

Schreiber, Sven; Breitung, Jörg

## Conference Paper

# Tests Of Non-Causality In A Frequency Band

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Time Series Econometrics, No. A23-V2

### Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Schreiber, Sven; Breitung, Jörg (2015) : Tests Of Non-Causality In A Frequency Band, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Time Series Econometrics, No. A23-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/113111>

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# TESTS OF NON-CAUSALITY IN A FREQUENCY BAND

<ANONYMIZED>

ABSTRACT. We extend the frequency-specific Granger-causality test of Breitung and Candelon (2006) to a more general null hypothesis that allows non-causality at unknown frequencies within an interval, instead of having to prespecify a single frequency. This setup corresponds better to most hypotheses that are typically analyzed in applied research and is easy to implement. We also discuss a test approach that departs from strict non-causality, given the impossibility of (non-trivial) non-causality over a continuum of frequencies. In an empirical application dealing with the dynamics of US temperatures and CO<sub>2</sub> emissions we find that emissions cause temperature changes only at very low frequencies with more than 30 years of oscillation.

Keywords: Granger causality, frequency domain, filter gain

JEL codes: C32 (multivariate times series), Q54 (global warming)

## 1. INTRODUCTION

The notion of empirical causality as predictive ability has a long history in science and was formalized by Granger (1969). It became very popular among practitioners due to the simplicity of its implementation in linear dynamic models, where a test for non-Granger-causality is equivalent to a joint exclusion test of lagged terms of the candidate variable. A generalization of this concept was later introduced by Geweke (1982), who noted that causal effects can vary between the different cycles of time series, where each cyclical component corresponds to a certain frequency of oscillation. However, the practical application of the test that Geweke's causality measure is zero at a certain frequency appeared to be quite difficult until Breitung and Candelon (2006, henceforth BC) pointed out that

---

*Date:* preliminary version March 2015 .

in the framework of a linear VAR the null hypothesis is equivalent to two linear restrictions that can be tested for example with a standard Wald test.

A drawback of the BC test is that the test is formulated in terms of a single frequency point that has to be specified a priori. In practice, however, many test statistics are calculated for a range of frequencies to gain insights into the relationship between the variables, although it is well-known that the classical test approach does not allow a rigorous joint interpretation of these set of statistics. Furthermore, the underlying (economic) theory usually does not provide a hypothesis for only a single frequency. For example, consider the following implication of the expectations hypothesis of the term structure as noted by Shiller (1979, p.1190), pointing out theoretical Granger non-causality of short-term interest rates for a range of higher frequencies:<sup>1</sup> "... excess [short-term] volatility implies a kind of forecastability for long rates."

In order to better reflect the hypotheses that come naturally from underlying theory we extend the frequency-specific test for Granger non-causality by formulating a generalized null hypothesis for a frequency interval. As a side effect of this work we also present a different representation of the original restrictions by BC, which turns out to be helpful for our present purpose.

Related literature. The BC test was used to analyze Granger-causal effects of money on inflation in a series of papers by Assenmacher-Wesche and Gerlach (Assenmacher-Wesche and Gerlach, 2008a, 2007, 2008b). They noted some moderate size distortions and applied the bootstrap as a small-sample correction, but given the lack of other tools at the time, they were forced to use the point-wise tests even though they analyzed frequency bands. Another use for output forecasting was shown by Lemmens, Croux, and

---

<sup>1</sup>His precise definition of volatility is "variance of short-term holding yields on long-term bonds", which is related to the short-term "percentage change in the long-term interest rate" (p. 1191) and thus to high-frequency fluctuations of long-term rates, but in a nonlinear way.

Dekimpe (2008), who concluded that the BC approach was the most efficient test among the ones considered. A concept which is closely related to frequency-specific Granger causality is “partial directed coherence”, see Baccalá and Sameshima (2001), where inference is also carried out point-wise.

## 2. SETUP AND NOTATION

Consider a standard vector autoregression (VAR) of order  $p$  in the two variables  $x_t$  and  $y_t$ :

$$(2.1) \quad A(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \mathbf{c} + \begin{bmatrix} u_{x,t} \\ u_{y,t} \end{bmatrix}, \quad t = p + 1, \dots, T$$

where  $\mathbf{u}_t = (u_{x,t}, u_{y,t})'$  are normally distributed white noise innovations with contemporaneous covariance matrix  $\Psi$ . We initially assume the polynomial  $A(L)$  to be stable with roots outside the unit circle such that both  $x_t$  and  $y_t$  will be stationary. The extended case of unit roots will also be discussed below. Further deterministic terms such as linear trends or seasonal dummies could be easily added.

Let  $y_t$  be the potential target variable that is Granger-caused by  $x_t$  under the alternative. Using some obvious notation we can write the second equation of the system as follows:

$$(2.2) \quad y_t = c_y + \sum_{j=1}^p \alpha_j y_{t-j} + \sum_{k=1}^p \beta_k x_{t-k} + u_{y,t}$$

BC (2006) showed that the hypothesis of no Granger causality at frequency  $\omega$ , or  $M_{y \rightarrow x}(\omega) = 0$ , can be imposed as two linear restrictions  $R(\omega)\beta = 0$ , where  $\beta = (\beta_1, \dots, \beta_p)'$  and

$$R(\omega) = \begin{bmatrix} \cos(\omega) & \cos(2\omega) & \cdots & \cos(p\omega) \\ \sin(\omega) & \sin(2\omega) & \cdots & \sin(p\omega) \end{bmatrix}$$

For a lag order of  $p = 1$  or  $p = 2$  there is only a trivial solution to this restriction, namely that  $\beta_1 = \beta_2 = 0$ . In these two cases therefore the hypothesis of Granger non-causality at a certain frequency  $\omega \in (0, \pi)$  automatically implies the standard case of no Granger causality at any frequency. We therefore require a higher lag order,  $p > 2$ , in order to make a frequency-specific analysis interesting.

