A Bootstrap Method to Test Granger-Causality in the Frequency Domain

Matteo Farnè[1] · **Angela Montanari[1]**

Abstract

We propose a bootstrap test for unconditional and conditional Granger-causality spectra in the frequency domain. Our test aims to detect if the causality at a particular frequency is systematically different from zero. In particular, we consider a stochastic process derived applying independently the stationary bootstrap to the original series. At each frequency, we test the sample causality against the distribution of the median causality across frequencies estimated for that process. Via our procedure, we infer about the relationship between money stock and GDP in the Euro Area during the period 1999–2017. We point out that the money stock aggregate M1 had a significant impact on economic output at all frequencies, while the opposite relationship is significant only at low frequencies.

Keywords Bootstrap tests · Granger-causality spectra · Money stock and GDP · Euro Area

1 Introduction

As a statistical concept, causality has a central role both from a theoretical and a practical point of view (see Berzuini et al. 2012). In time series analysis, it was first introduced by Wiener in the context of prediction theory (Wiener 1956) and then formalized by Granger in the context of linear regression modelling of stochastic processes (Granger 1969). Since then it is denoted as Granger-causality (GC). Specifically, Y_t is said to Granger-cause X_t if the past values of Y_t, i.e. $Y_{t-1}, Y_{t-2}, \ldots$, help predicting the current value of X_t.

✉ Matteo Farnè
matteo.farne@unibo.it

[1] Department of Statistical Sciences, Università di Bologna, Via delle Belle Arti 41, Bologna, Italy

The concept of causality was then extended from the time to the frequency domain. Pierce (1979) explicitly measured the effect of the lagged values of Y_t onto X_t through an R^2 indicator, defined frequency-wise to evaluate the intensity of the linear causal dependence at each frequency.

Geweke (1982) proposed to estimate a VAR model on $[X_t, Y_t]'$ by defining the causality from Y_t to X_t at each frequency as the proportional effect of the lagged values of Y_t onto the spectral density of X_t. This approach can be implemented even considering the conditional causality with respect to the lagged values of a third variable W_t (Geweke 1984). The described measures are respectively defined as unconditional and conditional Granger-causality spectra (Geweke 1982, 1984), and also discussed by Hosoya (1991) and Hosoya (2001). The conditional Granger-causality has then been reformulated by Chen et al. (2006).

While the use of GC in the time-domain dates back to the sixties, GC in the frequency domain has been employed in more recent years. In Lemmens et al. (2008) the causality structure of European production expectation surveys is analyzed by the methods of Pierce (1979) and Geweke (1982) comparatively, which require to study appropriate frequency-wise coefficients of coherence. The same approach is used in Tiwari (2014) for exploring the relationship between energy consumption and income in the United States. In this paper, we aim to study the mutual relationship between gross domestic product and money stock in the Euro Area by making inference on Granger-causality spectra. The problem has been widely addressed for the U.S. economy, which is a key player from an international macroeconomic perspective. In Belongia and Ireland (2016), for instance, the methodology of Friedman and Schwartz (1975), based on structural VAR models, is revisited and applied to U.S. data across the period 1967–2013. The same consideration applies to the Indian economy (see Nadig and Viswanathan 2019). On the contrary, since the European Central Bank was born in 1998, evidences and analyses regarding the Euro Area are less established and more controversial.

The advantage of frequency-domain GC lies in the disentanglement of the causality structure across a range of frequencies, while traditional time-domain GC only provides an overall indication of the presence of a causality relationship. In particular, GC measures at each frequency the intensity of the causal relationship between the components of the series at that particular frequency. For this reason, we aim to provide an inferential procedure to highlight at which specific frequency there exists causality among the data generating processes of the observed time series.

In Breitung and Candelon (2006), a parametric test (henceforth, BC test) for Granger-causality in the frequency domain is proposed. Its convergence rate is $O(T^{-1/2})$ (where T is the time series length) and its power is decreasing as the distance of the frequency of interest from $\frac{\pi}{2}$ increases (even if Yamada and Yanfeng 2014 show that the same test is still useful at extreme frequencies). The test is based upon a set of linear restrictions on the parameters of the (possibly cointegrated) VAR model best representing the series (we refer to Lütkepohl 2005 for VAR selection and estimation). Breitung and Schreiber (2018) generalize the BC procedure to test causality and delay within a frequency band.

BC test is widely used in the economic literature. In Croux and Reusens (2013), it is applied to test the predictive power of domestic stock prices for the economic activ-

ity of G-7 countries. Bozoklu and Yilanci (2013) consider the relationship between energy consumption and economic growth in the OECD area. The relationship between real and financial business cycles is analyzed in Gomez-Gonzalez et al. (2015). Wei (2015) deals with the link between commodity prices and monetary policies. Keho et al. (2015) test Wagner's law for some African economies. Two relevant applications for Turkish economy are described in Gül et al. (2018) and Tastan and Sahin (2020). Fromentin and Tadjeddine (2020) analyze the relationship between cross-border workers and financial instability in Luxembourg.

Some alternative testing approaches were also proposed. Hidalgo (2000) estimates VAR filters via generalized least squares and then accordingly derives a test statistics for GC in the presence of long-range dependence. Hidalgo (2005) extends the framework of Hidalgo (2000) to the multivariate case computing relevant quantiles under the null via resampling bootstrap. A consistent nonparametric test for nonlinear causality is proposed by Nishiyama et al. (2011). A nonparametric approach to test for Granger-causality in tail events is proposed by Hong et al. (2009) and extended by Candelon and Tokpavi (2016). The Philipps spectral estimator (Phillips 1988) is instead exploited to estimate causality both at frequency 0 and at the rest of frequencies in a cointegrated setting by Assenmacher-Wesche and Gerlach (2008). Such a method is used in Berger and Osterholm (2011) to test the relationship between money growth and inflation in the Euro Area. A comprehensive computational and inferential strategy for time-domain and frequency-domain GC spectra has been proposed in Barnett and Seth (2014).

The present work proposes a different approach to the classical testing framework of the no-causality hypothesis. Our null hypothesis of zero-causality, in fact, is tested frequency-wise comparing each causality to the $100 \times (1-\alpha)-$th percentile of the distribution of the median causality across frequencies under the stronger case of stochastic independence, where α is the significance level. The null distribution is approximated via the stationary bootstrap of Politis and Romano (1994). As empirically shown in Sect. 2.4, the median is chosen because under the null the individual causalities and the median causality are indistinguishable from a stochastic point of view. We thus derive the desired bootstrap quantile of the median causality across frequencies and we compare each observed causality to it. Our test is consistent provided that $T^{\frac{1}{3}} \to \infty$.

The paper is organized as follows. In the next Section the concept of Granger-causality in the frequency domain is recalled, our bootstrap inference approach is explained in detail and a simulation study which clarifies the features of our test is presented. In Sect. 3 we show the potentialities of our method in outlining the causal relationships between money and output in the Euro Area during the period 1999-2017. Finally, we conclude the paper with a discussion.

2 Granger-Causality Spectra: A Bootstrap Testing Approach

2.1 Definition

We now briefly recall the bases of Granger-causality spectral theory. We follow the approach in Ding et al. (2006), which we refer to for the details. Suppose that the

stochastic processes X_t and Y_t, jointly covariance-stationary, follow a non-singular $VAR(p)$ model. Defining the bivariate stochastic process $Z_t = [X_t, Y_t]'$, we have

$$Z_t = A_1 Z_{t-1} + \ldots + A_p Z_{t-p} + \epsilon_t, \tag{1}$$

where ϵ_t is a zero-mean bivariate stochastic process with covariance matrix Σ_2 and $A_1, \ldots, A_p$ are 2×2 coefficient matrices. We stress that ϵ_t has no auto-covariance structure at any lag different from 0.

Moving to the frequency domain, for each frequency ω we define the transfer function $P(\omega)$ of Z_t in (1) as

$$P(\omega) = \left(I_2 - \sum_{j=1}^{p} A_j e^{-ij\omega} \right)^{-1}, \quad -\pi \leq \omega \leq \pi, \tag{2}$$

which is invertible if and only if the roots of the equation $\det(I_2 - \sum_{j=1}^{p} A_j L^j) = 0$ (where L is the lag operator) lie within the unit circle. Setting $P(\omega) = \begin{bmatrix} P_{XX}(\omega) & P_{XY}(\omega) \\ P_{YX}(\omega) & P_{YY}(\omega) \end{bmatrix}$, the definition (2) allows to define in a compact way the model-based spectrum $h(\omega)$ as follows:

$$h(\omega) = P(\omega) \Sigma_2 P(\omega)^*, \quad -\pi \leq \omega \leq \pi,$$

where $*$ denotes the complex conjugate transpose.

Setting $\Sigma_2 = \begin{bmatrix} \sigma_2 & \upsilon_2 \\ \upsilon_2 & \gamma_2 \end{bmatrix}$, we need for computational reasons to define the transform matrix $S = \begin{bmatrix} 1 & 0 \\ -\frac{\upsilon_2}{\sigma_2} & 1 \end{bmatrix}$, from which we derive the transformed transfer function matrix $\tilde{P}(\omega) = P(\omega) \times S^{-1}$. The process $Z_t = [X_t, Y_t]'$ is normalized accordingly as $\tilde{Z}_t = \tilde{P}(L)[X_t, Y_t]'$ and becomes $\tilde{Z}_t = [\tilde{X}_t, \tilde{Y}_t]'$.

The unconditional Granger-causality spectrum of X_t (effect-variable) respect to Y_t (cause-variable) is then defined as (Geweke 1982)

$$h_{Y \to X}(\omega) = \ln \left(\frac{h_{XX}(\omega)}{\sigma_2 |\tilde{P}_{XX}(\omega)|^2} \right). \tag{3}$$

In the empirical analysis, the theoretical values of coefficient and covariance matrices will be replaced by the corresponding SURE estimates (Zellner 1962).

Moreover, we can define the conditional Granger-causality spectrum of X_t respect to Y_t given an exogenous variable W_t (conditioning variable). Suppose that we estimate a VAR on $[X_t, W_t]'$ with covariance matrix of the noise terms $\Sigma_{2'} = \begin{bmatrix} \sigma_{2'} & \upsilon_{2'} \\ \upsilon_{2'} & \gamma_{2'} \end{bmatrix}$ and transfer function $G(\omega)$ [defined as in (2)]. The corresponding normalized process of $[X_t, W_t]'$ (according to the procedure described above) is denoted by $[\tilde{X}_t, \tilde{W}_t]'$.

We then estimate a VAR on $[X_t, Y_t, W_t]'$ with covariance matrix of the noise terms

$$\Sigma_3 = \begin{bmatrix} \sigma_{XX} & \sigma_{XY} & \sigma_{XW} \\ \sigma_{YX} & \sigma_{YY} & \sigma_{YW} \\ \sigma_{WX} & \sigma_{WY} & \sigma_{WW} \end{bmatrix}$$

and transfer function $P'(\omega)$. Building the matrix

$$C(\omega) = \begin{bmatrix} G_{XX}(\omega) & 0 & G_{XW}(\omega) \\ 0 & 1 & 0 \\ G_{WX}(\omega) & 0 & G_{WW}(\omega) \end{bmatrix},$$

we can define $Q(\omega) = C^{-1}(\omega)P'(\omega)$, which is a sort of "conditional" transfer function matrix. The theoretical spectrum of $\tilde{X}_t$ can thus be written as

$$h_{\tilde{X}\tilde{X}}(\omega) = Q_{XX}(\omega)\sigma_{XX}Q_{XX}(\omega)^* \\ + Q_{XY}(\omega)\sigma_{YY}Q_{XY}(\omega)^* + Q_{XW}(\omega)\sigma_{WW}Q_{XW}(\omega)^*.$$

The conditional spectrum of X_t (effect-variable) respect to Y_t (cause-variable) given W_t (conditioning variable) is (Geweke 1984)

$$h_{Y \to X|W}(\omega) = \ln\left(\frac{h_{\tilde{X}\tilde{X}}(\omega)}{\sigma_{XX}|Q_{XX}(\omega)|^2}\right). \tag{4}$$

Both $h_{Y \to X}(\omega)$ and $h_{Y \to X|W}(\omega)$ range from 0 to ∞, with $-\pi \le \omega \le \pi$. $h_{Y \to X}(\omega)$ expresses the power of the relationship from Y to X at frequency ω, $h_{Y \to X|W}(\omega)$ expresses the strength of the relationship from Y to X at frequency ω given W. Therefore, the unconditional spectrum accounts for the whole effect of the past values of Y_t onto X_t, while the conditional spectrum accounts for the direct effect of the past values of Y_t onto X_t excluding the effect mediated by the past values of W_t.

