality under the null hypothesis. In this instance, the coefficient matrix Π_{f_1,f_2} is constant for the entire sample $[f_1,f_2]$. The null hypothesis for the causality test falls in the general framework of linear hypotheses of the form $H_0 : \mathbf{R}\pi_{f_1,f_2} = 0$, where $\mathbf{R}$ is a coefficient restriction matrix (of full row rank d). Given (f_1,f_2), the usual form of the Wald statistic for this null hypothesis is

$$\mathcal{W}_{f_2}(f_1) = \left(\mathbf{R}\hat{\pi}_{f_1,f_2}\right)' \left\{ \mathbf{R} \left[\hat{\mathbf{\Omega}}_{f_1,f_2} \otimes \left(\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \right] \mathbf{R}' \right\}^{-1} \left(\mathbf{R}\hat{\pi}_{f_1,f_2}\right). \tag{10}$$

3.1. Homoskedasticity

Under the assumption of homoskedasticity, the innovations are stationary, conditionally homoskedastic martingale differences satisfying either of the following two conditions.

Assumption A1. $\{\varepsilon_t, \mathcal{F}_t\}$ is a strictly stationary and ergodic martingale difference sequence (mds) with $\mathbb{E}\left(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\right) = \mathbf{\Omega}$ *a.s.* and positive definite Ω.

Assumption A2. $\{\varepsilon_t, \mathcal{F}_t\}$ is a covariance stationary mds with $\mathbb{E}\left(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\right) = \mathbf{\Omega}$ *a.s.*, positive definite $\mathbf{\Omega}$, and $\sup_t \mathbb{E}\left\|\varepsilon_t\right\|^{4+c} < \infty$ for some $c > 0$.

Lemma 3.1. Given the model (7), under assumption **A0** and **A1** or **A2** and the null (maintained) hypothesis of an unchanged coefficient matrix $\Pi_{f_1,f_2} = \Pi$ for all (fractional) subsamples $(f_1,f_2) \in \Lambda_0$ we have

(i) $\hat{\pi}_{f_1,f_2} \to_{a.s.} \pi_{f_1,f_2} = \pi$,
(ii) $\hat{\mathbf{\Omega}}_{f_1,f_2} \to_{a.s.} \mathbf{\Omega}$,
(iii) $\sqrt{T}\left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right) \Rightarrow \left[\mathbf{I}_n \otimes \mathbf{Q}\right]^{-1} \left[\frac{B(f_2) - B(f_1)}{f_w}\right]$,

where B is vector Brownian motion with covariance matrix $\mathbf{\Omega}\otimes\mathbf{Q}$, where $\mathbf{Q} = \mathbb{E}\left(\mathbf{x}_t\mathbf{x}_t'\right) > 0$, and $\hat{\pi}_{f_1,f_2}$ and $\hat{\mathbf{\Omega}}_{f_1,f_2}$ are the least squares estimators of π_{f_1,f_2} and $\mathbf{\Omega} = \mathbb{E}\left(\varepsilon_t\varepsilon_t'\right)$. The co-domain of the limit in (iii) is the subspace $D\left(\Lambda_0\right)$ of the Skorohod space $D[0,1]^{nk}$ equipped with the uniform metric. The finite dimensional distribution of the limit in (iii) for fixed f_2 and f_1 is $N\left(\mathbf{0}, \frac{1}{f_w}\mathbf{\Omega}\otimes\mathbf{Q}^{-1}\right)$.

From part (iii) and for fixed (f_1,f_2) the asymptotic covariance matrix of $\sqrt{T}\left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right)$ is $f_w^{-1}\left(\mathbf{\Omega}\otimes\mathbf{Q}^{-1}\right)$, which is dependent on the fractional window size f_w. The limit in (iii) may be interpreted as a double indexed process with parameters $(f_1,f_2) \in \Lambda_0$ that is linearly dependent on the single indexed stochastic process B. The proof of Lemma 3.1 (iii) uses an argument in which a standardized version of the estimation error (9) is written as an indexed composition of continuous functionals of the partial sum process $X_T^0(\cdot) := \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor T\cdot\rfloor}\xi_t$ which satisfies the weak convergence $X_T^0(\cdot) \Rightarrow B(\cdot)$ on the product Skorohod space $D[0,1]^{nk}$. This representation in terms of a continuous functional facilitates the asymptotic development of further continuous functionals such as the sup Wald statistic in the following parameters $(f_1,f_2) \in \Lambda_0$.

Proposition 3.1. Under **A0** and **A1** or **A2**, the null hypothesis $\mathbf{R}\pi_{f_1,f_2} = \mathbf{0}$, and the maintained null of an unchanged coefficient matrix $\Pi_{f_1,f_2} = \Pi$ for all subsamples, the subsample Wald process and sup Wald statistic converge weakly to the following limits

$$\mathcal{W}_{f_2}(f_1) \Rightarrow \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right]' \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right] \tag{11}$$

$$\mathcal{SW}_f(f_0) \Rightarrow \sup_{(f_1,f_2)\in\Lambda_0, f_2=f} \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right]' \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right] \tag{12}$$

where W_d is vector Brownian motion with covariance matrix $\mathbf{I}_d$ and d is the number of restrictions (the rank of R) under the null.

The limit process that appears in (11) is a quadratic functional of $W_d(\cdot)$. Its finite dimensional distribution for fixed f_1 and f_2 is χ^2_d, whereas the sup functional that appears in (12) involves the supremum of a continuous functional taken over $(f_1, f_2) \in \Lambda_0$ of the stochastic process $X^0_T(\cdot)$.

Consider a simple alternative hypothesis of causality with a structural break such that

$$\mathbf{y}_t = \begin{cases} \Pi\mathbf{x}_t + \varepsilon_t, & \text{if } t_1 \le t \le t_e \\ \Pi^*\mathbf{x}_t + \varepsilon_t, & \text{if } t_e < t \le t_2 \end{cases} \tag{13}$$

where $t_e = \lfloor Tf_e \rfloor$ is the break point. Let $\pi \equiv \text{vec}(\Pi)$ and $\pi^* \equiv \text{vec}(\Pi^*)$. We assume that $\mathbf{R}\pi = \mathbf{0}$ and $\mathbf{R}\pi^* = \mathbf{g} > \mathbf{0}$, i,e. the causality is switched on at t_e. Notice that model (13) collapses to model (8) under the null when $t_e = t_2$. Under the data generating process of (13) assuming $t_e < t_2$, the ordinary least squares estimator $\hat{\pi}_{f_1,f_2}$ becomes

$$\begin{aligned} \hat{\pi}_{f_1,f_2} = &\left[I_n \otimes \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t\mathbf{x}_t'\right]^{-1} \left[\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \xi_t\right] + \\ &\left[I_n \otimes \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t\mathbf{x}_t'\right]^{-1} \left[\left(I_n \otimes \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_e \rfloor} \mathbf{x}_t\mathbf{x}_t'\right)\pi + \left(I_n \otimes \sum_{t=\lfloor Tf_e \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t\mathbf{x}_t'\right)\pi^*\right]. \end{aligned} \tag{14}$$

We can show that under the assumptions of **A0** and **A1** or **A2**, the ordinary least squares estimator in (14) converges to a weighted average of π and π^* such that

$$\hat{\pi}_{f_1,f_2} \to^{a.s} \bar{\pi}_{f_1,f_2} \equiv \pi(f_e - f_1) + \pi^*(f_2 - f_e).$$

The error variance estimator $\hat{\Omega}_{f_1,f_2} \to^{a.s.} \Lambda + \Omega$, where Λ depends on the break location f_e and the coefficients π and π^*. This shows that the least squares estimates $\hat{\pi}_{f_1,f_2}$ and $\hat{\Omega}_{f_1,f_2}$ are inconsistent under the alternative of parameter instability, as in Rossi (2005, 2013). Under the alternative $\mathbf{R}\pi = \mathbf{0}$ and $\mathbf{R}\pi^* = \mathbf{g} > \mathbf{0}$, it follows that the test statistics $\mathcal{W}_{f_2}(f_1)$ and $\mathcal{SW}_{f_2}(f_1)$ have order $O_p(T)$ and diverge, in contrast to the well-defined limit distributions that apply under the null as given in Proposition 3.1. Following similar arguments to those in Phillips *et al.* (2015b), the break date (or the causality origination date) estimator $\hat{f}_e$ can be shown to be consistent provided that the critical values have order greater than $O_p(1)$ and smaller than $O_p(T)$. A full investigation of the asymptotic behavior of the test statistics under the alternative and the consistency of the causality dating algorithm will be provided in later work on a separate article.

3.2. Conditional Heteroskedasticity of an Unknown Form

In this subsection, the impact of conditional heteroskedasticity on the limiting distributions of the standard Wald and sup Wald statistics is investigated. The conditional heteroskedasticity case requires the following additional assumption.