In practice the system (2.1) would often be augmented with further variables  $\mathbf{z}_t$  to avoid spurious findings due to omitted variables, see BC for a discussion. Such an addition would lead to obvious augmentations of (2.2) with lagged (or in the case of exogenous variables, possibly contemporaneous and lagged) values  $\mathbf{z}_t$ , but would not affect our results in any other way. Therefore we focus on the bivariate case for ease of exposition.

### 3. AN ALTERNATIVE REPRESENTATION

For our purposes it is useful to represent the null hypothesis of no Granger causality at frequency  $\omega$  in a more convenient manner. Our representation is based on a decomposition that is similar to the well-known BN decomposition proposed by Beveridge and Nelson (1981) for the frequency  $\omega = 0$ . Let us first consider the test at frequency  $\omega = 0$  (long-run causality). In this case the null hypothesis boils down to  $\sum_{j=1}^p \beta_j = 0$ . Following Dickey and Fuller (1979) we decompose the polynomial  $\beta(L)$  as

$$\beta(L) = b_1^0 + (1 - L)\gamma^0(L)$$

where  $b_1^0 = \sum_{j=1}^p \beta_j$ ,  $\gamma^0(L) = \gamma_0^0 + \gamma_1^0 L + \dots + \gamma_{p-2}^0 L^{p-2}$  and  $\gamma_j^0 = -\sum_{i=j+2}^p \beta_i$ . Note that this decomposition is also used to obtain the Beveridge-Nelson decomposition. Accordingly, a test for causality at frequency  $\omega = 0$  is equivalent to testing  $b_1^0 = 0$  in the regression

$$(3.1) \quad y_t = c_y + \sum_{j=1}^p \alpha_j y_{t-j} + b_1^0 x_{t-1} + \sum_{k=1}^{p-1} \gamma_{k-1}^0 \Delta x_{t-k} + u_{y,t}.$$

In the following a similar approach is suggested for testing causality at frequencies  $0 < \omega < \pi$ . To this end we first present a suitable decomposition of the lag polynomial.

**Lemma 1.** *Let  $\beta(L) = 1 + \beta_1 L + \dots + \beta_{p-1} L^{p-1}$  with  $p \geq 3$ . Then for  $0 < \omega < \pi$  there exists a representation of the form*

$$(3.2) \quad \beta(L) = b_1^\omega + b_2^\omega L + \gamma^\omega(L) \nabla_\omega(L)$$

where  $\nabla_\omega(L) = 1 - 2 \cos(\omega)L + L^2$  and  $\gamma^\omega(L) = \gamma_0^\omega + \gamma_1^\omega L + \dots + \gamma_{p-3}^\omega L^{p-3}$ . The gain function  $|\beta(e^{i\omega})|^2$  is zero at frequency  $\omega$  if and only if  $b_1^\omega = 0$  and  $b_2^\omega = 0$ .

*Proof.* Comparing the coefficients at different lags yields the system of equations

$$\begin{pmatrix} 1 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 \cos(\omega) & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 \cos(\omega) & 1 & \dots & 0 & 0 & 0 \\ \vdots & & & \ddots & & & & \vdots & \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & -2 \cos(\omega) & 1 \end{pmatrix} \begin{pmatrix} b_1^\omega \\ b_2^\omega \\ \gamma_0^\omega \\ \vdots \\ \gamma_{p-3}^\omega \end{pmatrix}$$

For  $0 < \omega < \pi$  this linear system can be solved to obtain  $b_1^\omega$ ,  $b_2^\omega$  and  $\gamma_0^\omega, \dots, \gamma_{p-3}^\omega$ . Since  $\nabla_\omega(e^{i\omega}) = \nabla_\omega(e^{-i\omega}) = 0$ , the gain function results as

$$\begin{aligned} |\beta(e^{i\omega})|^2 &= \beta(e^{i\omega}) \beta(e^{-i\omega}) \\ &= (b_1^\omega)^2 + 2b_1^\omega b_2^\omega \cos(\omega) + (b_2^\omega)^2. \end{aligned}$$

It follows that  $|\beta(e^{i\omega})|^2 = 0$  if and only if  $b_1^\omega = b_2^\omega = 0$ .  $\square$

Accordingly, (2.2) can be re-written as

$$(3.3) \quad y_t = c_y + \sum_{j=1}^p \alpha_j y_{t-j} + b_1^\omega x_{t-1} + b_2^\omega x_{t-2} + \sum_{k=1}^{p-2} \gamma_{k-1}^\omega \nabla_\omega(L) x_{t-k} + u_{y,t},$$

for  $0 < \omega < \pi$ . Note that this representation requires a lag order of  $p \geq 3$ . From Lemma 1 it follows that the transfer function possesses a zero at frequency  $\omega$  if and only if  $b_1^\omega = 0$  and  $b_2^\omega = 0$ . Accordingly, the hypothesis that  $x_t$  is a Granger cause of  $y_t$  is equivalent to the joint null hypothesis  $H_0 : b_1^\omega = 0$  and  $b_2^\omega = 0$  in the representation (3.3).

The corresponding representation for frequency  $\omega = \pi$  is given by  $\nabla_\pi = 1 + L$  and causality at this frequency can be tested by replacing the difference operator  $\Delta$  in (3.1) with  $\nabla_\pi$ .

#### 4. TESTING WHEN THE FREQUENCY IS UNKNOWN

**4.1. Theory.** In many applications it is reasonable to assume that the frequency for which  $x_t$  is not a Granger cause for  $y_t$  is unknown but it is assumed that the frequency lies within some prespecified interval  $\omega \in \Omega_0 = [\omega_\ell, \omega_u]$ . Thus the relevant null hypothesis is

$$H_0^u : \text{There exists a frequency } \omega \in [\omega_\ell, \omega_u] \text{ such that } |\beta(e^{i\omega})|^2 = 0.$$

For testing such a null hypothesis it is natural to employ the minimum of the sequence of (Wald/LR/LM) test statistics for all test statistics associated with the grid of frequencies

$$\omega \in \Omega_0^\delta = \{\omega_\ell, \omega_\ell + \delta, \omega_\ell + 2\delta, \dots, \omega_u\}$$

where  $\delta$  denotes the frequency increment, say  $\pi/T$ . Let  $\lambda_T^\omega$  denote the causality test statistic at frequency  $\omega$ . The next theorem shows that asymptotically the significance level of the test can be controlled by using the usual critical value of the  $\chi_2^2$ -distribution of the test for a known frequency.