Granger-causality spectra $h_{Y \to X}(\omega)$ and $h_{Y \to X|W}(\omega)$ can be interpreted as follows. If $h_{Y \to X}(\omega) > 0$, it means that past values of Y_t help predicting X_t, and $\frac{1}{\omega}$ is a relevant cycle. If $h_{Y \to X|W}(\omega) > 0$, it means that past values of Y_t in addition to those of W_t help predicting X_t, and $\frac{1}{\omega}$ is a relevant cycle. Significant frequencies give us some hints on the relevant delay structure of the cause variable with respect to the effect variable.

2.2 Testing Framework

The inference on Granger-causality spectra in the frequency domain is still an open problem. In fact, differently from the corresponding time-domain quantities, the limiting distribution for unconditional and conditional spectra under the null hypothesis of no-causality is unknown (see Barnett and Seth 2014, Sect. 2.5). In spite of that, Breitung and Candelon (2006) test the nullity of unconditional and conditional GC at each frequency ω, imposing a necessary and sufficient set of linear restrictions to the

(possibly cointegrated) VAR model best fitting the series. The resulting test statistics is distributed under their null as a Fisher distribution with 2 and $T - 2p$ degrees of freedom (except for $\omega = \{0, \pi\}$ at which the distribution is $F_{(1,T-p)}$), where p is the VAR delay and T is the time series length.

On the contrary, as explained in the Introduction, our benchmark is GC spectrum under the null hypothesis of stochastic independence. At each frequency ω, we test in Fisher's sense the *null* hypothesis $H_0 : r(\omega) = 0$ where the functional $r(\omega)$ may be the unconditional GC $h_{Y \to X}(\omega)$, assuming that X_t is stochastically independent from Y_t, or the conditional GC $h_{Y \to X|W}(\omega)$, assuming that X_t is stochastically independent from Y_t conditionally on W_t. This equals to assume that under the null the bivariate stochastic process $[X_t, Y_t]$ is an independent process, unconditionally or conditionally on W_t.

In order to test the null hypothesis, we refer the estimated causality at each frequency to the distribution of $\hat{r}_{med}$, the estimated median of $r(\omega)$ across frequencies, approximated by the *stationary bootstrap* of Politis and Romano (1994). The median is chosen because under the null all causalities are indistinguishable with respect to the median causality from a stochastic point of view (see Sect. 2.4). We stress that unconditional and conditional spectra must be assessed separately, because their distributions are different under the null.

Our idea derives from Politis and Romano (1994), according to which each *Fréchet-differentiable* functional may be successfully approximated by the stationary bootstrap, and the resulting bias depends on the sum of the *Fréchet-differential* g_F evaluated at each X_t. The bootstrap series obtained via the stationary bootstrap of Politis and Romano (1994) are stationary Markov chains conditionally on the data. This means that each bootstrap series $X_1^*, \ldots, X_T^*$ is a Markov chain conditionally on $X_1, \ldots, X_T$.

Suppose that we apply the stationary bootstrap procedure independently to X_t and Y_t, obtaining the stationary bootstrap series X_t^* and Y_t^*. Computing unconditional Granger-causality spectra on those series equals to assess unconditional causalities under the assumption of stochastic independence, because X_t^* and Y_t^* result to be independent Markov chains. Therefore, testing each Granger-causality computed on the original series X_t, Y_t against the median causality computed across frequencies on X_t^* and Y_t^* is effective as a test for causality strength under the null.

For the conditional case, we proceed as follows. First, we estimate a VAR model on the observed series X_t and W_t jointly considered. Then, we apply the residual bootstrap (see Tibshirani and Efron 1993) to VAR residuals, thus obtaining the consistent residual bootstrap series X_t^* and W_t^*. This step is needed in order to avoid any bias which may arise ignoring the dependence structure between X_t and W_t. Finally, we apply independently the stationary bootstrap to the observed series Y_t to obtain the series Y_t^*. Computing conditional Granger-causality spectra on X_t^*, Y_t^* and W_t^* then equals to assess conditional causalities under the assumption that X_t is stochastically independent from Y_t given W_t. For this reason, we use the resulting distribution of the conditional median causality across frequencies computed on those series as our benchmark under the null hypothesis.

Let us consider $\hat{r}(\omega) = \hat{h}_{Y \to X}(\omega)$, which is defined as (3) where the coefficient matrices A_j, $j = 1, \ldots, p$, and the error covariance matrix Σ_2 are replaced by the

corresponding SURE estimates (Zellner [1962]). We know that SURE estimates $\hat{A}_j$, $j = 1, \ldots, p$, are rational functions of X_t, thus being *Fréchet-differentiable*. $\hat{\hat{P}}(\omega)$ and $\hat{\Sigma}_2$ are functions of the $\hat{A}_j$, thus being rational in turn; the same holds as a consequence for $\hat{h}(\omega)$. Therefore, $\hat{h}_{Y \to X}(\omega)$, the natural logarithm of a rational function of the data, is *Fréchet-differentiable*. At this point, even if the median is not *Fréchet-differentiable*, bootstrapping $\hat{r}_{med} = med(\hat{h}_{Y \to X}(\omega))$ under the null is still valid, because the distribution of $\hat{r}_{med}$ is indistinguishable from the causality distribution at any frequency under the null as T is enough large (see Sect. 2.4). As a consequence, according to Politis and Romano ([1994]), paragraph 4.3, we can estimate consistently any quantile of the distribution of the median of $\hat{h}_{Y \to X}(\omega)$ across frequencies under the null hypothesis via the stationary bootstrap. Considering $\hat{r}(\omega) = \hat{h}_{Y \to X|W}(\omega)$, which is defined as (4) where the coefficient matrices and the error covariance matrix are replaced by the corresponding SURE estimates, a similar reasoning can be carried out.

We stress that our aim is not to represent the common multivariate distribution function F of the process $[X_t, Y_t]$ or $[X_t, Y_t, W_t]$. That problem is an estimation one, which would be effectively solved by parametric or residual bootstrap. Our aim is to exploit the random process $[X_t^*, Y_t^*, W_t^*]$ to derive the bootstrap quantile $q_{r,1-\alpha}^*$, which satisfies

$$P\left(r_{med}^* \leq q_{r,1-\alpha}^*\right) = 1 - \alpha,$$

where α is the significance level and r_{med}^* is the bootstrap median across frequencies of unconditional or conditional GC under the assumption of stochastic independence. Since the stationary bootstrap procedure is valid for $\hat{r}_{med}$ under the null, $P(r_{med}^* \leq q_{r,1-\alpha}^*)$ approximates consistently $P(\hat{r}_{med} \leq q_{r,1-\alpha}|H_0)$ as $T^{\frac{1}{3}} \to \infty$.

In more detail, suppose that r is a *Fréchet-differentiable* functional, that is, there exists some influence function g_F, giving the influence of single observations on the functional r with respect to some distribution function F, such that

$$r(G) = r(F) + \int g_F d(G - F) + o(||G - F||)$$

with $\int g_F dF = 0$ ($||.||$ is the supremum norm). We define the mixing coefficient (which controls for the overall amount of time dependence) as

$$\alpha_X(k) = \sup_{A,B} |P(A, B) - P(A)P(B)|$$

where A and B vary over events in the σ-fields generated by $\{X_t, t \leq 0\}$, $\{X_t, t \geq k\}$. The following Theorem, expanded by Politis and Romano ([1994]), holds.

Theorem 1 *Suppose that $X_t^\dagger$ and $Y_t^\dagger$ are strictly stationary stochastic processes with distribution functions $F_{X^\dagger}$, $F_{Y^\dagger}$. Assume that, for some $d \geq 0$, $E(g_{F_{X^\dagger}}(X_1^\dagger)^{2+d}) < \infty$, $\sum_{k=1}^{\infty} \alpha_{X^\dagger}(k)^{\frac{d}{2+d}} < \infty$ and $\sum_{k=1}^{\infty} k^2 \alpha_{X^\dagger}(k)^{1/2-\tau} < \infty$, $0 < \tau < \frac{1}{2}$. Further assume*

that these assumptions also hold for $Y_t^\dagger$. Then, under H_0, i.e. if the distribution function of the stochastic process $Z_t^\dagger = [X_t^\dagger, Y_t^\dagger]$ can be factorized as $F_{X^\dagger} F_{Y^\dagger}$, it holds

$$\sup_{c \in \mathcal{R}} \left| P\left\{\sqrt{T}(r(F_{Z^\dagger}^*) - r(\hat{F}_{Z^\dagger})) \le c\right\} - P\left\{\sqrt{T}(r(\hat{F}_{Z^\dagger}) - r(F_{Z^\dagger})) \le c\right\} \right| \xrightarrow{prob} 0$$

for any Fréchet-differentiable functional r under the assumption $T^{\frac{1}{3}} \to \infty$.

The key to prove Theorem 1 is to bootstrap the cause and the effect series independently, and then show that any inference on a *Fréchet-differentiable* functional is weakly consistent.

For what concerns the conditional case, the following Theorem holds.

Theorem 2 *Suppose that $Z_t^{''} = [X_t^\dagger | W_t^\dagger, W_t^\dagger | X_t^\dagger]$ and $Y_t^\dagger$ are strictly stationary stochastic processes with distribution functions $F_{Z^{''}}$, $F_{Y^\dagger}$. Assume that, for some $d \ge 0$, $E(g_{F_{Y^\dagger}}(Y_1^\dagger)^{2+d}) < \infty$, $\sum_{k=1}^{\infty} \alpha_{Y^\dagger}(k)^{\frac{d}{2+d}} < \infty$ and $\sum_{k=1}^{\infty} k^2 \alpha_{Y^\dagger}(k)^{1/2-\tau'} < \infty$, $0 < \tau' < \frac{1}{2}$. Further assume that $0 < E(g_{F_{Z^{''}}}(Z_1^{''})^2) < \infty$ and $\int_{D_{Z^{''}}} \{F_{Z^{''}}(z^{''})(1 - F_{Z^{''}}(z^{''}))\}^{\tau^{''}/2} dz^{''}$, $\tau^{''} > 0$. Finally, assume that $W_t^\dagger | X_t^\dagger$ is stochastically independent from $Y_t^\dagger | X_t^\dagger$. Then, under H_0, i.e. if $X_t^\dagger | W_t^\dagger$ is stochastically independent from $Y_t^\dagger | W_t^\dagger$, for the stochastic process $Z_t^{'} = [X_t^\dagger | W_t^\dagger, Y_t^\dagger, W_t^\dagger | X_t^\dagger]$ it holds*

$$\sup_{c \in \mathcal{R}} \left| P\left\{\sqrt{T}(r(F_{Z'}^*) - r(\hat{F}_{Z'})) \le c\right\} - P\left\{\sqrt{T}(r(\hat{F}_{Z'}) - r(F_{Z'})) \le c\right\} \right| \xrightarrow{prob} 0$$

for any Fréchet-differentiable functional r under the assumption $T^{\frac{1}{3}} \to \infty$.

By assuming that the functional r is $g_{\tau'}-$ *Fréchet-differentiable* with respect to $F_{Y^\dagger}$ and $g_{\tau^{''}}-$ *Fréchet-differentiable* with respect to $F_{Z^{''}}$ we can ensure bootstrap consistency in weak sense for any conditional GC. Note that the assumption of stochastic independence between $W_t^\dagger | X_t^\dagger$ and $Y_t^\dagger | X_t^\dagger$ under the null is needed to avoid any circular causality effect under the null, such that the stationary bootstrap may be meaningfully applied. We refer to "Appendix B" for the proofs.

2.3 Testing Procedure

We now report in detail the testing procedures relative to the two functionals. For the functional $h_{Y \to X}(\omega)$, our bootstrap procedure is

– Simulate N independent stationary bootstrap series (X_t^*, Y_t^*) given the observed series (X_t, Y_t).
– On each simulated series (X_t^*, Y_t^*):

    1. estimate a VAR model on (X_t^*, Y_t^*) via SURE using BIC for model selection.
    2. at Fourier frequencies $f_i = \frac{i}{T}, i = 1, \ldots, [\frac{T}{2}]$, compute $\hat{h}_{Y^* \to X^*}(2\pi f_i)$.
    3. compute $med_{\{f_i, i=1, \ldots, T/2\}} \hat{h}_{Y^* \to X^*}(2\pi f_i)$.

– Then, compute $q_{uncond,1-\alpha}$, the $100 \times (1 - \alpha)$-th percentile of the bootstrap distribution at Step 3 across the N bootstrap series, where α is the significance level.
– Finally, at each f_i, flag $\hat{h}_{Y \to X}(2\pi f_i)$ as *significant* if larger than $q_{uncond,\alpha}$.