Assumption A3. $\left\{\varepsilon_t, \mathcal{F}_t\right\}$ is an mds satisfying the following conditions:

(i) ε_t is strongly uniformly integrable with a dominating random variable ε that satisifies $\mathbb{E}\left(\|\varepsilon\|^{4+c}\right) < \infty$ for some $c > 0$;
(ii) $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\right) \to_{a.s.} \mathbf{\Omega}$ where $\mathbf{\Omega}$ is positive definite with elements $\mathbf{\Omega} = \left(\Omega_{ij}\right)$;
(iii) $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{i,t}^2|\mathcal{F}_{t-1}\right)\varepsilon_{t-s} \to_{a.s.} 0$ and $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{i,t}\varepsilon_{j,t}|\mathcal{F}_{t-1}\right)\varepsilon_{t-s}\varepsilon_{t-s}' \to_{a.s.} \Gamma_s^{(i,j)}$ for $i,j = 1,\ldots,n$ and $s \geq 1$.

Strong uniform integrability is commonly assumed in cases of conditional and unconditional heterogeneity (see, for instance, Phillips and Solo (1992), Remarks 2.4(i) and 2.8 (i) and (ii)). Assumption **A3** implies that $\left\{\varepsilon_t\right\}$ is serially uncorrelated, unconditionally homoskedastic if $\mathbb{E}\left(\varepsilon_t\varepsilon_t'\right) = \mathbf{\Omega}$ for all t (and hence covariance stationary in that case), but is otherwise potentially conditionally heteroskedastic. **A3** allows, among other possibilities, stable ARCH or GARCH errors. Note that **A3**(i) is equivalent to assuming that

$$\sup_t \mathbb{E}\left\|\varepsilon_t\right\|^{4+c} < \infty \text{ for some } c > 0,$$

a condition that is often used in work involving heteroskedasticity (see, for example, Boswijk *et al.* (2016), Patilea and Rassi (2012), and Bodnar and Zabolotskyy (2011)). **A3**(iii) is required for Lemma 3.3(ii), and is similar to the condition used by Hannan and Heyde (1972, Theorem 2), Gonçalves and Kilian (2004), and Boswijk *et al.* (2016) for the univariate case.

Lemma 3.2. Under **A0** and **A3**, for all $f_2, f_1 \in [0,1]$ and $f_2 > f_1$,

(i) $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\varepsilon_t \to_{a.s} \mathbf{0}$,
(ii) $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\varepsilon_t\varepsilon_t' \to_{a.s} \mathbf{\Omega}$,
(iii) $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\varepsilon_t\varepsilon_s' \to_{a.s} \mathbf{0}$,
(iv) $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\mathbf{x}_t\varepsilon_t' \to_{a.s} \mathbf{0}$,
(v) $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\mathbf{x}_t\mathbf{x}_t' \to_{a.s} \mathbf{Q}$, where $\mathbf{Q}$ is defined as

$$\mathbf{Q} \equiv \begin{bmatrix} 1 & \mathbf{1}_p' \otimes \tilde{\Phi}_0' \\ \mathbf{1}_p \otimes \tilde{\Phi}_0 & \mathbf{I}_p \otimes \tilde{\Phi}_0\tilde{\Phi}_0' + \Theta \end{bmatrix} \text{ with } \Theta = \sum_{i=0}^{\infty}\begin{bmatrix} \Psi_i\mathbf{\Omega}\Psi_i' & \cdots & \Psi_{i+p-1}\mathbf{\Omega}\Psi_i' \\ \vdots & \ddots & \vdots \\ \Psi_i\mathbf{\Omega}\Psi_{i+p-1}' & \cdots & \Psi_i\mathbf{\Omega}\Psi_i' \end{bmatrix}.$$

In view of the covariance stationarity of ε_t, Lemma 3.2 holds for all possible fixed fractions of data with $(f_1, f_2) \in \Lambda_0$. However, this is not in general true under global covariance stationarity (Davidson, 1994) or nonstationary volatility settings, where the right-hand side of the statements in Lemma 3.2 may depend on f_1 and f_2.

Recalling that $\xi_t \equiv \varepsilon_t \otimes x_t$, partial sums of the time series $\left\{\xi_t\right\}$ continue to obey a martingale invariance principle under **A0** and **A3**, as in Theorem 3 of Brown (1971). This invariance result relies on the validity of the two standard conditions given in Lemma 3.3.

Lemma 3.3. Under **A0** and **A3**, the mds $\left\{\xi_t, \mathcal{F}_t\right\}$ satisfies the following Lindeberg and stability conditions:

(a) For every $\delta > 0$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\left\|\xi_t\right\|^2 .\mathbf{1}\left(\left\|\xi_t\right\| \geq \sqrt{T}\delta\right)|\mathcal{F}_{t-1}\right\} \xrightarrow{p} 0, \tag{15}$$

(b) $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left\{\xi_t\xi_t'|\mathcal{F}_{t-1}\right\} \to_{a.s} \mathbf{\Sigma}$, where $\mathbf{\Sigma} = \left\{\mathbf{\Sigma}^{(i,j)}\right\}_{i,j\in[1,n]}$ with block partitioned elements

$$\mathbf{\Sigma}^{(i,j)} = \begin{bmatrix} \Omega_{ij} & \mathbf{1}_p' \otimes \Omega_{ij}\tilde{\Phi}_0' \\ \mathbf{1}_p \otimes \Omega_{ij}\tilde{\Phi}_0 & \mathbf{I}_p \otimes \Omega_{ij}\tilde{\Phi}_0\tilde{\Phi}_0' + \Xi^{(i,j)} \end{bmatrix}$$

and

$$\mathbf{\Xi}^{(i,j)} \equiv \sum_{i=0}^{\infty}\begin{bmatrix} \Psi_i\Gamma^{(i,j)}_{h+j+i}\Psi_i' & \cdots & \Psi_{i+p-1}\Gamma^{(i,j)}_{h+j+i}\Psi_i' \\ \vdots & \ddots & \vdots \\ \Psi_i\Gamma^{(i,j)}_{h+j+i}\Psi_{i+p-1}' & \cdots & \Psi_i\Gamma^{(i,j)}_{h+j+i}\Psi_i' \end{bmatrix}.$$

Under Lemma 3.3, partial sums of $\left\{\xi_t\right\}$ satisfy $X_T^0(\cdot) := \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor T\cdot\rfloor}\xi_t \Rightarrow B(\cdot)$, so that

$$\frac{1}{\sqrt{T}}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\xi_t \Rightarrow B\left(f_2\right) - B\left(f_1\right). \tag{16}$$

The limit in (16) is a linear functional of the vector Brownian motion B with covariance matrix $\mathbf{\Sigma}$.

Lemma 3.4. Under **A0** and **A3**,

(a) $\hat{\pi}_{f_1,f_2} \to_{a.s.} \pi_{f_1,f_2}$,
(b) $\hat{\Omega}_{f_1,f_2} \to_{a.s.} \mathbf{\Omega}$,
(c) $\sqrt{T_w}\left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right) \Rightarrow f_w^{-1/2}\mathbf{V}^{-1}\left[\mathbf{B}\left(f_2\right) - \mathbf{B}\left(f_1\right)\right]$, where $\mathbf{V} = \mathbf{I}_n \otimes \mathbf{Q}$ and B is vector Brownian motion with covariance matrix $\mathbf{\Sigma}$.
(d) $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\hat{\xi}_t\hat{\xi}_t' \to_{a.s.} \mathbf{\Sigma}$, where $\hat{\xi}_t \equiv \hat{\varepsilon}_t \otimes \mathbf{x}_{t-1}$.

In the presence of conditional heteroskedasticity, the Wald and sup Wald statistic have non-standard, non-pivotal asymptotic distributions as detailed in the following result.

Proposition 3.2. Under the assumption of conditional heteroskedasticity of unknown form (**A0** and **A3**), the null hypothesis $\mathbf{R}\pi_{f_1,f_2} = \mathbf{0}$, and the maintained hypothesis of an unchanged coefficient matrix $\Pi_{f_1,f_2} = \Pi$ for all subsamples, the subsample Wald and sup Wald statistics have the following limits

$$\mathcal{W}_{f_2}\left(f_1\right) \Rightarrow \left[\frac{W_d\left(f_2\right) - W_d\left(f_1\right)}{\left(f_2 - f_1\right)^{1/2}}\right]' \mathbf{A}\mathbf{B}^{-1}\mathbf{A}' \left[\frac{W_d\left(f_2\right) - W_d\left(f_1\right)}{\left(f_2 - f_1\right)^{1/2}}\right],$$

$$\mathcal{SW}_f\left(f_0\right) \Rightarrow \sup_{(f_1,f_2)\in\Lambda_0, f_2 = f}\left[\frac{W_d\left(f_2\right) - W_d\left(f_1\right)}{\left(f_2 - f_1\right)^{1/2}}\right]' \mathbf{A}\mathbf{B}^{-1}\mathbf{A}' \left[\frac{W_d\left(f_2\right) - W_d\left(f_1\right)}{\left(f_2 - f_1\right)^{1/2}}\right],$$

where W_d is vector Brownian motion with covariance matrix $\mathbf{I_d}$, $\mathbf{A} = \mathbf{\Sigma}^{1/2}\mathbf{V}^{-1}\mathbf{R}'$, and $\mathbf{B} = \mathbf{R}\left(\mathbf{\Omega} \otimes \mathbf{Q}\right)\mathbf{R}'$.