**Theorem 1.** *Let  $\lambda_T^\omega$  denote the Wald/LM/LR test statistic for Granger causality at frequency  $\omega$  and  $\lambda_T^* = \inf\{\lambda_T^\omega | \omega \in \Omega_0^\delta\}$  with  $\delta = \pi/T$ . The  $(1 - \alpha)$  quantile of the  $\chi^2$  distribution with 2 degrees of freedom is denoted by  $\chi_{2,\alpha}^2$ . (i) Under the null hypothesis that there exists at least one frequency  $\omega^* \in \Omega_0$ , with  $\omega_l > 0$ ,*

$\omega_u < \pi$ , such that  $|\beta(e^{i\omega^*})|^2 = 0$ , then

$$\lim_{T \rightarrow \infty} P(\lambda_T^* > \chi_{2,\alpha}^2) \leq \alpha.$$

(ii) If under the null hypothesis there exists a single frequency with  $|\beta(e^{i\omega^*})|^2 = 0$  but  $|\beta(e^{i\omega})|^2 > 0$  for  $\omega \neq \omega^*$  in the interval  $\omega \in \Omega_0$ , then as  $T \rightarrow \infty$

$$\lim_{T \rightarrow \infty} P(\lambda_T^* > \chi_{2,\alpha}^2) = \alpha.$$

*Proof.* (i) As shown by BC the statistic  $\lambda_T^{\omega^*}$  for the simple test at the frequency  $\omega^*$  has a  $\chi^2$  limiting distribution with 2 degrees of freedom. Accordingly we have  $\lim_{T \rightarrow \infty} P(\lambda_T^{\omega^*} > \chi_{2,\alpha}^2) = \alpha$ . Since  $\lim_{\delta \rightarrow 0} \omega^* \in \Omega_0^\delta$ , as  $T \rightarrow \infty$  we have

$$\lambda_T^* = \inf\{\lambda_T^{\omega_\ell}, \lambda_T^{\omega_\ell + \delta}, \lambda_T^{\omega_\ell + 2\delta}, \dots, \lambda_T^{\omega_u}\} \leq \lambda_T^{\omega^*}$$

and, therefore,  $\lim_{T \rightarrow \infty} P(\lambda_T^* > \chi_{2,\alpha}^2) \leq \alpha$ .

(ii) If  $|\beta(e^{i\omega})|^2 > 0$  for  $\omega \neq \omega^*$  in the interval  $\omega \in \Omega_0$  then for all frequencies in this interval except  $\omega^*$  we have (cf. BC)  $\lambda_T^\omega = |O_p(T)|$  and, therefore,  $\lambda_T^* = \lambda_T^{\omega^*}$ . It follows that the statistic  $\lambda_T^*$  is asymptotically equivalent to a test statistic computed from the (correct) frequency  $\omega^*$ .  $\square$

It follows that the size of the test is controlled by using the minimal test statistic in the interval  $[\omega_\ell, \omega_u]$  and applying the test at the associated frequency as if this frequency were known. If there exists a single non-causal frequency in the interval  $[\omega_\ell, \omega_u]$ , then the test controls the size asymptotically, whereas the test becomes conservative whenever the interval contains more than one non-causal frequency.

The following corollary clarifies the extension to the special frequencies 0 and  $\pi$  where only a single restriction is tested and hence the limiting distribution has only one degree of freedom.



**Corollary 1.** *If  $0 \in \Omega_0$  or  $\pi \in \Omega_0$ , let  $\lambda_T^0$  and  $\lambda_T^\pi$  be the corresponding test statistics of the restriction  $b_1^0 = 0$  or  $b_1^\pi = 0$  in (3.3), and  $\chi_{1,\alpha}^2$  denotes the  $(1 - \alpha)$  quantile for 1 degree of freedom. Consider the three null hypotheses that  $|\beta(e^{i\omega^*})|^2 = 0$  for some  $\omega^*$  in (i)  $[0, \omega_u]$ ,  $\omega_u < \pi$ , (ii)  $[\omega_l, \pi]$ ,  $\omega_l > 0$ , and (iii)  $[0, \pi]$ , with relevant test statistics given by:*

$$\begin{aligned} \text{(i)} \quad \lambda_T^{0*} &= \inf \left\{ \frac{\chi_{2,\alpha}^2}{\chi_{1,\alpha}^2} \lambda_T^0, \lambda_T^\omega | \omega \in (\Omega_0^\delta \setminus 0) \right\}, \\ \text{(ii)} \quad \lambda_T^{\pi*} &= \inf \left\{ \frac{\chi_{2,\alpha}^2}{\chi_{1,\alpha}^2} \lambda_T^\pi, \lambda_T^\omega | \omega \in (\Omega_0^\delta \setminus \pi) \right\} \text{ and} \\ \text{(iii)} \quad \lambda_T^{0\pi*} &= \inf \left\{ \frac{\chi_{2,\alpha}^2}{\chi_{1,\alpha}^2} \lambda_T^0, \frac{\chi_{2,\alpha}^2}{\chi_{1,\alpha}^2} \lambda_T^\pi, \lambda_T^\omega | \omega \in (\Omega_0^\delta \setminus \{0, \pi\}) \right\}. \end{aligned}$$

*Under their respective null hypotheses these test statistics have the properties that were given for  $\lambda_T^*$  in theorem 1.*