For the functional $h_{Y \to X|W}(\omega)$, the procedure becomes

– Estimate a VAR on (X_t, W_t) via SURE using BIC for model selection.
– Simulate N residual bootstrap series (X_t^*, W_t^*) from VAR estimates (Tibshirani and Efron 1993).
– Simulate N stationary bootstrap series Y_t^* independently of (X_t^*, W_t^*) given the observed series Y_t.
– On each simulated series (X_t^*, Y_t^*, W_t^*):

1. estimate a VAR model on (X_t^*, W_t^*) and (X_t^*, Y_t^*, W_t^*) via SURE using BIC for model selection.
2. at Fourier frequencies $f_i = \frac{i}{T}, i = 1, \ldots, [\frac{T}{2}]$, compute $\hat{h}_{Y^* \to X^*|W^*}(2\pi f_i)$.
3. compute $med_{\{f_i, i=1,\ldots,T/2\}} \hat{h}_{Y^* \to X^*|W^*}(2\pi f_i)$.

– Then, compute $q_{cond,1-\alpha}$, the $100 \times (1 - \alpha)$-th percentile of the bootstrap distribution at Step 3 across the N bootstrap series.
– Finally, at each f_i, flag $\hat{h}_{Y \to X|W}(2\pi f_i)$ as *significant* if larger than $q_{cond,\alpha}$.

We provide an R package, called "grangers", performing these routines (see Declarations).

In addition, we extend our framework to test the nullity of $r(2\pi f_i), i = 1, \ldots, [\frac{T}{2}]$, across the entire frequency range. In order to do that, we apply the conservative Bonferroni correction, that is, we apply the testing procedure to each frequency with significance level $\frac{2\alpha}{T}$. In this way, we ensure that the overall level across frequencies is not larger than α under the null.

2.4 Test Features and Simulation Results

We now describe the performance of our test (with a significance level of 5%) in a number of situations, in comparison to BC test by Breitung and Candelon (2006). We do not use more competitors like Hidalgo (2000) and Nishiyama et al. (2011) for the following reasons. First of all, both approaches are intended to test the unconditional causality. Second, the assumption settings of both procedures are quite different and somehow more demanding: the former requires an explicit control of the spectral matrix, while the second requires particular conditions on the process distribution. Third, BC test has become the standard test for GC in the frequency domain for economic applications, as shown in the Introduction.

First of all, suppose that we simulate $N = 300$ replicates from a VAR process of length $T = 100, 300$ in the form (1) with $p = 1$, $\Sigma_2 = diag(1, 1)$ and no causality coefficients. The VAR delay is selected for each bootstrap setting by BIC criterion. Our first tested coefficient matrix A_1 is $A_{1,(jj)} = 0, 0.5, 1, j = 1, 2$. We observe that for $T = 300$ the estimated level is approximately 5% at all Fourier frequencies as long as $A_{1,(jj)} = 0$ (Case 1), both for the unconditional and the conditional GC. If $\Sigma_2 = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, the estimated level is still close to 5% when $\rho = 0.5$.

As $A_{1,(jj)}$ increases (Case 2), our test presents an increasing rejection rate particularly at low frequencies. This happens because our test signals any violation of the null hypothesis of stochastic independence. As $A_{1,(jj)}$ approaches 1, the power tends to 1 at the lowest frequency, as showed for the unconditional GC in Fig. 1. On the contrary, BC test has a level close to 5% over all frequencies as $A_{1,(jj)} = 0, 0.5$, and approaches 0.4 as $A_{1,(jj)} = 1$. A similar pattern is observed also for the conditional GC in Fig. 2.

In order to show that under the null hypothesis the distributions of the individual causalities are indistinguishable from the one of the median causality, we have calculated the Kullback-Leibler (KL) divergence (Kullback and Leibler 1951) between the density functions of each causality and the median causality, estimated via a Gaussian kernel with bandwidth equal to 0.25. The results for Case 1 (with $A_{1,(jj)} = 0$) are reported in Fig. 3. We note that the observed divergence is extremely small, considered that the KL-divergence between two simulated uniform distributions located in [0, 1] is around 0.2. The pattern of KL over frequencies may vary, but its magnitude is always very small, even if we set $T = 100$. Similar results are observed for the conditional case, or allowing for $\rho = 0.5$ and $A_{1,(jj)} = 0.5$.

Another relevant case we deal with is for $p = 1$ and $A_{1,(j2)} = 0.2, 0.5, 1, j = 1, 2$. This process has a causality which decreases as the frequency increases (Case 3). If $A_{1,(j2)}$ is at least 0.5, $T = 100$ is enough to reject the null hypothesis at all frequencies with probability 1. If $A_{1,(j2)} = 0.2$, $T = 300$ is required to reject the null at all frequencies, as for $T = 100$ the power is far from 1 and stands around 60% at the lowest frequency. At the same time, we observe in Figs. 4 and 5 that BC presents remarkable power losses at all frequencies even for $T = 300$, both for the unconditional and the conditional case.

Moreover, we consider the VAR models described in Breitung and Candelon (2006), Sect. 4. Those models have $p = 3$, $A_{k,(j2)} = 1$, $k = 1, 3$, $j = 1$ and $A_{k,(j2)} = -2\cos(\omega^*)$, $k = 2$, $j = 1$. Such coefficient structure results in a null causality at frequency ω^*. In addition, we test the sensitivity of the results to the condition number of the covariance matrix, setting $\Sigma_2 = diag(1, 1)$, $\Sigma_2 = diag(1, 0.2)$, $\Sigma_2 = diag(1, 5)$.

If $\omega^* = \frac{\pi}{2}$ and $\Sigma_2 = diag(1, 1)$ (Fig. 6), we observe that the power of our test is close to the BC one. This happens because Σ_2 has condition number 1, i.e. the process shows no collinearity. If $\Sigma_2 = diag(1, 0.2)$ (Fig. 7), BC test has a level approximately equal to 5% at ω^*, while ours is larger, because the two series are both dependent in distribution and collinear. On the contrary, if $\Sigma_2 = diag(1, 5)$ (Fig. 8), even BC test loses power at ω^* because both the magnitude and the variance of the estimated GC increase considerably. Setting a different ω^*, like $\frac{3}{4}\pi$, the exposed patterns remain the same (Fig. 9).

Setting instead $A_{k,(22)}, k = 1, 3$, to 0.25 and 0.5 equals to increase the magnitude of the VAR roots until the limit value of 1 (non-stationary case). In that case, consistently with our null hypothesis, we observe that the rejection rate of our test increases as $A_{k,(22)}$ increases (Figs. 10, 11).

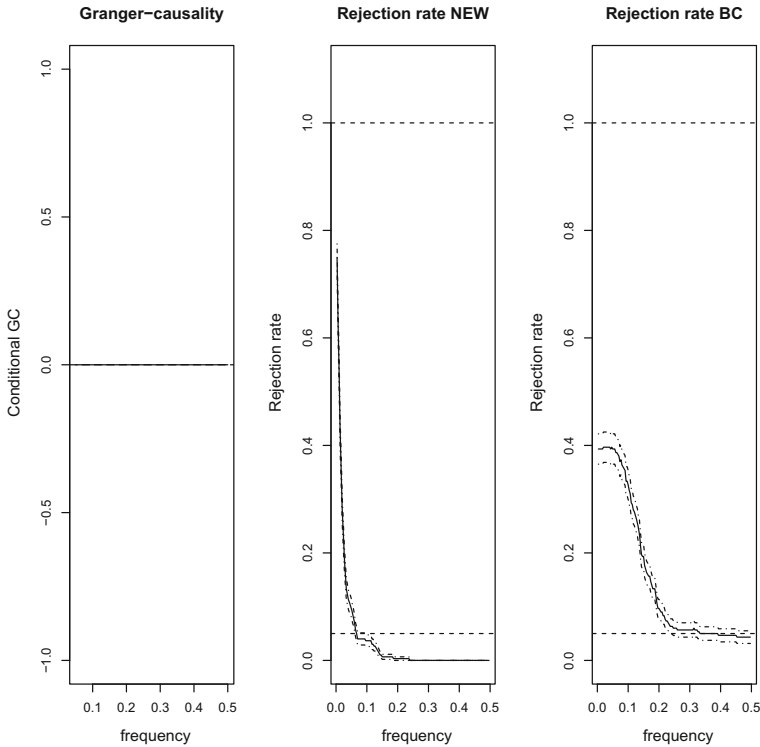To sum up, the rejection rate of our test depends on three factors:
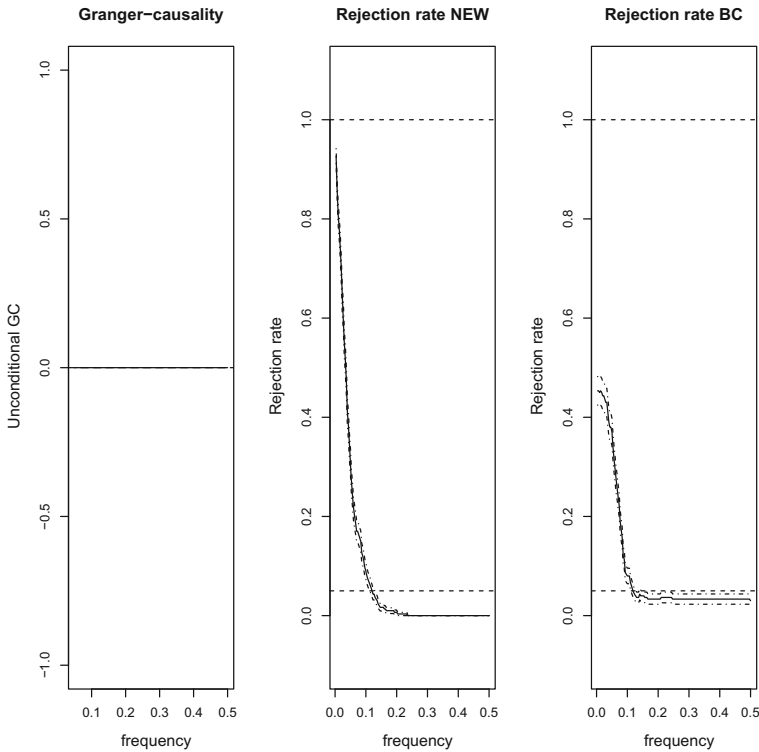
Fig. 1 Zero unconditional GC with $A_{1,(jj)} = 1$, $j = 1, 2$ (Case 2), $T = 300$. In solid line, GC shape is reported at left, rejection rates for our NEW test and BC test are reported at center and at right. One standard deviation bands are reported in dotdash line. The values $\alpha = 0.05$ and 1 are reported in dashed line

– the magnitude of VAR roots: the rejection rate is perturbed at low frequencies as the process is closer to non-stationarity;
– the true underlying spectral variability, which in turn depends on the relationship between the magnitude of causality and non-causality coefficients;
– the condition number of the autocovariance matrices R_j, $j \geq 0$, which impacts on the degree of stochastic dependence.

We also compute the rejection rates of the conservative test on all causalities jointly considered obtained by Bonferroni correction. The results for our test are reported in Table 1, the results for BC test are reported in Table 2. We note that our test is more powerful as soon as the underlying setting goes towards stochastic dependence.