The presence of conditionally heterogeneous errors affects the limit behavior of the standard Wald statistic, which no longer has the simple limit distribution (11). In consequence, use of the limit theory (12) for the sup Wald statistic may lead to invalid and distorted inference. A pivotal version of the statistic is obtained by suitable asymptotic covariance matrix estimation.

3.3. Heteroskedastic-Consistent Test Statistics

The heteroskedasticity consistent version of the Wald statistic is denoted by $\mathcal{W}^*_{f_2}(f_1)$ and defined as

$$\mathcal{W}^*_{f_2}(f_1) = T_w \left(\mathbf{R}\hat{\pi}_{f_1,f_2}\right)' \left[\mathbf{R}\left(\hat{\mathbf{V}}^{-1}_{f_1,f_2}\hat{\mathbf{\Sigma}}_{f_1,f_2}\hat{\mathbf{V}}^{-1}_{f_1,f_2}\right)\mathbf{R}'\right]^{-1} \left(\mathbf{R}\hat{\pi}_{f_1,f_2}\right), \tag{17}$$

where $\hat{\mathbf{V}}_{f_1,f_2} \equiv \mathbf{I}_n \otimes \hat{\mathbf{Q}}_{f_1,f_2}$ with $\hat{\mathbf{Q}}_{f_1,f_2} \equiv \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor} \mathbf{x}_t\mathbf{x}_t'$, and $\hat{\mathbf{\Sigma}}_{f_1,f_2} \equiv \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor} \hat{\xi}_t\hat{\xi}_t'$ with $\hat{\xi}_t \equiv \hat{\epsilon}_t \otimes \mathbf{x}_t$. The corresponding heteroskedasticity consistent sup Wald statistic is

$$\mathcal{SW}^*_f(f_0) := \sup_{(f_1,f_2)\in\Lambda_0, f_2=f} \left\{\mathcal{W}^*_{f_2}(f_1)\right\}.$$

As discussed following Lemma 3.1 and Proposition 3.1, for the purpose of deriving asymptotics the statistics $\mathcal{W}_{f_2}(f_1)$ and $\mathcal{W}^*_{f_2}(f_1)$ may be treated as functionals (indexed by the sample fractions (f_1,f_2)) of the stochastic process X^0_T. The sup statistics $\mathcal{SW}^*_f(f_0)$ and $\mathcal{SW}_f(f_0)$ are then composite functionals of X^0_T. In the Online Supplement this approach to the asymptotic theory uses the continuity of these functionals and the weak convergence $X^0_T(\cdot) \Rightarrow B(\cdot)$ to establish the limit theory.

Note that under the homoskedasticity assumption of **A1** or **A2**, the limit of the matrix $\hat{\mathbf{\Sigma}}_{f_1,f_2}$ that appears in the heteroskedastic consistent Wald statistic (17) would be given by $\mathbf{\Omega}\otimes\mathbf{Q}$ and the asymptotic covariance matrix would simplify as follows

$$\hat{\mathbf{V}}^{-1}_{f_1,f_2}\hat{\mathbf{\Sigma}}_{f_1,f_2}\hat{\mathbf{V}}^{-1}_{f_1,f_2} \to_{a.s} \left(\mathbf{I}_n\otimes\mathbf{Q}\right)^{-1}\left(\mathbf{\Omega}\otimes\mathbf{Q}\right)\left(\mathbf{I}_n\otimes\mathbf{Q}\right)^{-1} = \mathbf{\Omega}\otimes\mathbf{Q}^{-1}.$$

In this case, therefore, the heteroskedastic consistent test statistics, $\mathcal{W}^*_{f_2}(f_1)$ and $\mathcal{SW}^*_f(f_0)$, reduce to the conventional Wald and sup Wald statistics of $\mathcal{W}_{f_2}(f_1)$ and $\mathcal{SW}_f(f_0)$.

Proposition 3.3. Under the assumption of either homoskedasticity (**A0** and **A1** or **A2**) or conditional heteroskedasticity of unknown form (**A0** and **A3**), the null hypothesis $\mathbf{R}\pi_{f_1,f_2} = \mathbf{0}$, and the maintained hypothesis of an unchanged coefficient matrix $\Pi_{f_1,f_2} = \Pi$ for all subsamples, the subsample heteroskedastic consistent Wald and sup Wald statistics converge weakly to the following limits

$$\mathcal{W}^*_{f_2}(f_1) \Rightarrow \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2-f_1)^{1/2}}\right]' \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2-f_1)^{1/2}}\right],$$

$$\mathcal{SW}^*_f(f_0) \Rightarrow \sup_{(f_1,f_2)\in\Lambda_0, f_2=f} \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2-f_1)^{1/2}}\right]' \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2-f_1)^{1/2}}\right],$$

where W_d is vector Brownian motion with covariance matrix $\mathbf{I}_d$ and d is the number of restrictions (the rank of R) under the null.

The limit theory shows that the robust test statistics remain unchanged for both scenarios – homoskedasticity and conditional heteroskedasticity. The asymptotic distributions are the same as those of the Wald process and sup Wald statistic under the assumption of homoskedasticity, given in (11) and (12).

4. SIMULATION EXPERIMENTS

There is significant evidence to suggest that Wald tests suffer from size distortions in small samples (Guilkey and Salemi, 1982; Toda and Phillips, 1993, 1994). This section therefore reports a series of simulation experiments designed to assess the finite sample characteristics of the forward, rolling and recursive evolving causality tests proposed in Section 2. The prototype model used in the simulation experiments is the bivariate VAR(1) model:

$$\text{DGP}: \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{s_t} \\ 0 & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{18}$$

where ε_{1t} and ε_{2t} are i.i.d. $N(0,1)$. Assumption **A0** requires $\left|\phi_{11}\right| < 1$ and $\left|\phi_{22}\right| < 1$. For simplicity, the causal channel from y_1 to y_2 is shut down. Parameter ϕ_{s_t} controls the strength of the causal path running from y_{2t} to y_{1t}. Under the null hypothesis of no causality, $\phi_{s_t} = 0$. Under the alternative hypothesis, causation runs from y_{2t-1} to y_{1t} from observation $\lfloor f_e T \rfloor$ to $\lfloor f_f T \rfloor$. Let s_t be a causal indicator that takes the value unity for the causal period and zero otherwise such that

$$s_t = \begin{cases} 1, & \text{if } \lfloor f_e T \rfloor \leq t \leq \lfloor f_f T \rfloor \\ 0, & \text{otherwise} \end{cases}.$$

The autoregressive coefficient ϕ_{s_t} then equals $\phi_{12} s_t$.

Initial values of the data series (y_{11} and y_{21}) are set to unity. The lag length p in the regression model is fixed at one. The rolling window test procedure uses a window length taken to be the minimum window size, f_0. The experiments are repeated 1,000 times for each parameter constellation. We report the sizes (probability of rejecting at least one true null hypothesis) and powers (probability of rejecting at least one false null hypothesis) of the three procedures for these various cases.

4.1. Multiplicity

The multiplicity problem refers to the well-known fact that the probability of making a Type I error rises with the number of hypotheses being tested. In the current context, all three testing algorithms require that the test statistic be compared with the corresponding critical value for each observation starting from $\tau_0 \equiv \lfloor T f_0 \rfloor$ to T, so that the number of hypotheses tested over the sample period equals $T - \tau_0 + 1$. To control for any size distortion arising from this recursive testing, the bootstrap method proposed in Shi *et al.* (2018a) is used.

The details of the corrective bootstrap algorithm are set out below for a bivariate VAR(1) but the procedure is easily extended to higher dimensional systems.

Step 1: Using the full sample period, estimate a bivariate VAR(1) model which imposes the null hypothesis of no Granger causality from y_2 to y_1. The resultant equation is

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \widehat{\phi}_{11} & 0 \\ \widehat{\phi}_{12} & \widehat{\phi}_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{19}$$

in which e_{1t} and e_{2t} are the estimated residuals.

Step 2: Let τ_b be the number of observations in the window over which size is to be controlled. For a sample size $\tau_0 + \tau_b - 1$, generate a bootstrap sample given by

$$\begin{bmatrix} y_{1t}^b \\ y_{2t}^b \end{bmatrix} = \begin{bmatrix} \hat{\phi}_{11} & 0 \\ \hat{\phi}_{12} & \hat{\phi}_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1}^b \\ y_{2t-1}^b \end{bmatrix} + \begin{bmatrix} e_{1t}^b \\ e_{2t}^b \end{bmatrix} \tag{20}$$

with initial values $y_{11}^b = y_{11}$ and $y_{21}^b = y_{21}$, and where the residuals e_{1t}^b and e_{2t}^b are randomly drawn with replacement from the estimated residuals in Step 1.