*Proof.* If  $x_t$  is not Granger causal at frequencies 0 (or  $\pi$ ),  $\lambda_T^0$  (or  $\lambda_T^\pi$ ) has a limiting  $\chi^2$  distribution with 1 degree of freedom. Therefore  $\lim_{T \rightarrow \infty} P\left(\frac{\chi_{2,\alpha}^2}{\chi_{1,\alpha}^2} \lambda_T^0 > \chi_{1,\alpha}^2\right) | \omega^*=0 = \lim_{T \rightarrow \infty} P\left(\frac{\chi_{2,\alpha}^2}{\chi_{1,\alpha}^2} \lambda_T^\pi > \chi_{1,\alpha}^2\right) | \omega^*=\pi = \lim_{T \rightarrow \infty} P(\lambda_T^{\omega^*} > \chi_{2,\alpha}^2) | \omega^* \in (0, \pi)$ . Minimizing over the extended grid with the suitably scaled test statistics for frequencies 0 and  $\pi$  therefore also controls the size of the test.  $\square$

*Remark 1.* Note that the third hypothesis in corollary 1 involves all possible frequencies but is much less restrictive than the traditional test of Granger non-causality. The traditional test requires non-causality at all frequencies, while the third hypothesis above merely leaves the frequency of non-causality unspecified.

**4.2. Simulation study.** We assess the empirical characteristics of the test with a standard simulation study, where we use the following true models for  $t = 0 \dots T$ . For  $x$  we specify a univariate exogenous AR(1) process,

$$x_t = a_{1,x} x_{t-1} + u_x,$$

with two degrees of persistence,  $a_{1,x} \in \{0, 0.8\}$ . The pre-sample starting value is set to zero,  $x_{-1} = 0$ . The target variable  $y$  is not caused by  $x$  at a

TABLE 1. Specified frequency bands for the simulation of size and power

True non-causal frequency $w$	Bands for $H_0^u$ (size)	Bands for $H_0^u$ (power)
0	$[0, 0.5], [0, 1]$	$[0.5, 1.5], [1.5, 2.5]$
0.2	$[0.1, 0.5], [0, 1]$	$[0.5, 1.5], [1.5, 2.5]$
1.5	$[1, 2], [1.5, \pi]$	$[0, 1], [2, 3]$

frequency  $w$  for which we consider the values  $w \in \{0, 0.2, 1.5\}$ . We also allow persistence in the  $y$ -process and thus we have:

$$y_t = a_{1,y}y_{t-1} + \gamma_0x_{t-1} - \gamma_0 2 \cos(w)x_{t-2} + \gamma_0x_{t-3} + u_y,$$

where again  $a_{1,y} \in \{0, 0.8\}$  and zero starting values,  $y_{-1} = 0 = x_{-1} = x_{-2} = x_{-3}$ . We let the overall impact coefficient  $\gamma_0$  vary as  $\gamma_0 \in \{-1, 0.5, 1\}$ . Finally, the innovations are uncorrelated Gaussian white noise with normalized variance,  $u \sim NID(0, I_2)$ .

In table 1 we report the frequency bands that are considered as null hypotheses for the simulation of the size of the test. For the grid of tested frequencies we use a distance of  $\delta \approx 0.1$ , plus the special cases 0 and  $\pi$ , such that the grid contains the true non-causal frequencies. For the analysis of the power of the test we confine ourselves to situations where  $x$  is still non-causal at some frequency, but this frequency now lies outside the null band. In addition one could specify a true DGP without any non-causality. In the third column of table 1 we report the analyzed frequency bands for the simulated power of the test.

In table 2 we have collected the simulation results. It is apparent that the test is quite conservative and the rejection frequencies under the null never reach its nominal significance level for the considered sample sizes of  $T = 200$  and  $T = 500$ . This seems to happen especially when the frequency zero is included in the null band while the true frequency is at 0.2. In this scenario the null hypothesis is rejected in less than 1% of the simulations for a nominal level of 5%. Nevertheless, the power of the test is still quite

satisfactory. The only notable problems occur when the true impact of  $x$  is quite low ( $\gamma_0 = 0.5$ ) and the data are noisy ( $a_{1,x} = a_{1,y} = 0$ ); there the power may drop below 20%, but given the conservatism that we found the test is still far from being biased even in these extreme circumstances.

## 5. UNIT ROOTS AND COINTEGRATION

So far we have assumed a stable VAR system with all roots outside the unit circle. Considering the possibility of unit roots at frequency zero (real and positive unit roots), the analysis extends naturally to the case with  $I(1)$  variables that are not cointegrated, by differencing the corresponding variables and proceeding as before.

If the variables are cointegrated the analysis can be slightly adapted as already described in BC. The tests at non-zero frequencies are not (asymptotically) affected by the unit roots in a cointegrated system and can be performed in a standard fashion.

However, it may be worth noting that by definition of cointegration, at least one of the variables must be long-run causal at frequency zero.<sup>2</sup> Let the normalized cointegration coefficient be  $\kappa \neq 0$  such that  $y_t + \kappa x_t = e_t \sim I(0)$ . Then the null hypothesis of non-causality of  $x$  for  $y$  at frequency zero does not contradict the cointegration assumption only under the implicit hypothesis that  $y$  is long-run causal for  $x$ , i.e. that  $a_2 \neq 0$  in

$$\Delta x_t = c_x + a_1 x_{t-1} + a_2 y_{t-1} + \sum_{j=1}^{p-1} d_{x,j} \Delta x_{t-j} + \sum_{k=1}^{p-1} d_{y,k} \Delta y_{t-k} + u_{x,t},$$

because under cointegration it holds that either  $\kappa = a_1/a_2$  if  $x$  is caused by  $y$  at frequency zero, or else  $a_1 = a_2 = 0$ . In the latter case the error correction term would have to enter the equation for  $y$ , such that in (3.3) with

---

<sup>2</sup>As we saw before, the converse is not true in general. It makes sense to talk about causality at frequency zero even for  $I(0)$  variables.