Table 3 reports the mean computational time (with the relative standard deviation) of our inferential procedures (both for unconditional and conditional GC) for different values of the time series length T and the number of bootstrap samples N under the null hypothesis. We can see that the computational cost increases with the number of Fourier frequencies (equal to $T/2$). Considering that $N = 300$ is enough to obtain a sufficient approximation of critical values, our approach appears to be perfectly
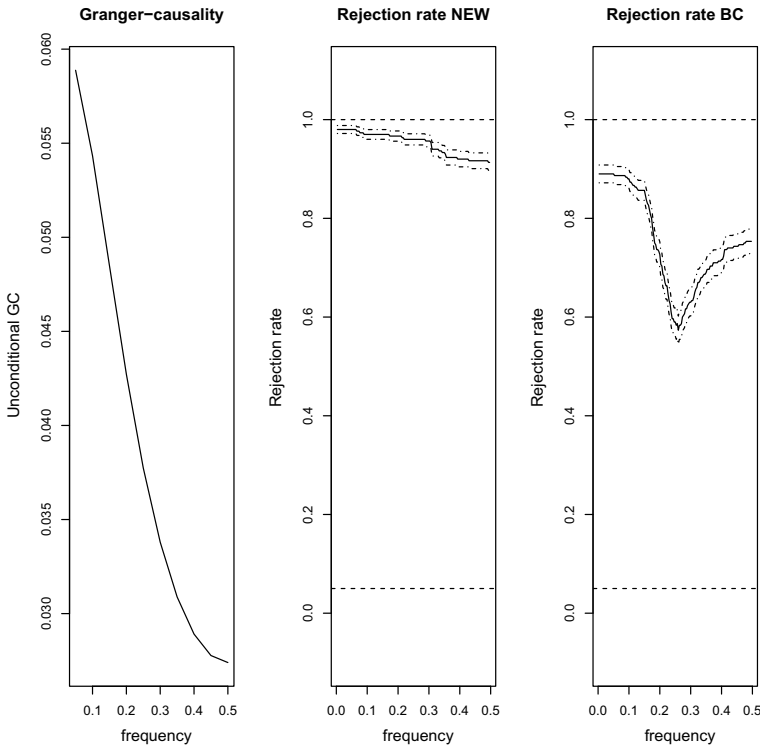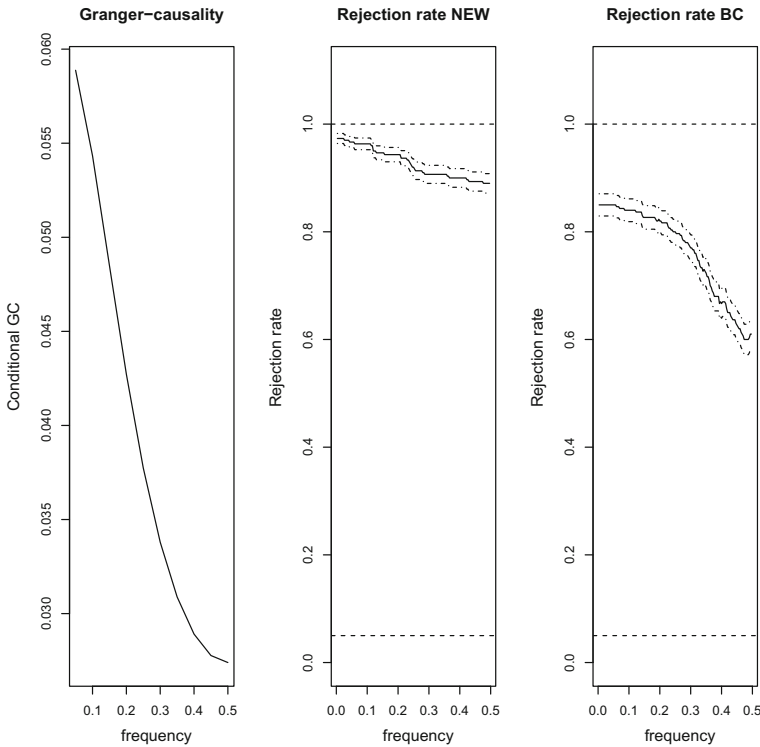
Fig. 2 Zero conditional GC with $A_{1,(jj)} = 1$, $j = 1, 2$ (Case 2), $T = 300$. In solid line, GC shape is reported at left, rejection rates for our NEW test and BC test are reported at center and at right. One standard deviation bands are reported in dotdash line. The values $\alpha = 0.05$ and 1 are reported in dashed line
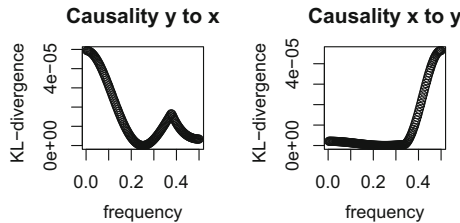
Fig. 3 An example of Kullback–Leibler divergence between the density of each individual causality and the median causality estimated via a Gaussian kernel with bandwidth equal to 0.25 for Case 1 with $A_{1,(jj)} = 0$, $j = 1, 2$, $T = 300$ and $N = 300$

feasible. For sets of very large series, provided that they are covariance stationary, we note that an appropriate subsample of Fourier frequencies can be easily drawn to keep computational times low.
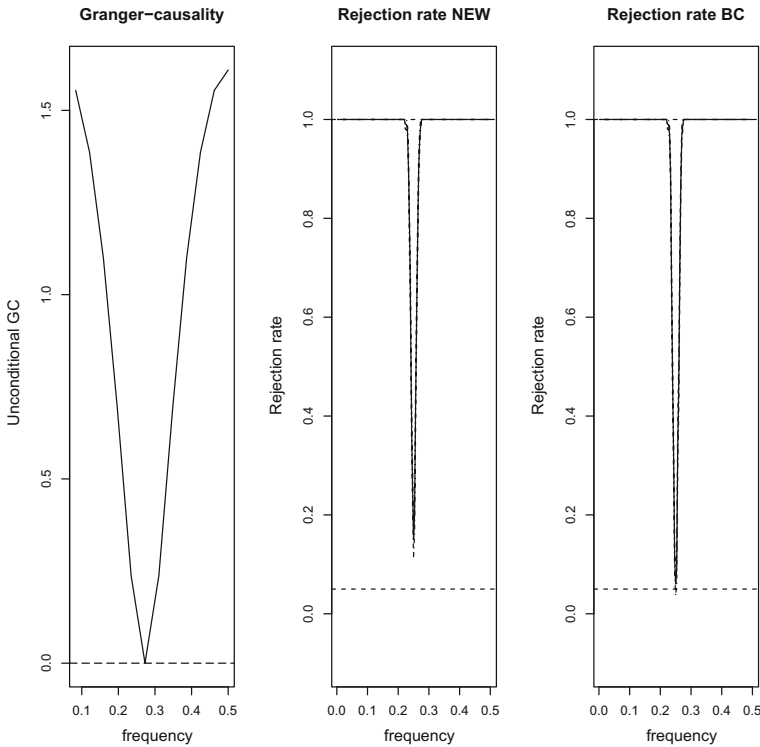
Fig. 4 Decreasing unconditional GC with $A_{1,(j2)} = 0.2$, $j = 1, 2$ (Case 3), $T = 300$. In solid line, GC shape is reported at left, rejection rates for our NEW test and BC test are reported at center and at right. One standard deviation bands are reported in dotdash line. The values $\alpha = 0.05$ and 1 are reported in dashed line

Table 1 Rejection rates of our test on all causalities jointly considered obtained by Bonferroni correction

Case	Unconditional GC	Conditional GC
Case 1 with $\Sigma_2 = diag(1, 1)$	0	0.01
Case 1 with $\rho = 0.5$	0.01	0.01
Case 2 with $A_{1,(jj)} = 0.5$	0.15	0.16
Case 2 with $A_{1,(jj)} = 1$	0.72	0.67
Case 3 with $A_{1,(j2)} = 0.2$ and $T = 100$	0.35	0.37
Case 3 with $A_{1,(j2)} = 0.2$ and $T = 300$	0.82	0.83

3 A Granger-Causality Analysis of Euro Area GDP, M3 and M1 in the Frequency Domain

In this section we study the co-movements of gross domestic product (GDP) and money stock (M3 and M1 aggregate) in the Euro Area. Differently from the Federal Reserve, the European Central Bank still considers the M3 growth rate as a policy

Fig. 5 Decreasing conditional GC with $A_{1,(j2)} = 0.2$, $j = 1, 2$ (Case 3), $T = 300$. In solid line, GC shape is reported at left, rejection rates for our NEW test and BC test are reported at center and at right. One standard deviation bands are reported in dotdash line. The values $\alpha = 0.05$ and 1 are reported in dashed line

Table 2 Rejection rates of BC test on all causalities jointly considered obtained by Bonferroni correction

Case	Unconditional GC	Conditional GC
Case 1 with $\Sigma_2 = diag(1, 1)$	0.02	0.01
Case 1 with $\rho = 0.5$	0.05	0.04
Case 2 with $A_{1,(jj)} = 0.5$	0.01	0.01
Case 2 with $A_{1,(jj)} = 1$	0.32	0.29
Case 3 with $A_{1,(j2)} = 0.2$ and $T = 100$	0.22	0.21
Case 3 with $A_{1,(j2)} = 0.2$ and $T = 300$	0.68	0.63

target (see Jung and Villanova 2020). We have thus considered M3 instead of M2, also because they are quite close in the Euro Area (see Darvas 2015). We test in the frequency domain both the *direct* link from one variable to the other one and the *indirect* link with respect to further explanatory variables like the inflation rate (HICP), the unemployment rate (UN), or the long-term interest rate (LTN). Published works on this research topic make use of time-domain methods: some of them use

Fig. 6 Case with $p = 3$, $A_{k,(j2)} = 1$, $k = 1, 3$, $j = 1$, $A_{2,(12)} = 0$, $\Sigma_2 = diag(1, 1)$, and $T = 300$. In solid line, GC shape is reported at left, rejection rates for our NEW test and BC test are reported at center and at right. One standard deviation bands are reported in dotdash line. The values $\alpha = 0.05$ and 1 are reported in dashed line

Table 3 Mean computational times in seconds of our NEW bootstrap testing approaches with the relative standard deviations under the null hypothesis for different values of the time series length T and the number of bootstrap samples N across 10 trials. Calculations are performed by a 64-bit Intel Core i7-8665U processor with 1.90 GHz base frequency and 16GB RAM	T	N	Unconditional GC		Conditional GC	
			Mean	STD	Mean	STD
	100	100	1.98	0.13	4.95	0.92
	100	300	6.06	0.30	13.23	0.42
	100	500	10.05	0.36	19.14	0.29
	300	100	3.27	0.27	4.81	0.30
	300	300	8.66	0.08	13.57	0.10
	300	500	15.29	1.48	23.38	1.32
	500	100	3.89	0.23	6.01	0.21
	500	300	11.86	0.49	17.33	0.14
	500	500	12.00	0.37	28.43	0.10

factor modelling (Cendejas et al. 2014), some others use likelihood methods (Andrés et al. 2006, Canova and Menz 2011), or large-dimensional VAR models (Giannone et al. 2013), or dynamic factor models (Breitung and Eickmeier 2006) or VAR models
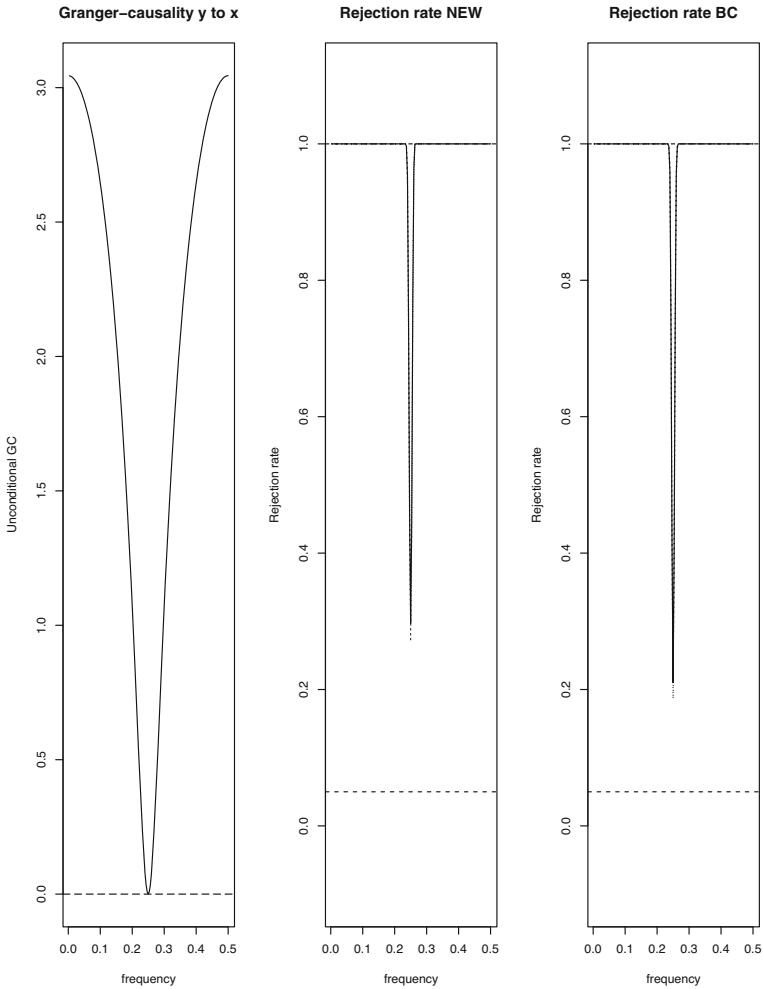
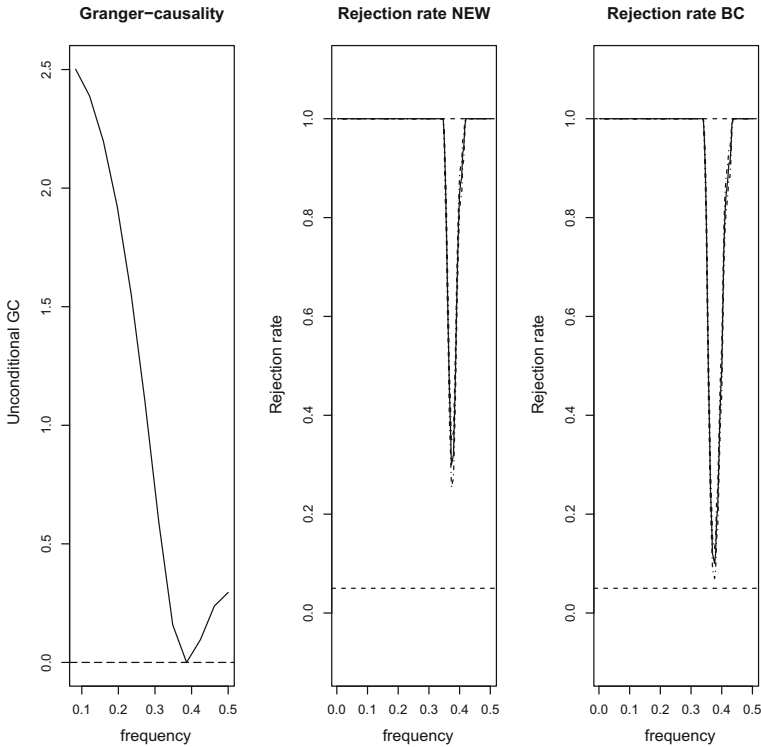**Fig. 7** Case with $p = 3$, $A_{k,(j2)} = 1$, $k = 1, 3$, $j = 1$ and $\omega^* = \frac{\pi}{2}$, $A_{2,(12)} = 0$, $\Sigma_2 = diag(1, 5)$, and $T = 300$. In solid line, GC shape is reported at left, rejection rates for our NEW test and BC test are reported at center and at right. One standard deviation bands are reported in dotdash line. The values $\alpha = 0.05$ and 1 are reported in dashed line