Step 3: Using the bootstrapped series, compute the test statistic sequences for the forward, $\left\{ \mathcal{W}_{1,t}^b \right\}_{t=\tau_0}^{\tau_0+\tau_b-1}$, rolling, $\left\{ \mathcal{W}_{t-\tau_0+1,t}^b \right\}_{t=\tau_0}^{\tau_0+\tau_b-1}$, and recursive evolving, $\left\{ \mathcal{SW}_t^b(\tau_0) \right\}_{t=\tau_0}^{\tau_0+\tau_b-1}$, algorithms respectively, and the associated maximum values of the test statistic sequences

$$\begin{aligned} \text{Forward (Thoma): } \mathcal{M}_{1,t}^b &= \max_{t\in[\tau_0,\tau_0+\tau_b-1]} \left(\mathcal{W}_{1,t}^b \right), \\ \text{Rolling (Swanson): } \mathcal{M}_{t-\tau_0+1,t}^b &= \max_{t\in[\tau_0,\tau_0+\tau_b-1]} \left(\mathcal{W}_{t-\tau_0+1,t}^b \right), \\ \text{Recursive evolving (PSY): } \mathcal{SM}_t^b(\tau_0) &= \max_{t\in[\tau_0,\tau_0+\tau_b-1]} \left(\mathcal{SW}_t^b(\tau_0) \right). \end{aligned} \tag{21}$$

Step 4: Repeat Steps 2 to 3 for $B = 499$ times.

Step 5: The critical values of the forward, rolling and recursive evolving procedures are now given by the 95% percentiles of the $\left\{ \mathcal{M}_{1,t}^b \right\}_{b=1}^B$, $\left\{ \mathcal{M}_{t-\tau_0+1,t}^b(\tau_0) \right\}_{b=1}^B$, and $\left\{ \mathcal{SM}_t^b(\tau_0) \right\}_{b=1}^B$ sequences respectively. Using these critical values, the probability of having at least one false positive detection over the sample period τ_b is 5%, thereby controlling for any potential size distortion due to multiplicity.[4]

4.2. Empirical Sizes and Powers

Table I reports the size (left panel) and power (middle and right panels) of the three procedures with different specifications of the persistence parameters $\{\phi_{11}, \phi_{22}\}$, the minimum window size f_0, and the sample size T. It is apparent in these results that the empirical size of all the procedures does not vary with the settings of the minimum window f_0 and the sample size T. Importantly, the empirical sizes of all three tests are close to nominal (5%) in most settings. This outcome shows the effectiveness of the bootstrap procedure in dealing with the multiplicity issue.[5] There is one exception. When $\{\phi_{11}, \phi_{22}\} = (0.5, 0.8)$, the empirical size for the rolling and recursive evolving procedures is close to twice the nominal size, while the size distortion for the forward algorithm is less severe. Unreported simulations suggest similar levels of size distortion for all three procedures when either one of the two autoregressive coefficients moves closer to unity. We conjecture that this size distortion is caused by the induced local-to-unity property (Phillips, 1987) of the data series in such cases. Interestingly and importantly for empirical practice, when the VAR model (1) is augmented by one lag as in Shi *et al.* (2018a), the empirical size of all the procedures under these parameter settings becomes close to the nominal size. This finding accords with known limit theory for subset testing in nonstationary VARs with augmented lags and merits further investigation in future work.

For the calculation of empirical powers in Table I, the causality (from $y_{2t} \to y_{1t}$) switches on in the middle of the sample ($f_e = 0.5$) and the relationship lasts for 20% of the sample with termination at $f_f = 0.7$. The causal strength

[4] We set $\tau_b = T$ in the simulation to control the size over the entire sample period. In the empirical application we control the size over a three-year period and this requires setting $\tau_b = \tau_0 + 35$.

[5] Unreported simulation results show that the forward, rolling and recursive evolving procedures are significantly over-sized when asymptotic critical values are used. The empirical sizes are respectively, 9%, 16% and 19% for the forward, rolling and recursive evolving procedures when $T = 100$, $(\phi_{11}, \phi_{22}) = (-0.5, 0.8)$, and $f_0 = 0.12$.

Table I. The empirical sizes and powers of the testing procedures. The parameter setting under the alternative is: $f_e = 0.5$ and $\mathcal{D} = 0.2$

	Size			Power ($\phi_{12} = 0.5$)			Power ($\phi_{12} = 0.8$)		
	Forward (Thoma)	Rolling (Swanson)	Recursive (PSY)	Forward (Thoma)	Rolling (Swanson)	Recursive (PSY)	Forward (Thoma)	Rolling (Swanson)	Recursive (PSY)
(ϕ_{11}, ϕ_{22}): $f_0 = 0.24$ and $T = 100$									
(0.5,0.5)	0.07	0.08	0.08	0.18	0.40	0.39	0.35	0.71	0.71
(0.5,0.8)	0.08	0.10	0.12	0.26	0.54	0.56	0.39	0.81	0.81
(-0.5,0.8)	0.05	0.05	0.05	0.29	0.49	0.53	0.48	0.79	0.81
(0.5,-0.8)	0.04	0.03	0.04	0.29	0.50	0.56	0.48	0.87	0.86
$\lfloor Tf_0 \rfloor$: $T = 100$ and $(\phi_{11}, \phi_{22}) = (-0.5, 0.8)$									
18	0.06	0.04	0.05	0.27	0.38	0.45	0.44	0.70	0.77
24	0.05	0.05	0.05	0.29	0.49	0.53	0.48	0.79	0.81
36	0.06	0.06	0.06	0.34	0.51	0.54	0.53	0.77	0.77
T: $f_0 = 0.24$ and $(\phi_{11}, \phi_{22}) = (-0.5, 0.8)$									
100	0.05	0.05	0.05	0.29	0.49	0.53	0.48	0.79	0.81
200	0.06	0.05	0.06	0.45	0.67	0.81	0.69	0.96	0.99
300	0.04	0.06	0.06	0.60	0.74	0.93	0.87	0.99	1.00

Note: Calculations are based on 1000 replications, with the 5% bootstrapped critical values of the statistics illustrated in Section 4.1.

Table II. The impact of causal characteristics on empirical powers of the testing procedures. The persistent parameters $(\phi_{11}, \phi_{22}) = (-0.5, 0.8)$, causality strength $\phi_{12} = 0.8$, the minimum window $f_0 = 0.24$, and $T = 100$

	Causality duration $\mathcal{D}$ $f_e = 0.5$			Causality location f_e $\mathcal{D} = 0.2$		
	$\mathcal{D} = 0.1$	$\mathcal{D} = 0.2$	$\mathcal{D} = 0.3$	$f_e = 0.3$	$f_e = 0.5$	$f_e = 0.7$
Forward	0.23	0.48	0.70	0.61	0.48	0.38
Rolling	0.35	0.79	0.95	0.81	0.79	0.79
Recursive	0.38	0.81	0.98	0.83	0.81	0.80

Note: Calculations are based on 1,000 replications, with the 5% bootstrapped critical values of the statistics illustrated in Section 4.1.

ϕ_{12} is 0.5 in the middle panel and 0.8 in the right panel. In Table II, we fix the causal strength (i.e., $\phi_{12} = 0.8$) and the persistence parameters (i.e., $(\phi_{11}, \phi_{22}) = (-0.5, 0.8)$) and investigate the impact of causal characteristics (the causal duration $\mathcal{D}$ and the location of the causal episode f_e) on the empirical powers. The results reported in Tables I and II show that, at least for the DGPs considered in this simulation exercise, the recursive evolving procedure has the highest power. The power advantage of the recursive evolving procedure is most obvious when the causal strength is moderate and the sample size is relatively large. For example, when $\phi_{12} = 0.5$, the power of the recursive evolving algorithm is 14% (19%) higher than that of the rolling window procedure when $T = 200$ ($T = 300$).

Furthermore, for all three algorithms, the power increases with the strength of the causal relationship ϕ_{12}, the sample size T, and the duration of the causal relationship $\mathcal{D}$. The powers increase slightly when the data becomes more persistent (i.e., ϕ_{22} rises form 0.5 to 0.8). There are no obvious changes in the powers when the persistent parameters change signs. The power of the forward procedure is higher when the change in causality happens early in the sample. By contrast, the location of the switch does not have an obvious impact on the performance of the rolling and recursive evolving algorithms.

The power of the forward procedure increases with the minimum window f_0. For both the rolling and recursive evolving algorithms, powers of these two procedures increase as f_0 rises from 0.18 to 0.24 but remain roughly the same or slightly lower when f_0 expands further to 0.36. This is consistent with our expectation that the additional

observations in the minimum window increase power only when they contain additional information of the causal relationship.[6] In practice the optimal choice of f_0 will depend on the strength and duration of the causal relationship and hence will be episode specific. At a minimum, f_0 needs to be large enough to ensure that there are enough observations to initiate the regression.