TABLE 2. Simulation results

$T = 200$						
Frequency band under $H_0$	$a_{1,x} = a_{1,y} = 0$			$a_{1,x} = a_{1,y} = 0.8$		
	$\gamma_0 =$	$\gamma_0 =$	$\gamma_0 =$	$\gamma_0 =$	$\gamma_0 =$	$\gamma_0 =$
	-1	0.5	1	-1	0.5	1
True frequency $w = 0$						
[0, 0.5]	.018	.012	.015	.018	.018	.017
[0, 1]	.015	.014	.015	.021	.017	.019
[0.5, 1.5]	.428	.134	.429	.992	.642	.990
[1.5, 2.5]	1	1	1	1	1	1
True frequency $w = 0.2$						
[0.1, 0.5]	.023	.032	.027	.022	.028	.022
[0, 1]	.009	.011	.007	.006	.006	.005
[0.5, 1.5]	.318	.115	.317	.953	.483	.957
[1.5, 2.5]	1	1	1	1	1	1
True frequency $w = 1.5$						
[1, 2]	.021	.020	.020	.022	.014	.019
[1.5, $\pi$ ]	.042	.035	.039	.036	.036	.043
[0, 1]	1	.994	1	1	1	1
[2, 3]	1	.999	1	1	.912	1
$T = 500$						
Frequency band under $H_0$	$a_{1,x} = a_{1,y} = 0$			$a_{1,x} = a_{1,y} = 0.8$		
	$\gamma_0 =$	$\gamma_0 =$	$\gamma_0 =$	$\gamma_0 =$	$\gamma_0 =$	$\gamma_0 =$
	-1	0.5	1	-1	0.5	1
True frequency $w = 0$						
[0, 0.5]	.012	.015	.014	.015	.016	.015
[0, 1]	.017	.013	.015	.017	.017	.017
[0.5, 1.5]	.837	.296	.846	1	.976	1
[1.5, 2.5]	1	1	1	1	1	1
True frequency $w = 0.2$						
[0.1, 0.5]	.022	.027	.019	.016	.027	.016
[0, 1]	.007	.009	.005	.004	.006	.003
[0.5, 1.5]	.684	.217	.685	1	.895	1
[1.5, 2.5]	1	1	1	1	1	1
True frequency $w = 1.5$						
[1, 2]	.017	.017	.014	.018	.017	.015
[1.5, $\pi$ ]	.031	.029	.034	.037	.032	.036
[0, 1]	1	1	1	1	1	1
[2, 3]	1	1	1	1	1	1

Notes: Empirical rejection frequencies, nominal significance level 0.05, 5000 replications. The value "1" means unity up to a precision of six decimal digits. Power is raw (not size-adjusted).

$\nabla_0(L) = 1 - L$  we would have  $\alpha_1 \neq 0$  and  $b_1^0 = \kappa\alpha_1$ , which contradicts non-causality of  $x$ . This shows that a test of  $b_1^0 = 0$  is asymptotically equivalent to a test of  $\alpha_1 = 0$  (given a rate- $T$ -consistent estimate  $\hat{\kappa}$ ), but only under the maintained hypothesis of cointegration ( $a_1 \neq 0$ ). Under this maintained hypothesis the coefficient is effectively attached to the  $I(0)$  error term  $e_t$  and inference is standard. In contrast, a joint test of no long-run causality between two  $I(1)$  variables in both directions would be equivalent to a test of the null of no cointegration, with non-standard but also well-known inference procedures.

## 6. AVERAGE CAUSALITY OVER A BAND RANGE

It is impossible to impose  $H_0^{all}$  in terms of the linear VAR model (3.3) because a (lag) polynomial can only have a finite number of roots, hence  $b_1^\omega = b_2^\omega = 0$  cannot hold for infinitely many  $\omega$ . In fact, in order to factor out a second  $\nabla_{\omega^{**} \neq \omega^*}$  polynomial from the  $\gamma^{\omega^*}(L)$  polynomial with another non-causal frequency  $\omega^{**}$ ,  $p \geq 5$  would be required, and in general the number of non-causal frequencies is bounded by  $(p - 1)/2$ .

This impossibility of strict non-causality in a frequency band is a fundamental property that is related to the well-known result that the ideal (band pass, low pass, or high pass) filter in the frequency domain implies a two-sided filter in the time domain. Such a two-sided filter is often and adequately called non-causal, because it destroys the temporal ordering and thus invalidates analyses of predictive properties.

To illustrate this problem, consider the one-sided version of the flexible band-pass filter suggested by Christiano and Fitzgerald (2003), where we

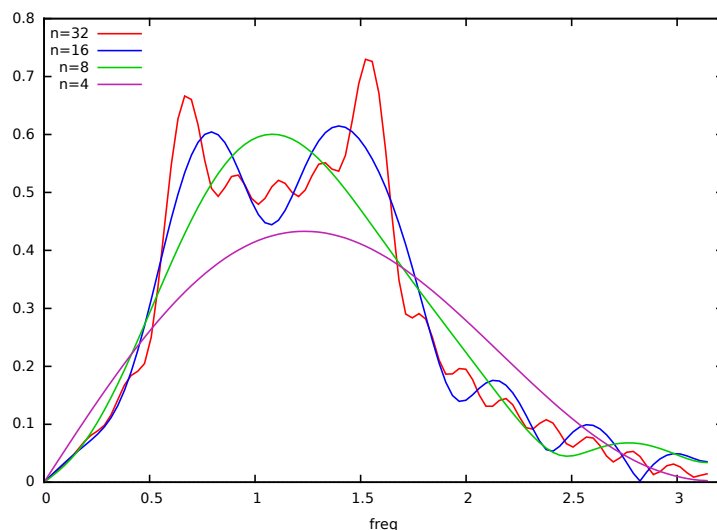


FIGURE 6.1. Christiano-Fitzgerald one-sided and truncated filter gain

focus on the incarnation based on the random-walk assumption.<sup>3</sup> The frequency response (gain) function of the filter for various filter length truncation values of  $n$  and for a pass band of frequencies corresponding to 4 to 10 periods of oscillation (frequencies  $[0.63, 1.57]$ ) is shown in figure 6.1. We can see that the filter deviates considerably from an ideal filter with zero gain up to 0.63 and beyond 1.57, and raising the number of lagged terms only seems to make the gain flatter in the pass band, but not in the stop bands.