with time-varying parameters (Psaradakis et al. 2005). On the contrary, we apply the inferential framework for GC in the frequency domain developed in Sect. 2. In this way, we provide explicit inference on unconditional and conditional GC. HICP, UN and LTN are used as conditioning variables, with the aim to discount for the mediating power of each of the three variables with respect to the relationship between output and money supply.
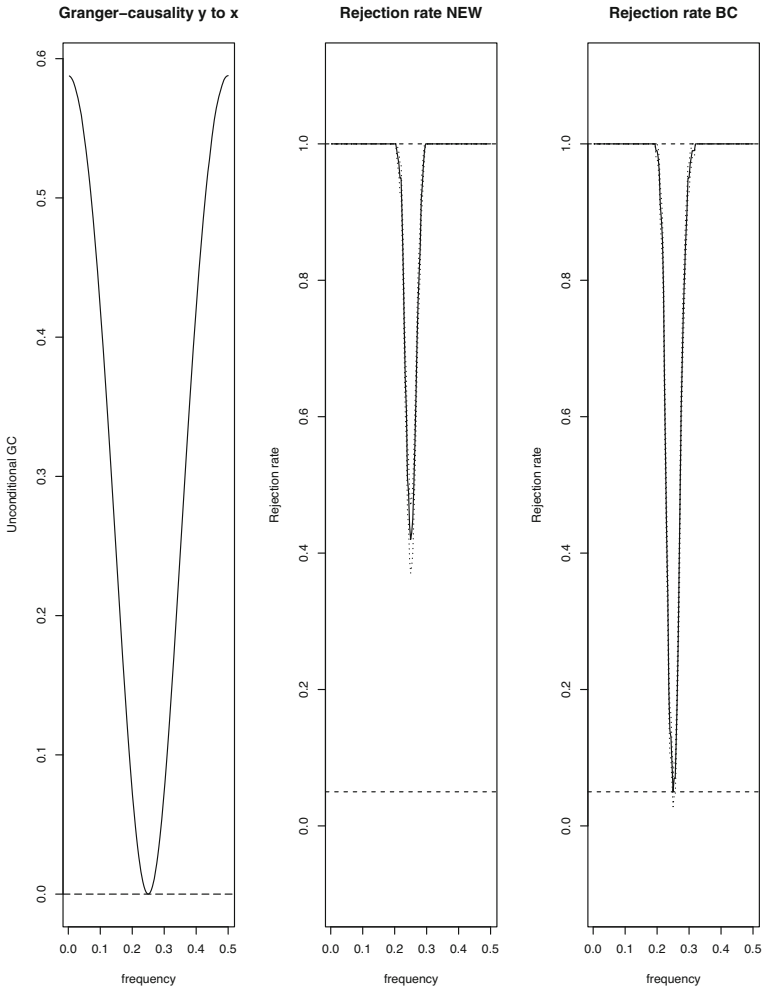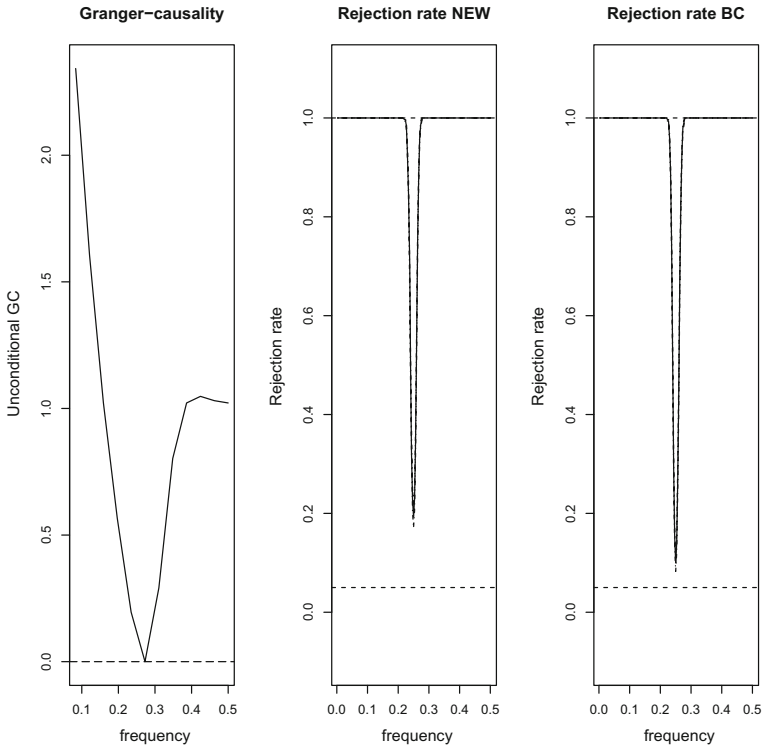
Fig. 8 Case with $\omega^* = \frac{\pi}{2}$, $A_{2,(12)} = 0$, $\Sigma_2 = diag(1, 0.2)$, and $T = 300$. In solid line, GC shape is reported at left, rejection rates for our NEW test and BC test are reported at center and at right. One standard deviation bands are reported in dotdash line. The values $\alpha = 0.05$ and 1 are reported in dashed line

3.1 Data Preparation

We have considered the time series of GDP at market price in the Euro Area (chain linked volumes in Euro) and the monetary aggregate M3 and M1 (outstanding amount of loans to the whole economy excluded the monetary and financial sector, all currencies combined). M3 is also called "broad money", M1 "narrow money".

Since our goal is to focus on the effect of monetary policy on output, we restrict our analysis to the period 1999–2017, when the ECB has taken actual decisions on the Euro Area. We can thus denote our series by GDP_t, $M3_t$, $M1_t$, $HICP_t$, UN_t, LTN_t, where $t = 1, \ldots, 76$ (there are 76 quarters from Winter 1999 to Autumn

Fig. 9 Case with $\omega^* = \frac{3}{4}\pi$, $A_{2,(12)} = 0$, $\Sigma_2 = diag(1,1)$, and $T = 300$. In solid line, GC shape is reported at left, rejection rates for our NEW test and BC test are reported at center and at right. One standard deviation bands are reported in dotdash line. The values $\alpha = 0.05$ and 1 are reported in dashed line

2017). The data are drawn from the ECB Real Time Research database where national figures are aggregated according to a changing composition of the Euro Area across time (see Giannone et al. 2012). We consider quarterly series as the GDP is regularly published on a quarterly base. The other series (all but GDP) are therefore made quarterly by averaging. We refer to "https://www.ecb.europa.eu/stats" and ECB (2012) for technical and computational details.

According to the Dickey-Fuller test, the logarithmic transforms of GDP_t, $M3_t$, $M1_t$ are non-stationary, as well as the three conditioning variables $HICP_t$, UN_t and LTN_t. Therefore, following Friedman and Schwartz (1975), we pass all series by Hodrick-Prescott filter (Hodrick and Prescott 1997), with the canonical value of $\lambda = 1600$, in order to remove any trend and to extract cyclical components. We do not use Baxter-King filter (Baxter and King 1999), as suggested in Belongia and Ireland (2016), because we have not enough end of sample data. Cycle extraction is performed via the R package "mFilter".

Figures 12 and 13 contain the plots of GDP_t, $M3_t$, $M1_t$, $HICP_t$, UN_t, LTN_t respectively. Left figures contain the original series and the estimated trend, while right figures contain the estimated cycles. The patterns of the estimated ACF and CCF are
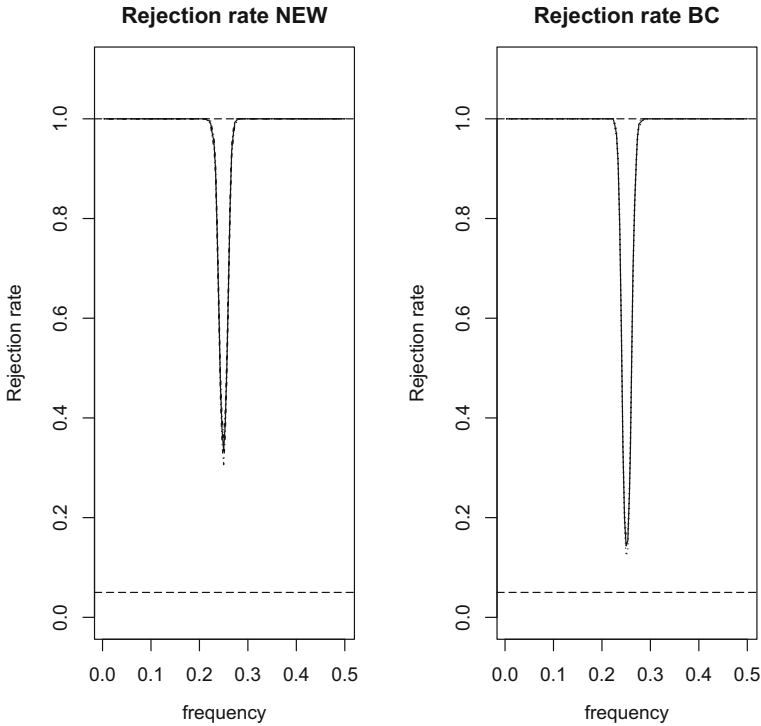
Fig. 10 Case with $\omega^* = \frac{\pi}{2}$, $A_{p,(22)} = 0.25$, $p = 1, 3$, and $T = 300$. In solid line, GC shape is reported at left, rejection rates for our NEW test and BC test are reported at center and at right. One standard deviation bands are reported in dotdash line. The values $\alpha = 0.05$ and 1 are reported in dashed line

reported and commented in "Appendix A". Since our ultimate goal is to infer about the cause-effect relationship of money stock and economic output, we test at each frequency Granger-causality spectra via our approach. Due to the use of Fast Fourier Transform, the frequencies used are the following: $f_i = \frac{i}{80}, i = 1, \ldots, 40$, because $T = 76$. The frequency range is re-scaled to [0, 2] for the quarterly frequency of our series.

Relevant VAR models, estimated including an intercept by the R package "vars" (Pfaff et al. 2008), are selected by the Bayesian Information Criterion (BIC), imposing a maximum of four lags. BIC is used because we know that BIC is correctly estimating the unknown number of delays, while AIC may overestimate it, thus increasing the probability to estimate non-stationary VAR models. In any case, the roots of all estimated characteristic polynomials are strictly smaller than one. The number of bootstrap samples is 1000.

Note that for computational reasons BC test cannot be computed for $p = 1$. Besides, its p-value is constant across frequencies (except for $\omega = \{0, \pi\}$) for $p = 2$. BC test requires a large number of delays, while ours works for all values, given that the resulting VAR is stationary and non-singular. Therefore, we can not compare directly our test to BC test on real data, because BC is not useful for all cases with $p \leq 2$.

Fig. 11 Case with $\omega^* = \frac{\pi}{2}$, $A_{p,(22)} = 0.5$, $p = 1, 3$. In solid line, the rejection rates for our NEW test and BC test are reported at left and at right respectively. One standard deviation bands are reported in dotdash line. The values $\alpha = 0.05$ and 1 are reported in dashed line. Here we omit the GC shape because the process is non-stationary

3.2 Causality Results

We start describing VAR estimates on the couple GDP-M3. Our lag selection procedure chooses 2 lags. In the GDP_t equation, GDP_{t-1} and GDP_{t-2} are heavily significant, while $M3_{t-1}$ and $M3_{t-2}$ slightly are (at 5% and 10% respectively). This results in a GC spectral shape which is approximately constant and slightly significant across frequencies. In the $M3_t$ equation, $M3_{t-1}$ is heavily significant, while GDP_{t-2} is at 10%. The corresponding GC shape is significant in the first half of the frequency range.