4.3. The Heteroskedastic-Consistent Tests

We consider two forms of conditional heteroskedasticity, namely, a GARCH(1,1) model and a stochastic volatility (SV) model, both of which are standard in the literature (Shephard, 1996; Gonçalves and Kilian, 2004; Deo, 2000; Cavaliere *et al.*, 2014). The GARCH model is

$$\begin{aligned}\varepsilon_{it} &= h_{it}^{1/2} v_{it} \text{ with } v_{it} \sim^{i.i.d} \text{N}(0,1),\\ h_{it} &= \alpha_0 + \alpha_1 \varepsilon_{it-1}^2 + \beta_1 h_{it-1}.\end{aligned}$$

As in Gonçalves and Kilian (2004), different levels of volatility persistence are considered, given by $\alpha_1 + \beta_1 = \{0.5, 0.95, 0.99\}$. The unconditional volatility of the residual is normalized to unity. The stochastic volatility model is

$$\begin{aligned}\varepsilon_{it} &= \eta_{it} \exp(h_{it}),\\ h_{it} &= \lambda h_{it-1} + 0.5 v_{it},\end{aligned}$$

where $(\eta_{it}, v_{it}) \sim^{i.i.d} \text{N}(0, diag(\sigma_v^2, 1))$. The model parameters (λ, σ_v) are set to be either $(0.951, 0.314)$ or $(0.936, 0.424)$ as in Shephard (1996) and Gonçalves and Kilian (2004).[7]

The sizes and powers of the heteroskedastic-consistent tests under the DGP of (18) and with the above two specifications of conditional heteroskedasticity are reported in Table III. The persistence parameters are $(\phi_{11}, \phi_{22}) = (-0.5, 0.8)$. The minimum window size $f_0 = 0.24$. We set the sample size $T = 100$ and $\phi_{12} = 0.8$ in the left panel. In the right panel, we consider the case with a moderate causal strength $\phi_{12} = 0.5$ and sample size $T = 200$.

The general conclusion from Table III is that the heteroskedastic-consistent tests behave very much like the those in the homoskedastic case. For the DGPs and the types of causal switching considered here, it appears that the recursive evolving procedure provides overall best performance. The results remain largely unchanged for all the different parameter settings of the GARCH and SV models.

5. CAUSALITY, THE YIELD CURVE SLOPE AND REAL ECONOMY ACTIVITY

The slope of the yield curve, defined as the difference between zero-coupon interest rates on 3-month Treasury bills and 10-year Treasury bonds, is a potentially important explanatory variable in the prediction of real economic activity (Harvey, 1988). Empirical evidence of the ability of the slope of the yield curve to forecast macroeconomic activity, including real economic growth or recessions, was provided in the 1980s and 1990s for several countries (Stock and Watson, 1989; Estrella and Hardouvelis, 1991; Estrella and Mishkin, 1998; Dotsey, 1998; Estrella and Mishkin, 1997; Plosser and Rouwenhorst, 1994). More recent work in the context of predicting real activity and recessions suggests that the slope of the yield curve still retains its predictive power (Estrella, 2005; Chauvet and Potter, 2002; Ang *et al.*, 2006; Wright, 2006; Estrella and Trubin, 2006; Rudebusch and Williams, 2009; Kauppi and Saikkonen, 2008).

While most of the early literature focused on the ability of the yield curve to predict real activity, it is reasonable to conjecture that feedback effects are present from real activity to monetary policy and therefore to the yield

[6] Recall that the duration of the causality episode is 0.2. When the minimum window f_0 exceeds that duration, the regression contains a mix of causal and non-causal observations.

[7] They are obtained by matching the SV model to real exchange rate data (Shephard, 1996).

Table III. The empirical performance of the heteroskedastic-consistent tests under the null DGP with conditional heteroskedasticity of GARCH(1,1) and stochastic volatility. The parameter settings are: $(\phi_{11}, \phi_{22}) = (-0.5, 0.8), f_0 = 0.24, f_e = 0.5$, and $D = 0.2$

	$T = 100, \phi_{12} = 0.8$			$T = 200, \phi_{12} = 0.5$		
	Forward (Thoma)	Rolling (Swanson)	Recursive (PSY)	Forward (Thoma)	Rolling (Swanson)	Recursive (PSY)
Size						
GARCH						
$\alpha_1 = 0.05, \beta_1 = 0.45$	0.06	0.05	0.05	0.06	0.05	0.06
$\alpha_1 = 0.05, \beta_1 = 0.90$	0.05	0.04	0.05	0.06	0.05	0.05
$\alpha_1 = 0.05, \beta_1 = 0.94$	0.05	0.04	0.05	0.06	0.05	0.05
Stochastic volatility						
$\lambda = 0.951, \sigma_v = 0.314$	0.05	0.06	0.06	0.03	0.05	0.05
$\lambda = 0.936, \sigma_v = 0.424$	0.04	0.06	0.06	0.04	0.05	0.05
Power						
GARCH						
$\alpha_1 = 0.05, \beta_1 = 0.45$	0.34	0.70	0.71	0.32	0.53	0.65
$\alpha_1 = 0.05, \beta_1 = 0.90$	0.34	0.69	0.71	0.33	0.54	0.64
$\alpha_1 = 0.05, \beta_1 = 0.94$	0.33	0.69	0.70	0.33	0.54	0.62
Stochastic volatility						
$\lambda = 0.951, \sigma_v = 0.314$	0.18	0.61	0.63	0.19	0.57	0.62
$\lambda = 0.936, \sigma_v = 0.424$	0.19	0.60	0.62	0.23	0.58	0.63

Note: Calculations are based on 1,000 replications, with the 5% bootstrapped critical values of the statistics illustrated in Section 4.1.

curve (Estrella and Hardouvelis, 1991; Estrella and Mishkin, 1997; Estrella, 2005; Huse, 2011). Consequently a substantial body of empirical work in this area has been conducted in terms of VAR models (Ang and Piazzesi, 2003; Evans and Marshall, 2007; Diebold *et al.*, 2006), which provides ample precedence to support the use of VAR models to establish the direction of Granger causality in these macroeconomic relationships.

In the present application a four-variable VAR model is used to test for changes in Granger causal relationships between the slope of the yield curve and the macroeconomy using United States data. The variables included are proxies for the macroeconomy: real economic activity (y_t), inflation (π_t), the monetary policy interest rate (i_t), and the yield curve spread (S_t). The decision to use a four-variable VAR model means that the curvature (or bow) of the yield curve is omitted from the model. There have been attempts to devise theoretical links between the curvature of the yield curve and the macroeconomy (Dewachter and Lyrio, 2006; Modena, 2008; Mönch, 2012); but there is little evidence to support the nature of the relationship. In view of the ambivalent evidence, it was decided not to include the curvature in the VAR.

5.1. Data

Real economic activity is proxied by the annual growth rate of (real) industrial production. Inflation is measured from the core consumer price index and calculated as log differences (multiplied by 1200). The policy rate is measured using the effective Federal funds rate. Term spread is defined as the difference between the 3-month Treasury bill rate and the 10-year government bond rate. All the data are obtained from the Federal Reserve Bank of St. Louis FRED[8] at the monthly frequency. The data start from January 1980 to March 2015 ($T = 423$).

The data are plotted in Figure 1. The left panel plots the annualized growth rate of industrial production (left axis) and inflation (right axis). The right panel plots the Federal funds rate and the slope of the yield curve. Official NBER recession periods that coincide with the sample period, namely 1980:M01-M07, 1981:M07-1982:M11, 1990:M07-1991:M03, 2001:M03-M11 and 2007:M12-2009:M06 are marked in grey. Industrial production falls sharply during recessions. After the 2008-2009 recession, the growth rate of industrial production rebounds quickly and is relatively stable until the end of the sample. Inflation fluctuates around the 2% level and shows a persistent

[8] Website: wwww.research.stlouisfed.org/fred2/.

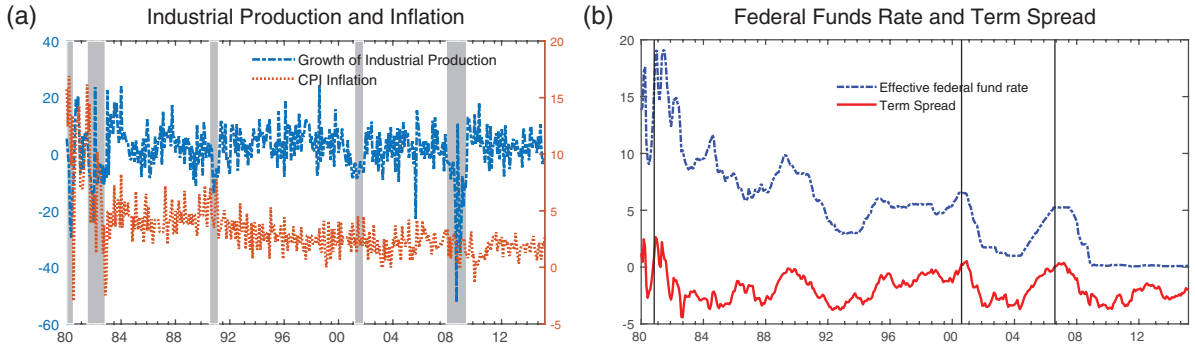

Figure 1. Time series plots of real economic activity and inflation (left panel) and the Federal funds rate and the slope of the yield curve (right panel) in the United States. Also shown are official NBER recession periods shaded in grey, namely, 1980:M01-M07, 1981:M07-1982:M11, 1990:M07-1991:M03, 2001:M03-M11 and 2007:M12-2009:M06. The vertical lines mark the generally accepted dates of the onset of an inverted yield curve given by 1980M11, 2000M08 and 2006M08 respectively

decline towards the end of the sample period, consistent with the deflationary conditions prevalent in the United States economy after the Global Financial Crisis and the movement of the Federal funds rate to the zero lower bound.