Acknowledging this impossibility of implementing strict Granger non-causality over a frequency range, we are still interested in the hypothesis that the degree of causality in some frequency band is different from

<sup>3</sup>That filter is given by (their equation (4)):

$$y_t^F = 0.5B_0y_t + B_1y_{t-1} + \dots + B_{t-2}y_2 + \tilde{B}_{t-1}y_1,$$

where  $B_0 = (b - a)/\pi$ ,  $a = 2\pi/l_u = \omega_l$ ,  $b = 2\pi/l_l = \omega_u$ ,  $B_{j \geq 1} = (\sin(jb) - \sin(ja))/(\pi j)$ ,  $\tilde{B}_{t-1} = -0.5B_0 - \sum_{i=1}^{t-2} B_i$ . Here  $l_u$  and  $l_l$  denote the bounds of the wavelengths (periods of oscillation). This one-sided filter uses all available past data, but for practical purposes one can choose a certain number of coefficients  $n$  and apply the following truncated filter (to observations satisfying  $t > n - 1$ ), where for asymptotic arguments  $n$  can grow up to a rate  $T$ :

$$y_t^F = 0.5B_0y_t + B_1y_{t-1} + \dots + B_{n-2}y_{t-(n-2)} + \tilde{B}_{n-1}y_{t-(n-1)}$$

that in another frequency range. For example one could analyze whether short-term interest rates are really less causal for long-term rates at high frequencies than at low frequencies as predicted by theory, or whether a certain business-cycle indicator has higher causality for output growth at business-cycle frequencies compared to lower frequencies. It is natural to choose the case of equality as the null hypothesis and treat the inequality case as the alternative.

- $H_0^e$  : The degree of Granger causality of  $x$  for  $y$  in the frequency band  $\Omega_0 = [\omega_l, \omega_u]$  is equal to that in another band  $\Omega_1$  (with non-overlapping bands  $\Omega_0 \cap \Omega_1 = \emptyset$ ).
- $H_1^{more}$ : The Granger causality of  $x$  for  $y$  is higher in  $\Omega_0$  than in  $\Omega_1$ .

$H_0^e$  includes the traditional case of non-causality at all frequencies. The additional alternative hypothesis that causality is lower in  $\Omega_0$  than in  $\Omega_1$  is obviously equivalent to  $H_1^{more}$  with the bands  $\Omega_0$  and  $\Omega_1$  reversed, hence we do not address it explicitly. We believe that the combined inequality hypothesis that Granger causality is different between  $\Omega_0$  and  $\Omega_1$  (lower or higher) without specifying the direction is not interesting for meaningful applications. Note that there may or may not be frequencies with strict non-causality present in  $\Omega_0$  or  $\Omega_1$ , the presence of which is neither necessary nor sufficient for  $H_0^e$  to be rejected.

As the relevant measure of Granger causality we use the frequency response or gain function of the causal filter that the VAR system represents and by which it transforms innovations in  $x$  to reactions of  $y$ . If we denote with  $g_\omega^{xy}(A(L))$  this gain at frequency  $\omega$ , the degree of Granger causality in the band  $\Omega_0$  is given by  $g_{\Omega_0}^{xy}(A(L)) = \int_{\omega_l}^{\omega_u} g_\omega^{xy}(A(L)) d\omega$ . The central element of the hypothesis  $H_0^e$  (and  $H_0^{eu}$ ) is whether  $g_{\Omega_0}^{xy}(A(L)) = g_{\Omega_1}^{xy}(A(L))$ .

We will consider an isolated unit impulse in  $x_t$  or equivalently in  $u_{x,t}$  which is the reduced-form (non-structural) innovation, without changing  $u_{y,t}$  or  $y_t$  contemporaneously, even though  $Cov(u_t)$  may be non-diagonal.

This approach is natural for a Granger causality analysis.<sup>4</sup> Viewing the VAR model as a linear filter, the implied gain function is the Fourier transform of its impulse response function (IRF) which can be derived from the moving average representation of the model,  $(x_t, y_t)' = A(L)^{-1}(u_{x,t}, u_{y,t})'$  with  $A(L)^{-1} \equiv \Phi(L) = I + \sum_{i=1}^{\infty} \Phi_i$ . Standard textbook results show (e.g. Lütkepohl, 2007) that the coefficient matrices of the MA form are computable recursively by

$$\Phi_{i \geq 1} = \sum_{j=1}^i \Phi_{i-j} A_j$$

(where  $\Phi_0 = I$ ). The interesting responses of  $y$  to an impulse in  $x$  are given by the south-west corners of  $\Phi_{i \geq 1}$ , let the corresponding scalar infinite lag polynomial be  $\phi(L) = \phi_1 + \phi_2 L + \dots$ . The aggregate gain in the band  $\Omega_0$  is then given by  $g_{\Omega_0}^{xy}(A(L)) = \int_{\omega_l}^{\omega_u} |\phi(e^{i\omega})|^2 d\omega$ , and analogously for the band  $\Omega_1$ . Given estimated VAR coefficients  $\hat{A}(L)$  we can derive estimates  $\hat{g}_F^{xy}$  for  $F = \Omega_0, \Omega_1$  by approximating the infinite-order  $\hat{\phi}(L)$  with a truncated version  $\tilde{\phi}_n(L)$  up to order  $n$  and apply the Fourier transform to  $\tilde{\phi}_n(L)$ .

Deriving the sampling distribution of  $\hat{g}_F^{xy}$  is related to the problem of estimating the distribution of the IRF which is typically addressed in the literature with bootstrapping methods (cf. Kilian, 1998) because it is analytically intractable. Here the analytic problem is aggravated by the additional Fourier transform step, but conceptually the bootstrap approach covers this problem as well.