Concerning the couple GDP-M1, our VAR lag selection procedure chooses 2 lags. In the GDP_t equation, GDP_{t-1}, GDP_{t-2} and $M1_{t-2}$ are heavily significant. The related unconditional GC shape is significant across the entire frequency range. In the $M1_t$ equation, only $M1_{t-1}$ is heavily significant, while GDP_{t-1} has a p-value of 12%. The resulting GC spectral shape is thus significant in the first third of the frequency range.

In Figs. 14 and 15, unconditional and conditional GC spectra from M3 to GDP and viceversa are reported. The same spectra from M1 to GDP and viceversa are in Figs. 16 and 17 respectively. In dashed, our bootstrap threshold at 5% is outlined. In
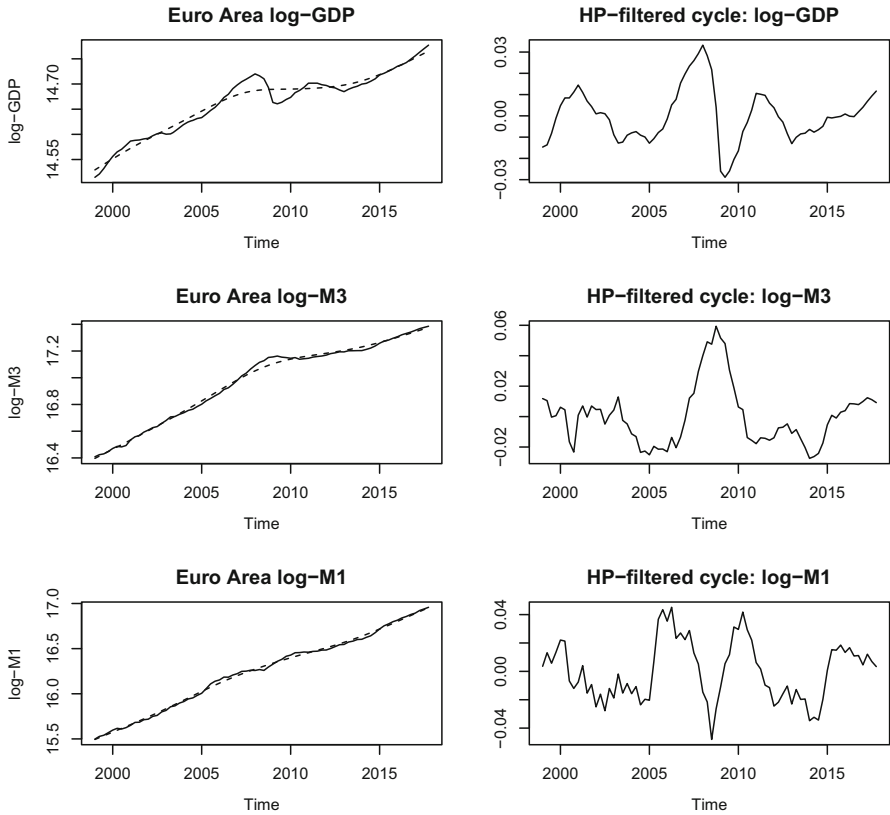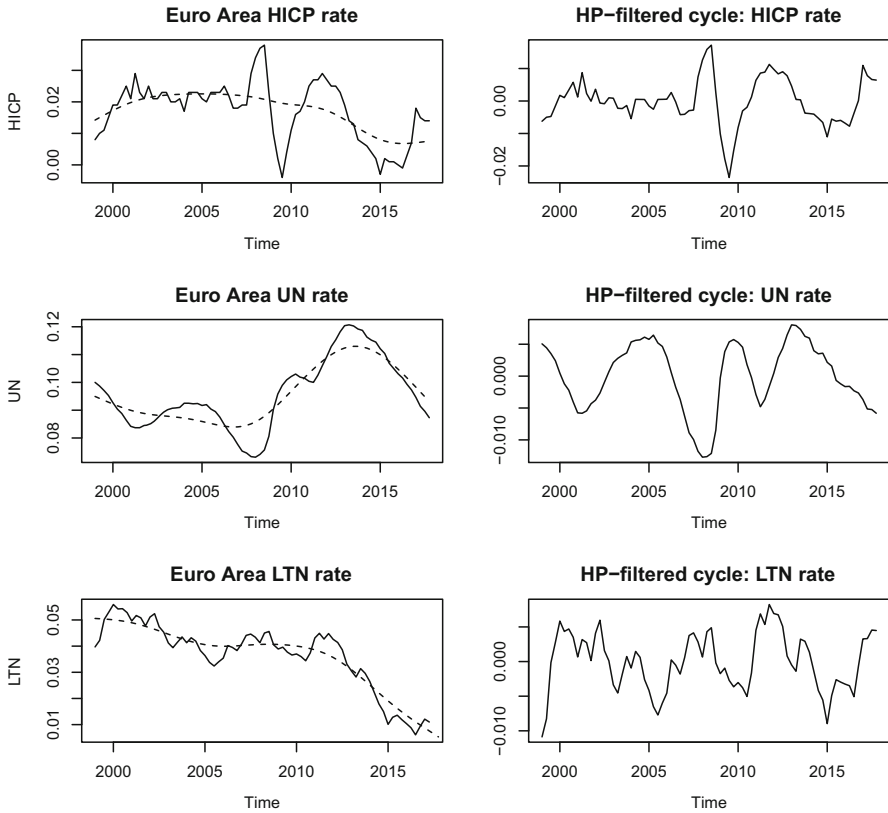
Fig. 12 Left figures contain quarterly Euro Area GDP, M3 and M1 in logs across the period 1999–2017. In dashed the estimated trend by Hodrick-Prescott filter ($\lambda = 1600$). Right figures contain the estimated cycle by the same method

dotdash, the same threshold for the overall test obtained by Bonferroni correction is depicted.

We first comment conditional GC spectra for the couple GDP-M3. Conditioning on HICP, the level of significance of $M3_{t-1}$ and $M3_{t-2}$ increases in the GDP_t equation. This results in a GC decreasing across frequencies and significant across the entire frequency range. In the $M3_t$ equation, the level of significance of GDP_{t-2} increases to 5%. As a result, GC is significant until the period of 1 year. Conditioning on UN, in the GDP_t equation the level of significance of $M3_{t-1}$ is 5% while $M3_{t-2}$ is no longer significant. This results in a GC which is no longer significant. In the $M3_t$ equation, GDP_{t-2} is no longer significant, resulting in a non-significant GC everywhere. Conditioning on LTN, the level of significance is 5% for $M3_{t-1}$ and 10% for $M3_{t-2}$ in the GDP_t equation. The corresponding GC is significant across the entire frequency range. In the $M3_t$ equation, GDP_{t-2} is significant at 5%, causing again GC to be significant almost everywhere.

We now comment conditional GC spectra for the couple GDP-M1. Conditioning on HICP, in the GDP_t equation $M1_{t-2}$ is still heavily significant. The GC spectral

Fig. 13 Left figures contain quarterly Euro Area HICP, UN, LTN rates across the period 1999–2017. In dashed the estimated trend by Hodrick-Prescott filter ($\lambda = 1600$). Right figures contain the estimated cycle by the same method

shape is almost the same as the unconditional one. In the $M1_t$ equation, the level of significance is quite smaller, so that causalities are significant only in the first quarter of the frequency range. Conditioning on UN, $M1_{t-2}$ is still heavily significant in the GDP_t equation. The conditional GC magnitude is slightly weaker than the corresponding unconditional one. In the $M1_t$ equation, GDP_{t-1} has a p-value of 20% and the related GC shape is significant in the first half of the frequency range. Conditioning on LTN, $M1_{t-2}$ is still significant at 1% in the GDP_t equation, causing GC shape to be significant over all frequencies. In the $M1_t$ equation, GDP_{t-1} has a p-value of 26%. As a consequence, we observe significance at low frequencies only.

Concerning the overall test on all causalities, we observe the absence of any significance in four cases out of sixteen: the unconditional GC from M3 to GDP, the conditional GC from M3 to GDP both on UN and LTN, and the conditional GC from GDP to M3 on UN. We remark that this test is conservative in nature: however, it allows to adequately contextualize the significance of individual tests. In particular, we can claim that, for those four cases, the GC spectrum is indistinguishable at all

M3 to GDP

M3 to GDP on HICP
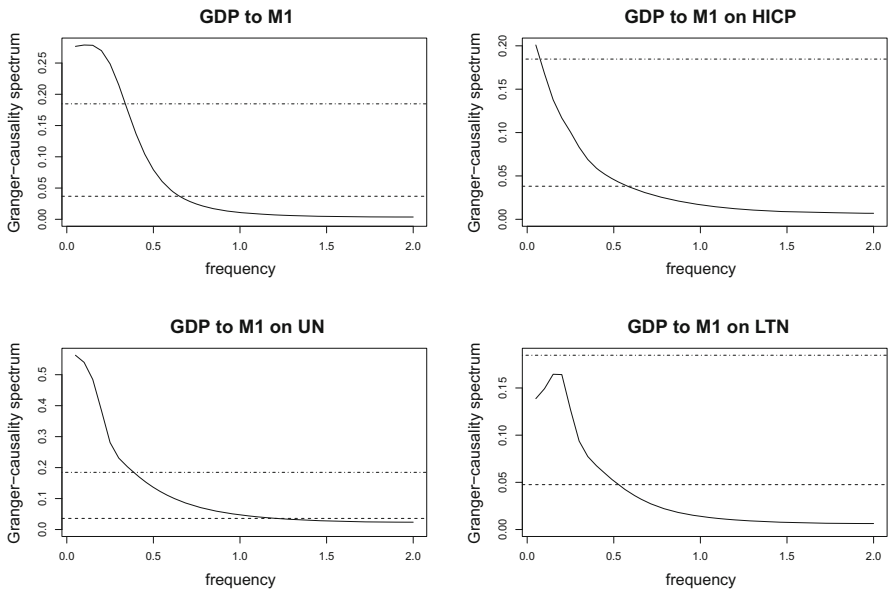
M3 to GDP on UN

M3 to GDP on LTN
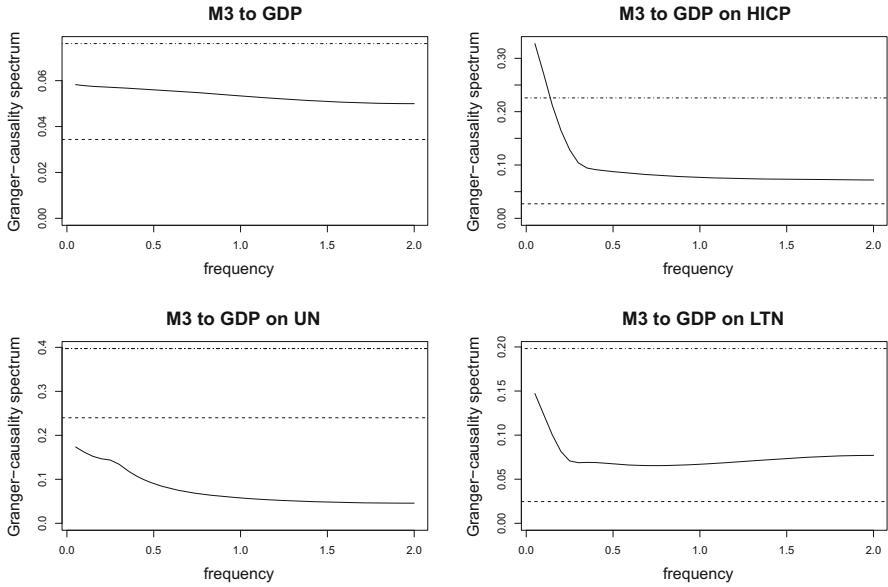
Fig. 14 GC spectra from M3 to GDP: unconditional, conditional on HICP, UN and LTN respectively. In dashed, our bootstrap threshold at 5% is outlined. In dotdash, the same threshold for the overall test obtained by Bonferroni correction is depicted

GDP to M3

GDP to M3 on HICP

GDP to M3 on UN

GDP to M3 on LTN
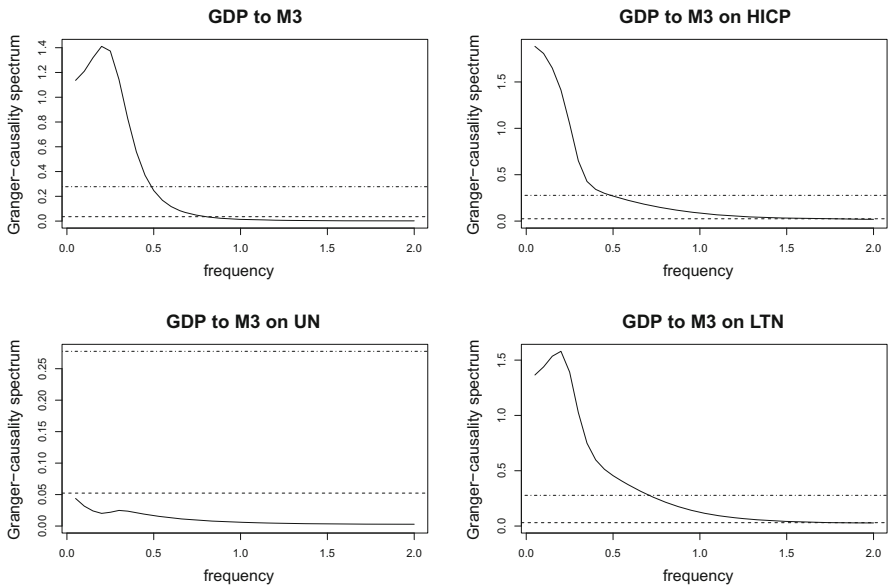
Fig. 15 GC spectra GDP to M3: unconditional, conditional on HICP, UN and LTN respectively. In dashed, our bootstrap threshold at 5% is outlined. In dotdash, the same threshold for the overall test obtained by Bonferroni correction is depicted