Since the yield curve is typically upward sloping, the slope factor, defined as the difference between the zero-coupon interest rates on 3-month Treasury bills and 10-year Treasury bonds, usually takes a negative value. Steeper yield curves are represented by lower values of the slope factor. If the yield curve becomes inverted then the slope factor will be positive and the dates of the onset of an inverted yield curve are shown by vertical lines. Notable instances are in 2000 (when a recession followed) and in 2006 (when the inverted yield curve was not immediately followed by a recession). A final feature of Figure 1b is the settling of the effective funds rate at zero for the latter part of the sample period after 2009Q1, the zero lower bound period of monetary policy.

In estimating the VAR and implementing tests of Granger causality, the lag order is assumed the same for all subsamples and selected using the Bayesian information criteria (BIC) for the whole sample period with a maximum potential lag length 12. The selected lag order is three. When implementing the recursive testing procedure the minimum window size is $f_0 = 0.2$, which contains 84 observations. This constant window size is also used for the rolling procedure. The critical values are obtained from bootstrapping procedure illustrated in Section 4.1 with 499 replications. The empirical size is 5% and is controlled over a three-year period.[9]

A sensitivity analysis is conducted using a minimum window size of $f_0 = 0.25$ and with empirical sizes controlled over a two year period. Additionally, we repeat the empirical analysis using quarterly output gap as a proxy for real economic activity. These results are collected in the online supplement (Shi *et al.*, 2018b).

5.2. Yield Curve Slope to Real Economic Activity

The time-varying Wald test statistics for causal effects from the slope of the yield curve to real economic activity, along with their bootstrapped critical values, are displayed in Figure 2. The three rows illustrate the sequences of test statistics obtained from the forward recursive, rolling window and recursive evolving procedures respectively, while the columns of the figure refer to the two different assumptions of the residual error term (homoskedasticity and heteroskedasticity) for the VAR. Sequences of the test statistics start from December 1986, the shaded areas are the NBER recession periods, vertical lines are the dates of the onset of an inverted yield curve and the dates of causal episodes are also shown.

[9] We estimate the VAR model under the null with the whole sample and simulate $84 + 35 = 119$ observations for each bootstrapped sample. The same critical value is applied to all tests over the sample period.

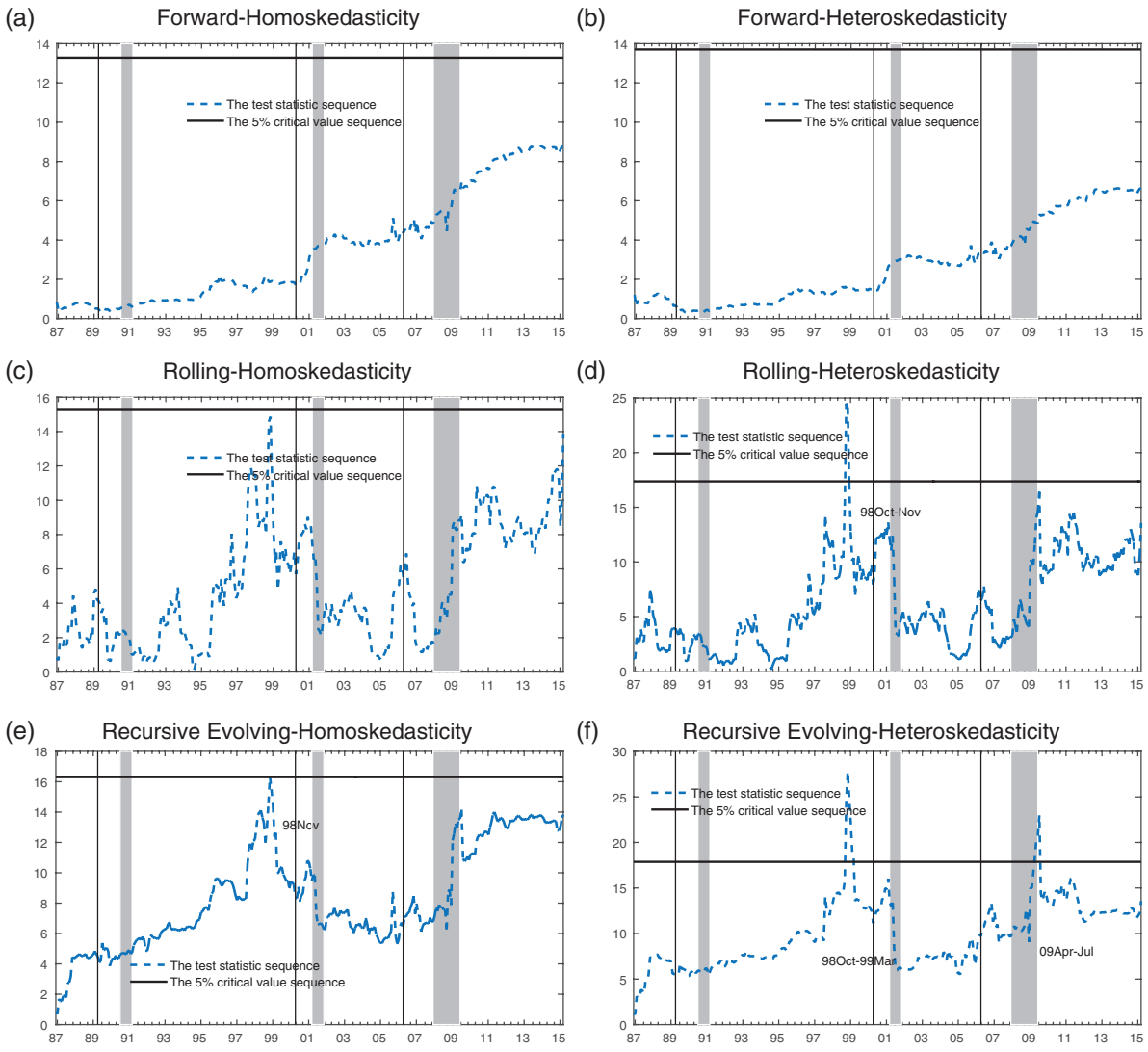

Figure 2. Tests for Granger causality running from the yield curve slope to the growth rate of industrial production. The shaded areas are the NBER recession periods, the vertical lines are the dates of the onset of an inverted yield curve and causal periods are shown in text

Panels (a) and (b) of Figure 2 indicate that the test statistics of the predictive power of the slope of the yield curve for real economic activity are always below their critical values at the end of the sample period in March 2015. Consequently, the null hypothesis of no Granger causality from the yield curve slope to the industrial production over the whole sample period cannot be rejected. This result highlights the danger of using Wald tests of Granger causality indiscriminately over the full sample period. The fact that the slope of the yield curve has little predictive power towards the end of the sample is to be expected given that this period is characterised by interest rates at the zero lower bound. The relative lack of information encoded in the slope of the yield curve during this period, therefore, is bound to have a significant distorting influence on inference based on the entire sample.

An even stronger result is provided by the forward recursive Wald test (both homoskedastic and heteroskedastic) of causality to the effect that there is no causal relationship (or change in the causal relationship) between the slope

of the yield curve and real economic activity over the entire sample period. This conclusion appears to be at odds not only with our priors but also with all existing evidence of the usefulness of the slope of the term structure in predicting real economic activity.

By contrast, the rolling and recursive evolving procedures (panels (c) to (f) of Figure 2) paint a different picture from that of an unequivocal failure to reject the null hypothesis of no predictability. Instead, a far more dynamic causal relationship between the slope of the yield curve and real economic activity is revealed. Furthermore, the difference between the homoskedastic and heteroskedastic tests is quite obvious for the rolling and recursive evolving procedures. Under the homoskedastic assumption, the rolling window procedure does not find any episodes of causality running from the yield curve slope to the growth rate of industrial production over the entire sample period; and the recursive evolving procedure detects one episode that occurred in November 1998 lasting for 1 month. By contrast, the heteroskedastic-consistent tests find stronger evidence of causality. Under the heteroskedastic assumption, both the rolling and recursive evolving procedures detect one episode in 1998, which starts in October 1998 and ends in November 1998 (March 1999) according to the rolling (recursive evolving) procedure. The duration suggested by the recursive evolving algorithm is longer than that suggested by the rolling procedure.