**Lemma 2.** *Given consistent and asymptotically normally distributed estimates of the VAR coefficients  $\hat{a} = \text{vec}(\hat{A}(L))$  with corresponding variance-covariance matrix  $\hat{\Psi}_a$ , a consistent test of  $H_0^c$  against  $H_1^{\text{more}}$  is given by considering the simulated distribution of  $\hat{g}_{\Omega_0}^{xy} - \hat{g}_{\Omega_1}^{xy} = \hat{g}$  and rejecting if its  $1 - \alpha$  quantile is positive.*

<sup>4</sup>The result is identical to the effect of a structural innovation  $\epsilon_{1,t}$  that only affects  $x_t$ , i.e. applying a Choleski identification strategy with an upper-triangular impact matrix. However, the concept of Granger causality focuses on variables and not on shocks. Considering shocks that move  $x_t$  and  $y_t$  simultaneously would mean to introduce instantaneous causality, which is another concept that does not fit well with the notion of Granger causality, which is based on predictive ability.



The distribution can be generated with a parametric bootstrap, by drawing repeatedly from the  $k$ -variate normal distribution  $N(\hat{a}, \hat{\Psi}_a)$ , where  $k = 4p$ , and calculating  $\tilde{\phi}_n(L)$  and  $\hat{g}$  for each simulation draw, with  $n \rightarrow \infty$  as  $T \rightarrow \infty$ .

*Proof.* The consistency of the bootstrap for the distribution of  $\hat{\phi}(L)$  follows from the existing literature. Some small-sample bias corrections may also be applied. We have  $\lim_{T \rightarrow \infty} \tilde{\phi}_n(L) = \hat{\phi}(L)$ . The gain function is a differentiable function of the impulse response for stable VAR systems and hence integrable over frequencies. The theoretical gain difference  $g$  is freely varying and in particular not bounded by 0. Hence the bootstrap carries over to an evaluation of  $\hat{g}$ .  $\square$

Sometimes it may also be of interest to analyze  $H_0^c$  after  $H_0^u$  has been already established, i.e. conditional on the assumption that a frequency  $w$  exists at which  $x$  is not causal for  $y$ . In this case the south-west corner of  $I - A(L)$  would be given by  $\beta(L) = \gamma^w(L)\nabla_w(L)$  which implies for example  $\phi_1 = \gamma_0^w$ . In this restricted VAR the number of parameters to be estimated drops from  $4p$  to  $4p - 2$ . If  $H_0^c$  &  $H_0^u$  are to be jointly tested with this two-stage method, a Bonferroni-type approach can be used to control the overall size of the procedure and to avoid pre-testing distortions.

## 7. ILLUSTRATION

Many economic relationships where the Granger causality should theoretically vary with the frequency are related to expectation formation. Apart from the previously mentioned example of the term structure of interest rates another classical example would be the smoothing of consumption according to the persistent movement of permanent income which is essentially the expected net present value of all resources over the life cycle. In applications where expectations matter, special care must be taken to distinguish between the concepts of Granger causality and structural causality. In the consumption-smoothing theory, the persistent component

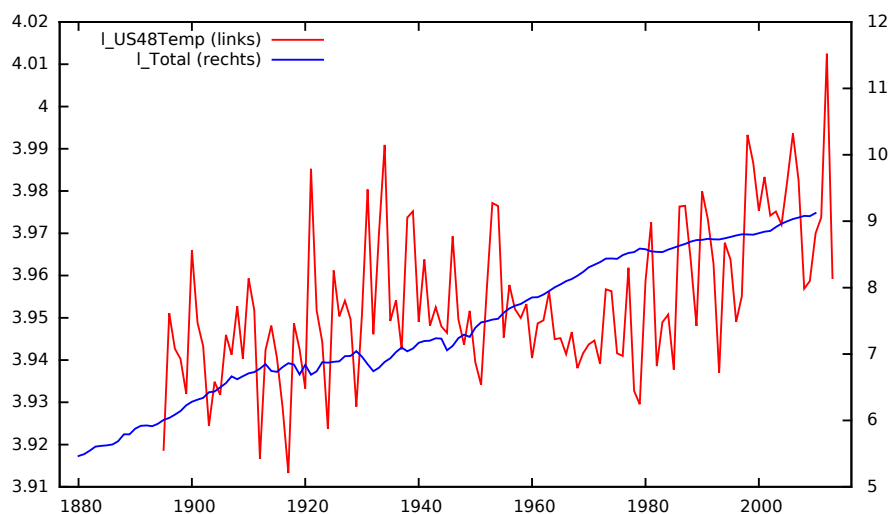


FIGURE 7.1. Time series of US continental (48 states) temperatures (log degrees Fahrenheit) and total CO2 emissions (log millions of metric tons).

of expected income is the underlying cause of parallel movements of consumption. In empirical studies realized income must be used instead, given that expected future income is not contemporaneously observable. This choice effectively reverses the temporal ordering of cause and effect, with the result that consumption would turn out as Granger-causal for income at low frequencies.

If properly interpreted, these difficulties do not invalidate the use of Granger causality analyses, but to avoid confusion we instead present an example where expectations are unlikely to matter. We analyze US temperature measurements and greenhouse gas emissions for Granger non-causality. We use the annualized data 1895 to 2013, because emissions are not available quarterly or monthly.

The temperature data are from the US National Climatic Data Center (Climate at a Glance), while the CO2 emissions data are the “Total” series from Boden, Marland, and Andres (2014) ranging from 1751 to 2010. The jointly available sample is thus 1895-2010 ( $T = 116$ ) and we use log transforms.

First we determine the lag order of the bivariate VAR in log-levels. The Akaike information criterion suggests just two lags, but the third lag is also significant at the 10% level, and as explained above at least three lags are needed in order to distinguish causality at different frequencies. Hence we choose  $p = 3$ . In principle it would be possible to consider a more complicated lag structure, restricting some of the intermediate lag coefficients in certain equations to zero, but we do not pursue this strategy in this illustration.