M1 to GDP

M1 to GDP on HICP

M1 to GDP on UN

M1 to GDP on LTN
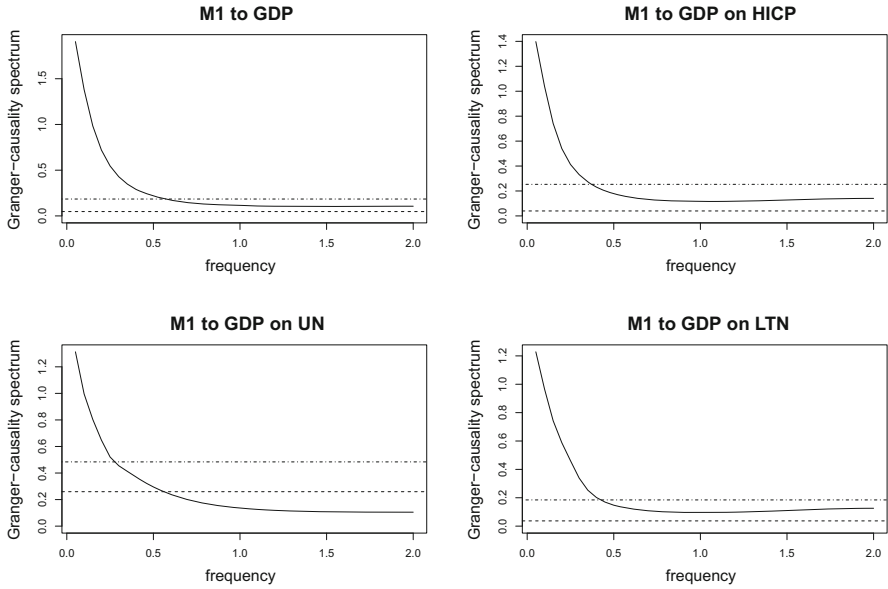
Fig. 16 GC spectra M1 to GDP: unconditional, conditional on HICP, UN and LTN respectively. In dashed, our bootstrap threshold at 5% is outlined. In dotdash, the same threshold for the overall test obtained by Bonferroni correction is depicted

GDP to M1

GDP to M1 on HICP

GDP to M1 on UN

GDP to M1 on LTN

Fig. 17 GC spectra GDP to M1: unconditional, conditional on HICP, UN and LTN respectively. In dashed, our bootstrap threshold at 5% is outlined. In dotdash, the same threshold for the overall test obtained by Bonferroni correction is depicted

frequencies with a significance level of at most 5% from the one we would observe in case of stochastic independence between output and money stock.

4 Conclusions

In this paper, we have developed a bootstrap testing procedure for Granger-causality (GC) in the frequency domain. In particular, our test is based on the null hypothesis of stochastic independence, which embeds the traditional one of no Granger-causality. Via our test, we highlight those frequencies which present a Granger-causality systematically larger than the median causality across frequencies observed under the null hypothesis. Such a goal is reached by appropriately exploiting the stationary bootstrap procedure of Politis and Romano (1994). We both deal with unconditional GC and conditional GC with respect to third variables. A simulation study has shown that our procedure is complementary with respect to the test of Breitung and Candelon (2006).

Our test has a general validity, as it only requires the stationary bootstrap of Politis and Romano (1994) to be consistent on the data generating process of interest, which requires $T^{\frac{1}{3}} \to \infty$, where T is the time series length. Therefore, our procedure may find application in several fields like macroeconomics, neuroscience, meteorology, seismology and finance. Among those, monetary economics is a very suitable application field, as the time series of interest often present a rich causality structure, and the need rises to disambiguate among relevant frequencies the significant ones for causality.

For this reason, we have applied our methodology to Euro Area money stock and GDP. From an empirical point of view, we have been able to say that the relationship between money supply and output is present in the Euro Area across the period 1999–2017. We have provided evidence that M3 (M1) in some cases reacts to economic shocks, in some others it acts as a policy shock with respect to economic output. We have observed that the link between GDP and M1 is much stronger in both directions than the link between GDP and M3. In particular, the causal relationship from M1 to GDP appears to be significant at all frequencies, while the opposite one is significant at low frequencies only. This allows us to conclude that the intensity of the causal link from money to output appears to be stronger than the reverse one in the Euro Area, reinforcing the idea, outlined among others in Musso et al. (2019), that M1 has a strong predictive power for GDP in the Euro Area.

Declarations

A Auto and Cross Covariance Functions

Figures 18 and 19 show the ACF of the estimated cycles by Hodrick-Prescott filter ($\lambda = 1600$) for the series of Euro Area GDP, M3, M1, HICP, UN and LTN across the period 1999–2017. The patterns are very similar across series: positive for the first 4-5 quarters, negative for all quarters around 2 years and non-significant elsewhere. UN shows a rebound for the quarters around 5 years. LTN is no longer significant after 2 quarters. Figure 20 shows the CCF for the couples GDP-M3 and GDP-M1. Their pattern is similar: we have positive correlation around 0 and negative correlation at sides around the lag of 2 years.
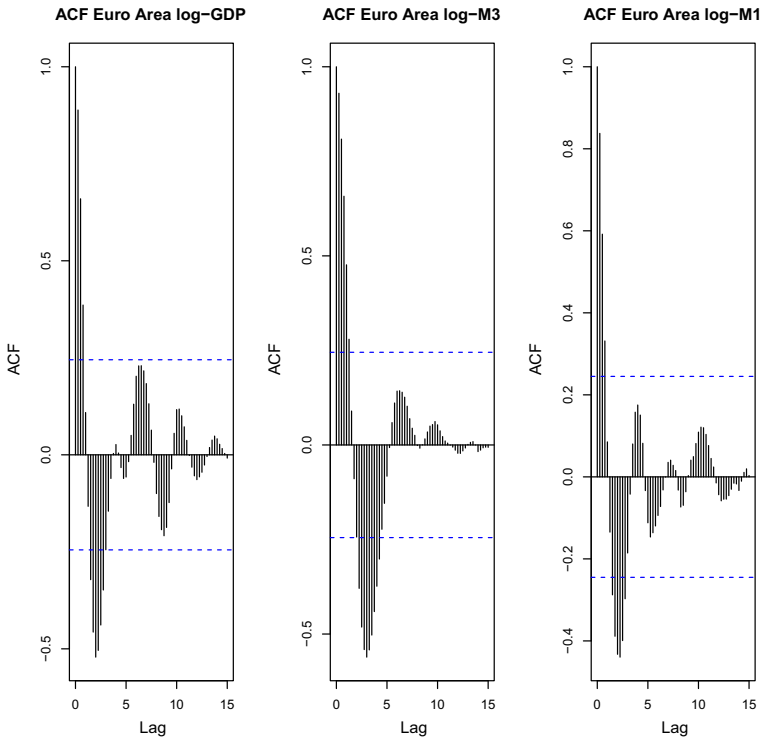
Fig. 18 Quarterly ACF patterns of the extracted cycles by Hodrick–Prescott filter ($\lambda = 1600$) of Euro Area GDP, M3 and M1 in logs across the period 1999–2017

B Proofs

B.1 Proof of Theorem 1

Let us define the stochastic process $Z_t^{\dagger} = [X_t^{\dagger}, Y_t^{\dagger}]$. We assume that $X_t^{\dagger}$, $Y_t^{\dagger}$ are stochastically independent, which causes the null hypothesis H_0 to hold. This is like assuming that the distribution function $F_{Z^{\dagger}}$ can be factorized as $F_{X^{\dagger}} F_{Y^{\dagger}}$. In addition, we assume that $X_t^{\dagger}$ and $Y_t^{\dagger}$ are strictly stationary.

By Politis and Romano (1994) (paragraph 4.3) we know that

$$\sqrt{T}(r(\hat{F}_{X^{\dagger}}) - r(F_{X^{\dagger}})) = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} g_F(X_i) + o(\sqrt{T}||\hat{F}_{X^{\dagger}} - F_{X^{\dagger}}||), \qquad (5)$$

where $\hat{F}_{X^{\dagger}}$ is the empirical density function of $X_t^{\dagger}$ and $F_{X^{\dagger}}$ is the corresponding true distribution function. The same equation holds for $Y^{\dagger}$.

If, for some $d \geq 0$, $E(g_{F_{X^{\dagger}}}(X_1^{\dagger}))^{2+d} < \infty$, and if it holds $\sum_{k=1}^{\infty} \alpha_{X^{\dagger}}(k)^{\frac{d}{2+d}} < \infty$, then $\frac{1}{\sqrt{T}} \sum_{i=1}^{T} g_{F_{X^{\dagger}}}(X_i^{\dagger})$ is asymptotically normal with mean 0 and variance
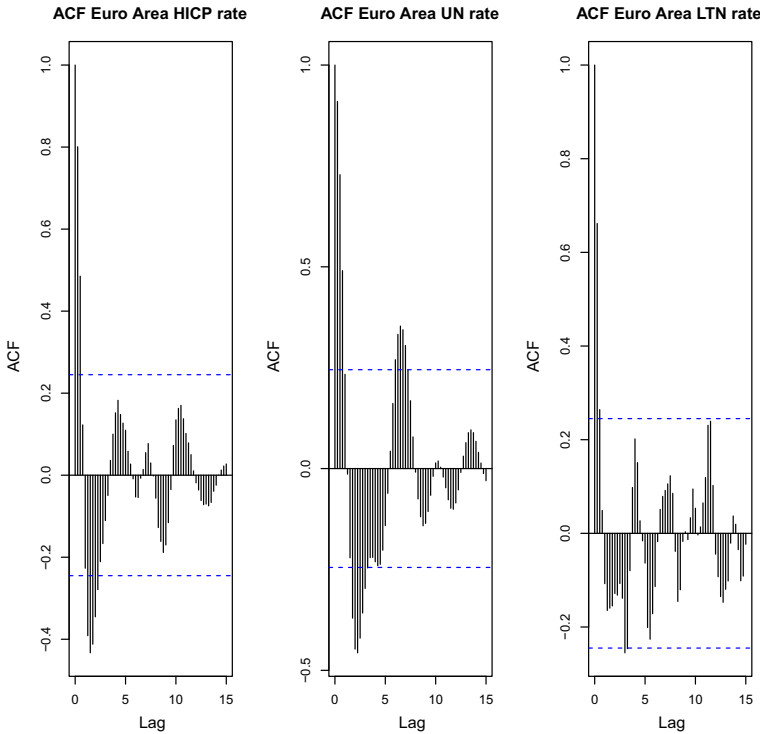
Fig. 19 Quarterly ACF patterns of the extracted cycles by Hodrick-Prescott filter ($\lambda = 1600$) of Euro Area HICP, UN and LTN rates across the period 1999–2017

$$E(g_{F_{X^\dagger}}(X_i^\dagger)^2) + 2\sum_{k=1}^{\infty} cov(g_{F_{X^\dagger}}(X_1^\dagger), g_{F_{X^\dagger}}(X_{1+k}^\dagger)). \tag{6}$$

The same equation holds for $Y^\dagger$ if $E(g_{F_{Y^\dagger}}(Y_1^\dagger))^{2+d} < \infty$ and $\alpha_{Y^\dagger}(k)^{\frac{d}{2+d}} < \infty$ respectively.

At the same time, if $\sum_{k=1}^{\infty} k^2\alpha_{X^\dagger}(k)^{1/2-\tau} < \infty$, $\sum_{k=1}^{\infty} k^2\alpha_{Y^\dagger}(k)^{1/2-\tau} < \infty$ for some $0 < \tau < 1/2$, the stochastic processes $\hat{F}_{X^\dagger} - F_{X^\dagger}$, $\hat{F}_{Y^\dagger} - F_{Y^\dagger}$ converge in supremum norm to a Gaussian process having continuous paths and mean 0. Therefore, $\sqrt{T}(r(\hat{F}_{X^\dagger}) - r(F_{X^\dagger}))$ is asymptotically normal with mean 0 and variance (6). The same results hold for $r(\hat{F}_{Y^\dagger})$.