The heteroskedastic-consistent recursive evolving algorithm detects an additional episode in 2009, starting in April and terminating in July. This difference in these empirical results highlights the efficacy of the recursive evolving algorithm and the importance of taking the potential heteroskedasticity in the data into consideration when conducting Granger causality tests.

5.3. Real Economic Activity to the Yield Curve Slope

Figure 3 displays the time-varying Wald test statistics for causal effects running from real economic activity to the slope of the yield curve. The first interesting feature of the results reported in panels (a) and (b) of Figure 3 is that a simple Granger causality test based on the entire sample would suggest no evidence against the null hypothesis of no causality. Casual inspection of the other graphics in this figure would suggest that the unambiguous conclusion of no causal relationship from economic activity to the slope of the yield curve would be understating the balance of evidence.

The second conclusion to emerge from these results is that the empirical findings are sensitive to the assumptions made about the variance of the VAR errors. The strongest discrepancy occurs with the recursive evolving algorithm shown in Figure 3(e, f) where there is a dramatic decrease in the number of causal episodes when using the heteroskedastic-consistent version of the test.[10] Our third observation is that, just as in the case when testing causality running from the slope of the yield curve to real economic activity, the recursive evolving procedure detects more episodes of causality than the forward and rolling algorithms. The forward procedure suggests no causality, despite strong rejection of the null hypothesis by the other two methods over the period of 2012–2015. The recursive evolving method also detects an episode in 1989–1991, running from 1989:M03 to 1991:M12.

The results obtained from this empirical examination of causal links between the slope of the yield curve and real economic activity suggest that the recursive evolving algorithm is most able to detect causal changes. This result is consistent with the lessons from the simulation experiments and from the results reported in the online supplement when quarterly data is used for the empirical application.

6. CONCLUSION

The recursive evolving test procedure introduced here provides a mechanism for detecting and dating changes in Granger causal relationships. The approach uses sequences of the supremum norm of Wald statistics. Variants of the test that are robust to departures from homoskedasticity are also examined. Limit distributions of the tests are obtained and shown to have simple forms that are amenable to computation for the purpose of providing critical

[10] Consequently, we focus on the heteroskedastic-consistent tests for sensitivity analysis presented in the online supplement.

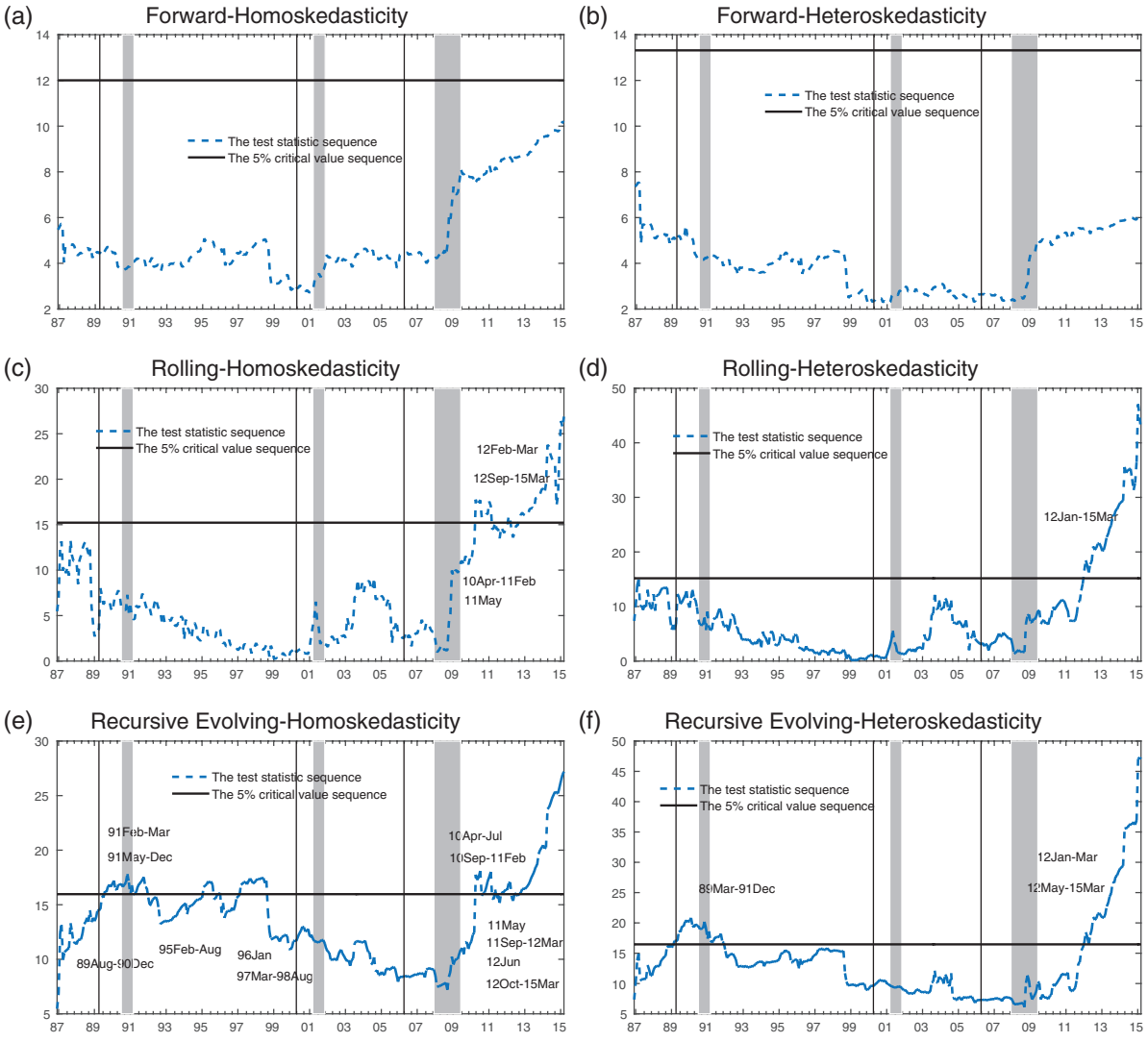

Figure 3. Tests for Granger causality running from the growth rate of industrial production to the yield curve slope. The shaded areas are the NBER recession periods, the vertical lines are the dates of the onset of an inverted yield curve and causal periods are shown in text

values. The recursive evolving procedure is compared to forward recursive and rolling window tests. The simulation findings indicate that the recursive evolving approach has superior change detection performance among the three methods in finite samples with the models considered here.

The tests are used to investigate causal relationships between the slope of the yield curve and real economic activity (proxied by either the output gap or industrial production) using United States data over 1980-2015. The empirical application builds on earlier findings in the literature concerning bidirectional causal effects between the yield slope on real economic activity. The results are consistent with much of the earlier literature. However, their most striking feature is the fact that causal relationships show considerable sensitivity to the subsample period. They reveal how endogenous detection of switches in causal effects can provide useful insights about how the nature of the macroeconomic impact of the yield curve slope can change over time. However, they also point to fragilities that can arise in the indiscriminate application of the tests over arbitrarily chosen subsamples.

ACKNOWLEDGEMENTS

We thank the Editor and three referees for helpful comments which identified an error in the earlier version of the article and led to many improvements. This research was supported under Australian Research Council's Discovery Projects funding scheme (project number DP150101716). Phillips acknowledges support from the NSF under Grant No. SES 12-85258.

SUPPORTING INFORMATION

Additional material may be found online in the supporting information tab for this article.

REFERENCES

Ang A, Piazzesi M. 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* **50**(4): 745–787.

Ang A, Piazzesi M, Wei M. 2006. What does the yield curve tell us about GDP growth? *Journal of Econometrics* **131**(1-2): 359–403.

Arora V, Shi S. 2016. Energy consumption and economic growth in the United States *Applied Economics* **48**(39): 3763–3773.

Balcilar M, Ozdemir ZA, Arslanturk Y. 2010. Economic growth and energy consumption causal nexus viewed through a bootstrap rolling window *Energy Economics* **32**(6): 1398–1410.

Benati L, Goodhart C. 2008. Investigating time-variation in the marginal predictive power of the yield spread *Journal of Economic Dynamics and Control* **32**(4): 1236–1272.

Billio M, Getmansky M, Lo AW, Pelizzon L. 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors *Journal of Financial Economics* **104**(3): 535–559.

Bodnar T, Zabolotskyy T. 2011. Estimation and inference of the vector autoregressive process under heteroscedasticity *Theory of Probability and Mathematical Statistics* **83**: 27–45.

Bollerslev T. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return *The Review of Economics and Statistics* **69**(3): 542–547.

Boswijk HP, Cavaliere G, Rahbek A, Taylor AR. 2016. Inference on co-integration parameters in heteroskedastic vector autoregressions *Journal of Econometrics* **192**(1): 64–85.

Brown BM. 1971. Martingale central limit theorems *The Annals of Mathematical Statistics* **42**(1): 59–66.

Cavaliere G, Rahbek A, Robert Taylor A. 2014. Bootstrap determination of the co-integration rank in heteroskedastic VAR models *Econometric Reviews* **33**(5-6): 606–650.