For these data it is natural to suspect cointegration so we run the Johansen test, with an unrestricted constant to deal with the trending data. The highest eigenvalue is 0.16 and the p-value of the trace test of no cointegration yields 0.0098, such that there is evidence for cointegration at the nominal 1% significance level. Notice that the error correction term is insignificant in the emissions equation (p-value of 0.12), hence emissions do not seem to be caused by temperatures at frequency zero. This is plausible, but it also means that we cannot test the restriction of no long-run causality running in the other direction –from emissions on temperatures– without affecting the cointegration property of the system, as discussed above. Therefore the causality at frequency zero is already established, and our interpretation focuses on non-zero frequencies. Figure 7.2 shows the frequency-wise test results.

For any frequency band up to roughly 0.2, corresponding to wavelengths down to roughly 31 periods (years) the minimal test statistics would exceed the critical value, and hence for those frequency bands we would reject the null hypothesis that there exists a frequency without Granger causality. Of course the frequency band below 0.2 is very close to zero, and with this effective sample of  $T = 113$  it is very difficult to distinguish cycles of 30 periods from even lower frequencies. Hence some leakage from the zero frequency is expected. For any frequency bands containing higher frequencies (shorter wavelengths) we would not be able to reject the corresponding

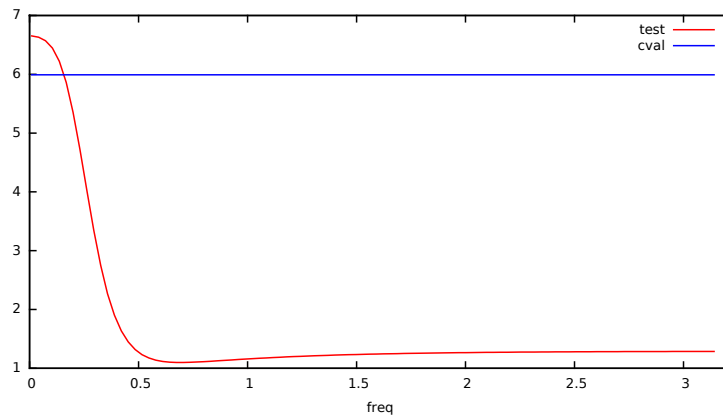


FIGURE 7.2. Frequency-wise causality test from log total emissions on log US continental temperatures. System with 3 lags. The horizontal line is the critical value of the  $\chi^2$  distribution with two degrees of freedom at the 5% level. The lowest tested frequency here is 0.01, see the text for the zero frequency.

null hypothesis. The overall conclusion is thus that Granger causality from emissions to temperatures varies across frequencies.

## 8. SUMMARY

In this paper we have shown that tests of Granger non-causality can also be specified in terms of frequency bands or intervals instead of single frequency points. The result is a rigorous testing framework enabling standard inference that circumvents the ad-hoc procedures of joint testing with unknown statistical properties. The implementation is easy because in practice the relevant test statistic is just the minimum over the specified frequency band, apart from a special but equally easy treatment of the frequencies 0 and  $\pi$ . In a preliminary simulation study the test performed satisfactorily albeit conservatively.

However, given that strict non-causality over a range of frequencies is impossible in this (linear) framework except if there is no causality at all, accepting the null hypothesis still means that some causality exists in the band of the null hypothesis. For practical purposes it may therefore be advisable to keep the specified frequency band reasonably short. Other

types of hypothesis are also conceivable, dealing with a minimum amount of causality under the null instead of strict non-causality. We plan to address these additional hypotheses in the next revision of this manuscript.

Our empirical application with long time series of CO<sub>2</sub> emissions and earth surface temperatures demonstrated that varying degrees of Granger causality in the frequency domain are of practical relevance. In the introduction we already mentioned the case of the term structure of interest rates where such varying connections are also expected. In addition, according to the economic hypothesis of consumption smoothing a similar result about differing impacts of short- versus long-term fluctuations might hold between income and consumption. We believe that many more potential applications in economics and perhaps other disciplines are likely to exist.

#### REFERENCES

- ASSENMACHER-WESCHE, K., AND S. GERLACH (2007): "Money at Low Frequencies," *Journal of the European Economic Association*, 5(2-3), 534–542.
- (2008a): "Interpreting euro area inflation at high and low frequencies," *European Economic Review*, 52, 964–986.
- (2008b): "Money growth, output gaps and inflation at low and high frequency: Spectral estimates for Switzerland," *Journal of Economic Dynamics and Control*, 32, 411–435.
- BACCALÁ, L. A., AND K. SAMESHIMA (2001): "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, 84, 463–474.
- BEVERIDGE, S., AND C. R. NELSON (1981): "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle," *Journal of Monetary Economics*, 7, 151–174.

- BODEN, T. A., G. MARLAND, AND R. J. ANDRES (2014): "Global, Regional, and National Fossil-Fuel CO<sub>2</sub> Emissions," .
- BREITUNG, J., AND B. CANDELON (2006): "Testing for short- and long-run causality: A frequency-domain approach," *Journal of Econometrics*, 132, 363–378.
- CHRISTIANO, L. J., AND T. J. FITZGERALD (2003): "The Band Pass Filter," *International Economic Review*, 44(2), 435–465.
- 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.
- GEWEKE, J. (1982): "Measurement of Linear Dependence and Feedback Between Multiple Time Series," *Journal of the American Statistical Association*, 77, 304–324.
- GRANGER, C. W. J. (1969): "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, 37, 424–438.
- KILIAN, L. (1998): "Small-Sample Confidence Intervals For Impulse Response Functions," *The Review of Economics and Statistics*, 80(2), 218–230.
- LEMMENS, A., C. CROUX, AND M. G. DEKIMPE (2008): "Measuring and testing Granger causality over the spectrum: An application to European production expectation surveys," *International Journal of Forecasting*, 24, 414–431.
- LÜTKEPOHL, H. (2007): *New Introduction to Multiple Time Series Analysis*. Springer.
- SHILLER, R. J. (1979): "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," *Journal of Political Economy*, 87(6), 1190–1219.