Moreover, the distribution of $\sqrt{T}(r(\hat{F}_{X^\dagger}) - r(F_{X^\dagger}))$ is approximated via the distribution of $\sqrt{T}(r(F_{X^\dagger}^*) - r(\hat{F}_{X^\dagger}))$, where $F_{X^\dagger}^*$ is the empirical density function obtained via stationary bootstrap. This holds because the two distributions converge to the same Gaussian process under previous weak dependence assumptions, provided that $T^{\frac{1}{3}} \to \infty$. The same result holds considering $r(\hat{F}_{Y^\dagger})$.

At this point, since we assumed the stochastic independence of $X_t^\dagger$ and $Y_t^\dagger$, the weak dependence assumptions on $E(g_F)$, α_X and $\sum_{k=1}^{\infty} \alpha(k)^{\frac{d}{2+d}}$ are transmitted to
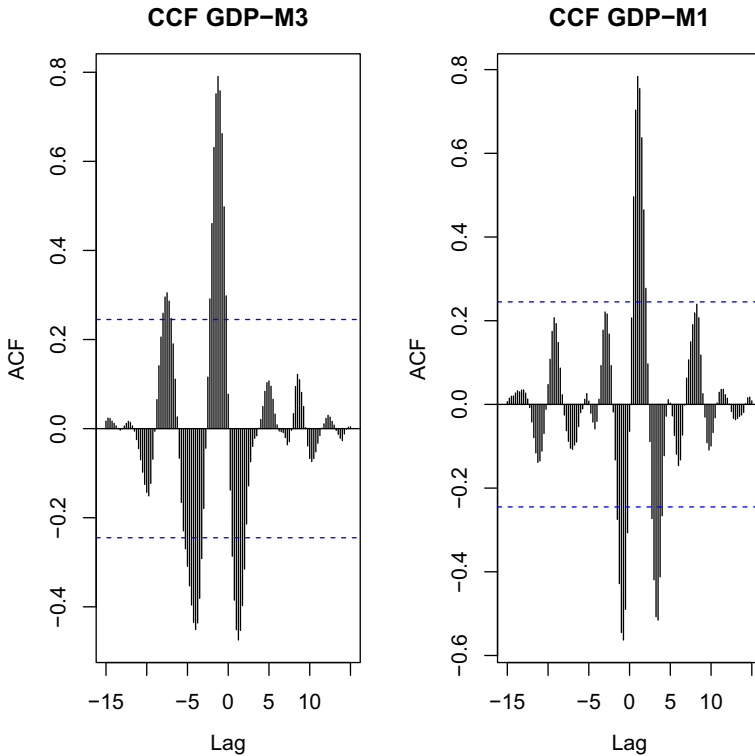
Fig. 20 Quarterly CCF patterns of the extracted cycles by Hodrick–Prescott filter ($\lambda = 1600$) of the couples GDP-M3 and GDP-M1 in logs measured over the Euro Area across the period 1999–2017

the whole process $Z_t^{\dagger}$. Therefore, for any *Fréchet-differentiable* functional r we can write

$$\sup_{c \in \mathcal{R}} \left| P\left\{ \sqrt{T}(r(F_{Z^{\dagger}}^{*}) - r(\hat{F}_{Z^{\dagger}})) \leq c \right\} - P\left\{ \sqrt{T}(r(\hat{F}_{Z^{\dagger}}) - r(F_{Z^{\dagger}})) \leq c \right\} \right| \xrightarrow{prob} 0$$

under the assumption $T^{\frac{1}{3}} \to \infty$.

B.2 Proof of Theorem 2

Under the assumptions of Theorem 2, the stochastic processes Z'' and $Y^{\dagger}$ are independent in distribution. Therefore, the distribution function of any statistics $r(Z'', Y^{\dagger})$ is the product of the distribution function of $r(Z'')$ and the one of $r(Y^{\dagger})$. Under the two assumptions on Z'', we can apply Theorem 3.6 in Shao and Tu (1995), which prescribes that any re-sampling bootstrap approximation of $r(Z'')$, like the one we adopt based on residual bootstrap, is consistent in weak sense as $T \to \infty$. Since in our case the median of the functional r across frequencies is stochastically indistinguishable from the functional r at any frequency, we can apply the Theorem in paragraph 4.3 of

Politis and Romano (1994) to claim that the bootstrap approximation of $F_{Y^\dagger}$ is also consistent in weak sense as $T^{\frac{1}{3}}$ diverges. From these arguments, the thesis follows. In fact, defined $Z^{'} = [Z^{''}, Y^\dagger]$, for any *Fréchet-differentiable* functional r we can write

$$\sup_{c \in \mathcal{R}} \left| P\left\{ \sqrt{T}(r(F_{Z'}^*) - r(\hat{F}_{Z'})) \le c \right\} - P\left\{ \sqrt{T}(r(\hat{F}_{Z'}) - r(F_{Z'})) \le c \right\} \right| \xrightarrow{prob} 0$$

under the assumption $T^{\frac{1}{3}} \to \infty$.

References

Andrés, J., David López-Salido, J., & Vallés, J. (2006). Money in an estimated business cycle model of the euro area. *The Economic Journal*, *116*, 457–477.

Assenmacher-Wesche, K., & Gerlach, S. (2008). Interpreting euro area inflation at high and low frequencies. *European Economic Review*, *52*, 964–986.

Barnett, L., & Seth, A. K. (2014). The MVGC multivariate granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, *223*, 50–68.

Baxter, M., & King, R. G. (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, *81*(4), 575–593.

Belongia, M. T., & Ireland, P. N. (2016). Money and output: Friedman and Schwartz revisited. *Journal of Money, Credit and Banking*, *48*(6), 1223–1266.

Berger, H., & Osterholm, P. (2011). Does money growth granger cause inflation in the euro area? Evidence from out-of-sample forecasts using Bayesian VARs. *Economic Record*, *87*, 45–60.

Berzuini, C., Dawid, P., & Bernardinelli, L. (2012). *Causality: Statistical perspectives and applications.* Wiley.

Bozoklu, S., & Yilanci, V. (2013). Energy consumption and economic growth for selected OECD countries: Further evidence from the granger causality test in the frequency domain. *Energy Policy*, *63*, 877–881.

Breitung, J., & Candelon, B. (2006). Testing for short-and long-run causality: A frequency-domain approach. *Journal of Econometrics*, *132*, 363–378.

Breitung, J., & Eickmeier, S. (2006). Dynamic factor models. *Allgemeines Statistisches Archiv*, *90*(1), 27–42.

Breitung, J., & Schreiber, S. (2018). Assessing causality and delay within a frequency band. *Econometrics and Statistics*, *6*, 57–73.

Candelon, B., & Tokpavi, S. (2016). A nonparametric test for granger causality in distribution with application to financial contagion. *Journal of Business & Economic Statistics*, *34*(2), 240–253.

Canova, F., & Menz, T. (2011). Does money matter in shaping domestic business cycles? An international investigation. *Journal of Money, Credit and Banking*, *43*, 577–607.

Cendejas, J. L., Castañeda, J. E., & Muñoz, F.-F. (2014). Business cycle, interest rate and money in the euro area: A common factor model. *Economic Modelling*, *43*, 136–141.

Chen, Y., Bressler, S. L., & Ding, M. (2006). Frequency decomposition of conditional granger causality and application to multivariate neural field potential data. *Journal of Neuroscience Methods*, *150*, 228–237.

Croux, C., & Reusens, P. (2013). Do stock prices contain predictive power for the future economic activity? A Granger causality analysis in the frequency domain. *Journal of Macroeconomics*, *35*, 93–103.

Darvas, Z. (2015). Does money matter in the euro area? Evidence from a new Divisia index. *Economics Letters*, *133*, 123–126.

Ding, M., Chen, Y., & Bressler, S.L. (2006). Granger causality: Basic theory and application to neuroscience. Handbook of time series analysis: Recent theoretical developments and applications, 437.

ECB. (2012). Manual on MFI balance sheet statistics, April 2012. Technical report.

Friedman, M. & Schwartz, A. J. (1975). Money and business cycles. In *The state of monetary economics* (pp. 32–78). NBER.

Fromentin, V., & Tadjeddine, Y. (2020). Cross-border workers and financial instability: A frequency domain causality analysis applied to the Luxembourg financial centre. *Applied Economics Letters*, *27*(4), 280–285.

Geweke, J. F. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, *77*, 304–313.

Geweke, J. F. (1984). Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association*, *79*, 907–915.

Giannone, D., Henry, J., Lalik, M., & Modugno, M. (2012). An area-wide real-time database for the euro area. *Review of Economics and Statistics*, *94*, 1000–1013.

Giannone, D., Lenza, M., & Reichlin, L. (2013). Money, credit, monetary policy and the business cycle in the euro area: What has changed since the crisis. Technical report, ECB.

Gomez-Gonzalez, J. E., Villamizar-Villegas, M., Zarate, H. M., Amador, J. S., & Gaitan-Maldonado, C. (2015). Credit and business cycles: Causal effects in the frequency domain. *Ensayos sobre Política Económica*, *33*, 176–189.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424–438.

Gül, H., Özer, M., et al. (2018). Frequency domain causality analysis of tourism and economic activity in Turkey. *European Journal of Tourism Research*, *19*, 86–97.

Hidalgo, J. (2000). Nonparametric test for causality with long-range dependence. *Econometrica*, *68*, 1465–1490.

Hidalgo, J. (2005). A bootstrap causality test for covariance stationary processes. *Journal of Econometrics*, *126*, 115–143.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar us business cycles: An empirical investigation. *Journal of Money, Credit, and Banking,* 1–16.

Hong, Y., Liu, Y., & Wang, S. (2009). Granger causality in risk and detection of extreme risk spillover between financial markets. *Journal of Econometrics*, *150*, 271–287.

Hosoya, Y. (1991). The decomposition and measurement of the interdependency between second-order stationary processes. *Probability Theory and Related Fields*, *88*, 429–444.

Hosoya, Y. (2001). Elimination of third-series effect and defining partial measures of causality. *Journal of Time Series Analysis*, *22*, 537–554.

Jung, A., & Villanova, H. C. (2020). The empirical properties of euro area m3, 1980–2017. *The Quarterly Review of Economics and Finance*.

Keho, Y., et al. (2015). Revisiting Wagner's law for selected African countries: A frequency domain causality analysis. *Journal of Statistical and Econometric Methods*, *4*(4), 1–4.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Lemmens, A., Croux, C., & Dekimpe, M. G. (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. (2005). *New introduction to multiple time series analysis*. Cambridge University Press.

Musso, A., et al. (2019). The predictive power of real m1 for real economic activity in the euro area. *Economic Bulletin Boxes* 3.

Nadig, A., & Viswanathan, T. (2019). An empirical analysis of money supply, inflation and output: The case of India. *International Journal of Public Sector Performance Management*, *5*(3–4), 516–530.

Nishiyama, Y., Hitomi, K., Kawasaki, Y., & Jeong, K. (2011). A consistent nonparametric test for nonlinear causality-specification in time series regression. *Journal of Econometrics*, *165*(1), 112–127.

Pfaff, B., et al. (2008). VAR, SVAR and SVEC models: Implementation within R package VARs. *Journal of Statistical Software*, *27*(4), 1–32.

Phillips, P. C. (1988). Spectral Regression for Cointegrated Time Series. Cowles Foundation Discussion Papers 872, Cowles Foundation for Research in Economics, Yale University.

Pierce, D. A. (1979). R 2 measures for time series. *Journal of the American Statistical Association*, *74*, 901–910.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical association*, *89*, 1303–1313.

Psaradakis, Z., Ravn, M. O., & Sola, M. (2005). Markov switching causality and the money-output relationship. *Journal of Applied Econometrics*, *20*, 665–683.

Shao, J., & Tu, D. (1995). *The jackknife and bootstrap*. Springer.

Tastan, H. (2015). Testing for spectral granger causality. *The Stata Journal*, *15*(4), 1157–1166.

Tastan, H., & Sahin, S. (2020). Low-frequency relationship between money growth and inflation in Turkey. *Quantitative Finance and Economics*, *4*(1), 91.

Tibshirani, R. J., & Efron, B. (1993). An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, *57*, 1–436.

Tiwari, A. K. (2014). The frequency domain causality analysis between energy consumption and income in the united states. *Economia Aplicada*, *18*, 51–67.

Wei, Y. (2015). The informational role of commodity prices in formulating monetary policy: A reexamination under the frequency domain. *Empirical Economics*, *49*(2), 537–549.

Wiener, N. (1956). The theory of prediction. *Modern Mathematics for Engineers*, *1*, 125–139.

Yamada, H., & Yanfeng, W. (2014). Some theoretical and simulation results on the frequency domain causality test. *Econometric Reviews*, *33*, 936–947.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, *57*(298), 348–368.