Chauvet M, Potter S. 2002. Predicting a recession: evidence from the yield curve in the presence of structural breaks *Economics Letters* **77**(2): 245–253.

Chauvet M, Potter S. 2005. Forecasting recessions using the yield curve *Journal of Forecasting* **24**(2): 77–103.

Chauvet M, Senyuz Z. 2016. A dynamic factor model of the yield curve components as a predictor of the economy *International Journal of Forecasting* **32**(2): 324–343.

Chen H, Cummins JD, Viswanathan KS, Weiss MA. 2014. Systemic risk and the interconnectedness between banks and insurers: An econometric analysis *Journal of Risk and Insurance* **81**(3): 623–652.

Chinn M, Kucko K. 2015. The predictive power of the yield curve across countries and time *International Finance* **18**(2): 129–156.

Davidson J. 1994. *Stochastic Limit theory: An Introduction for Econometricians*. Oxford: Oxford University Press.

Deo RS. 2000. Spectral tests of the martingale hypothesis under conditional heteroscedasticity *Journal of Econometrics* **99**(2): 291–315.

Dewachter H, Lyrio M. 2006. Macro factors and the term structure of interest rates *Journal of Money, Credit, and Banking* **38**(1): 119–140.

Diebold FX, Rudebusch GD, Aruoba SB. 2006. The macroeconomy and the yield curve: a dynamic latent factor approach *Journal of econometrics* **131**(1-2): 309–338.

Dotsey M. 1998. The predictive content of the interest rate term spread for future economic growth *Economic Quarterly (Sum)*: 31–51.

Elder J. 2004. Another perspective on the effects of inflation uncertainty *Journal of Money, Credit, and Banking* **36**(5): 911–928.

Engle RF, Bollerslev T. 1986. Modelling the persistence of conditional variances *Econometric Reviews* **5**(1): 1–50.

Estrella A. 2005. Why does the yield curve predict output and inflation? *The Economic Journal* **115**(505): 722–744.

Estrella A, Hardouvelis GA. 1991. The term structure as a predictor of real economic activity *The Journal of Finance* **46**(2): 555–576.

Estrella A, Mishkin FS. 1997. The predictive power of the term structure of interest rates in Europe and the United States: implications for the European Central Bank *European Economic Review* **41**(7): 1375–1401.

Estrella A, Mishkin FS. 1998. Predicting US recessions: financial variables as leading indicators *Review of Economics and Statistics* **80**(1): 45–61.
Estrella A, Rodrigues AP, Schich S. 2003. How stable is the predictive power of the yield curve? Evidence from Germany and the United States *Review of Economics and Statistics* **85**(3): 629–644.
Estrella A, Trubin MR. 2006. The yield curve as a leading indicator: some practical issues *Current Issues in Economics and Finance* **12**(5): 1.
Evans CL, Marshall DA. 2007. Economic determinants of the nominal treasury yield curve *Journal of Monetary Economics* **54**(7): 1986–2003.
Giacomini R, Rossi B. 2006. How stable is the forecasting performance of the yield curve for output growth? *Oxford Bulletin of Economics and Statistics* **68**(s1): 783–795.
Gonçalves S, Kilian L. 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form *Journal of Econometrics* **123**(1): 89–120.
Granger C, Newbold P. 1986. *Forecasting Economic Time Series: Economic Theory, Econometrics and Mathematical Economics*. San Diego, CA: Academic Press.
Granger CW. 1969. Investigating causal relations by econometric models and cross-spectral methods *Econometrica*: 424–438.
Granger CW. 1988. Some recent developments in a concept of causality *Journal of Econometrics* **39**(1-2): 199–211.
Guilkey DK, Salemi MK. 1982. Small sample properties of three tests for granger-causal ordering in a bivariate stochastic system *The Review of Economics and Statistics*: 668–680.
Hamilton JD. 2011. Calling recessions in real time *International Journal of Forecasting* **27**(4): 1006–1026.
Hannan E, Heyde C. 1972. On limit theorems for quadratic functions of discrete time series *The Annals of Mathematical Statistics* **43**(6): 2058–2066.
Harvey CR. 1988. The real term structure and consumption growth *Journal of Financial Economics* **22**(2): 305–333.
Haubrich JG, Dombrosky AM. 1996. Predicting real growth using the yield curve *Economic Review-Federal Reserve Bank of Cleveland* **32**(1): 26.
Huse C. 2011. Term structure modelling with observable state variables *Journal of Banking & Finance* **35**(12): 3240–3252.
Kauppi H, Saikkonen P. 2008. Predicting US recessions with dynamic binary response models *The Review of Economics and Statistics* **90**(4): 777–791.
Leybourne S, Kim T-H, Taylor AR. 2007. Detecting multiple changes in persistence *Studies in Nonlinear Dynamics & Econometrics* **11**(3): Article 2.
Modena M. 2008. An Empirical Analysis of the Curvature Factor of the Term Structure of Interest Rates. Technical Report.
Mönch E. 2012. Term structure surprises: the predictive content of curvature, level, and slope *Journal of Applied Econometrics* **27**(4): 574–602.
Nelson DB. 1991. Conditional heteroskedasticity in asset returns: a new approach: 347–370.
Newbold P, Hotopp SM. 1986. Testing causality using efficiently parametrized vector ARMA models *Applied Mathematics and Computation* **20**(3–4): 329–348.
Patilea V, Rassi H. 2012. Adaptive estimation of vector autoregressive models with time-varying variance: application to testing linear causality in mean *Journal of Statistical Planning and Inference* **142**(11): 2891–2912.
Phillips PCB. 1987. Towards a unified asymptotic theory for autoregression *Biometrika* **74**(3): 535–547.
Phillips PCB, Shi S, Yu J. 2015a. Testing for multiple bubbles: historical episodes of exuberance and collapse in the S&P 500 *International Economic Review* **56**(4): 1043–1078.
Phillips PCB, Shi S, Yu J. 2015b. Testing for multiple bubbles: limit theory of real-time detectors *International Economic Review* **56**(4): 1079–1134.
Phillips PCB, Solo V. 1992. Asymptotics for linear processes *The Annals of Statistics* **20**(2): 971–1001.
Phillips PCB, Wu Y, Yu J. 2011. Explosive behavior in the 1990s Nasdaq: when did exuberance escalate asset values? *International Economic Review* **52**(1): 201–226.
Phillips PCB, Yu J. 2011. Dating the timeline of financial bubbles during the subprime crisis *Quantitative Economics* **2**(3): 455–491.
Plosser CI, Rouwenhorst KG. 1994. International term structures and real economic growth *Journal of Monetary Economics* **33**(1): 133–155.
Psaradakis Z, Ravn MO, Sola M. 2005. Markov switching causality and the money–output relationship *Journal of Applied Econometrics* **20**(5): 665–683.
Rossi B. 2005. Optimal tests for nested model selection with underlying parameter instability *Econometric Theory* **21**(5): 962–990.
Rossi B. 2013. Advances in forecasting under instability. In *Handbook of Economic Forecasting*, Vol. 2: Elsevier; 1203–1324.
Rudebusch GD, Williams JC. 2009. Forecasting recessions: the puzzle of the enduring power of the yield curve *Journal of Business & Economic Statistics* **27**(4): 492–503.
Shephard N. 1996. Statistical aspects of ARCH and stochastic volatility *Monographs on Statistics and Applied Probability* **65**: 1–68.

Shi S, Hurn S, Phillips PCB. 2018a. Causal Change Detection in Possibly Integrated Systems: Revisiting the Money–Income Relationship. Technical report, Available at SSRN: https://ssrn.com/abstract=3237213.
Shi S, Phillips PCB, Hurn S. 2018b. Supplement to the Paper: Change Detection and the Causal Impact of the Yield Curve. https://doi.org/10.1111/jtsa.12427.
Stern DI. 2000. A multivariate cointegration analysis of the role of energy in the US macroeconomy *Energy Economics* **22**(2): 267–283.
Stock JH, Watson MW. 1989. Interpreting the evidence on money–income causality *Journal of Econometrics* **40**(1): 161–181.
Stock JH, Watson MW. 1999. Business cycle fluctuations in US macroeconomic time series *Handbook of Macroeconomics* **1**: 3–64.
Swanson NR. 1998. Money and output viewed through a rolling window *Journal of Monetary Economics* **41**(3): 455–474.
Thoma MA. 1994. Subsample instability and asymmetries in money–income causality *Journal of Econometrics* **64**(1–2): 279–306.
Toda HY, Phillips PCB. 1993. Vector autoregressions and causality *Econometrica* **61**(6): 1367–1393.
Toda HY, Phillips PCB. 1994. Vector autoregression and causality: a theoretical overview and simulation study *Econometric Reviews* **13**(2): 259–285.
Wright J. 2006. The Yield Curve and Predicting Recessions. Technical report. Board of Governors of the Federal Reserve System.
Yogo M. 2004. Estimating the elasticity of intertemporal substitution when instruments are weak *Review of Economics and Statistics* **86**(3): 797–810.