Chilean Journal of Statistics
Vol. 6, No. 1, April 2015, 49–65

Time Series
Research Paper

Sensitivity of the causality in variance tests to GARCH$(1,1)$ processes

Farrukh Javed[1,*] and Panagiotis Mantalos[2]

[1]Department of Statistics, Lund University, SE 220 07, Lund Sweden.
[2]Department of Statistics, Örebro University, SE 701 82, Örebro, Sweden.

(Received: 24 April 2014 · Accepted in final form: 22 April 2015)

Abstract

This paper studies the impact of a number of volatile data sets on volatility spillover tests. We investigate a type of data generating process, AR(1)-GARCH(1,1), with an extensive set of Monte Carlo simulations. It is found that causation pattern, due to causality between two series, is influenced by the intensity of volatility clustering. Two testing procedures are applied for testing causality in the variance. We notice a severe size and power distortion when the clustering parameter is high and when the process is near integration. Furthermore, whenever there is a severe size distortion, there is a serial autocorrelation in the standardized residuals. This is seen when the asymptotic distribution of the statistics is used to define a critical region. So, instead of relying on the asymptotic distribution, we calculate the percentiles of the test statistic with the null hypothesis of no spillover effect and use them as a critical region for both size and power. We observe a significant improvement in the results.

Keywords: Causality · GARCH · Spillover · Volatility.

Mathematics Subject Classification: Primary 62M10 · Secondary 91B84.

1. Introduction

The recent financial crunch and its ramifications all around the world highlight the need to understand the linkages across financial markets. If the volatility of one economic fundamental in a specific market changes the behavior of another market, both the two markets are interrelated with respect to volatility. This mechanism is well known as volatility spillover. Volatility spillover can be seen as the transmission of volatility across markets. Such spillover effects, across different markets and assets, have recently been studied by e.g., Asgharian and Bengtsson (2006), Bollio and Pelizzon (2003), Forbes and Rigobon (2002), Kanas (2000), Li et al. (2008), Rigobon and Sack (2003). One way of measuring spillover effects is via establishing cross-correlation between the processes. This approach is easy to implement compared to applying multivariate GARCH models, which requires simultaneous modeling of inter- and intra-series dynamics (Cheung and Ng, 1996).

*Corresponding author. Email: Farrukh.Javed@stat.lu.se

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

Stock returns often exhibit volatility clustering, such as high volatility "*today*" tends to give high volatility "*tomorrow*". Such clusters result in persistence of the amplitudes of price changes. The clustering mechanism also tells us something about the predictability of volatility. If large changes in financial markets tend to be followed by more large changes, and small changes by small change, then volatility must be highly predictable. Good insight into such behavior is very helpful for investors developing investment strategies. For instance, from a portfolio perspective, a substantial increase in volatility has a negative effect on risk averse investors, and whenever there is a crisis, risk averse investors choose to adjust their portfolios either by selling highly risky assets or by adopting other approaches such as hedging. As asset prices are interrelated due to a number of common driving forces (e.g., macroeconomic development, central bank's policy over interest rates or risk preferences), analyzing single series in isolation may amount to ignore important information about its true behaviour.

Javed and Mantalos (2012) have observed an impact of volatility structure of time series on the performance of information criteria. In Mantalos and Shukur (2005), strong GARCH effects have been shown to affect tests for autocorrelation in the stationary dynamic system of equations. Phenomena such as these inspired us to study the impact of volatility clustering on tests for causality in the simple GARCH framework. This is the central theme of this paper and, to our knowledge, it has not been analyzed in previous studies. Van Dijk et al. (2005) discuss how the performance of the causality in variance tests proposed by Cheung and Ng (1996) and Hong (2001) are affected by the presence of breaks. Van Dijk et al. (2005) found severe size distortion in the presence of breaks that were ignored. In Hecq (1996), the author noticed how information criteria determine, in the presence of five integrated highly volatility GARCH errors, an optimal lag length in univariate time series and causality tests. The author also stresses that one should be cautious in the use of ADF unit root tests as well as in the augmented EngleGranger cointegration test for large volatile parameters. The Granger causality (Granger, 1969, 1980) tests are popularly known methods for investigating the direction of causality in various set of time series due to its simplicity and easy applicability. Having noticed the importance of GARCH process with respect to the level of persistence of volatility, in this work, we study the Granger causality in variance in the presence of GARCH effects at high and medium levels of persistence with the sum of parameters (α, β) indicating the level of the conditional heteroskedastic process moving towards integration.

A financial market is said to be complete if it is arbitrage free, that is no profit can be made by taking an advantage of a price difference between two or more markets. According to Ross (1989), in an arbitrage free economy the variance of price change is directly related to the rate of information flow into the market. War, regime changes, or economic crises could be some of the possible reasons for volatile behaviour. This issue led us to investigate the effect of GARCH processes on the causality in variance tests. We simulated a number of financial series, based on various volatile structures, with GARCH(1,1) errors. An extensive set of data generating processes was used to reach the above-mentioned objective. We consider high and medium level GARCH$(1, 1)$ processes with different parameter combinations.

Causality in variance tests are used for assessing the causation pattern between two time series. The tests require the residuals of the tested series to be serially independent (Cheung and Ng, 1996). However, some financial time series models do not account for serial autocorrelation. In such cases, serial autocorrelation might influence the ability of the tests to detect causality in mean and variance. We investigate various volatile behavior of time series by means of Monte Carlo simulations. The causality tests appear to be sensitive with respect to the parameters of simulated series. Whenever there is a severe size distortion, there is a serial autocorrelation in the standardized residuals. Moreover, a change in the shape of the distribution of the test statistics is also noticed. We therefore calculated the empirical critical values based on the 90% and 95% percentiles of the one-way and two-way causality test statistics (Q_1, Q_2, S_1 and S_2) for all the simulated processes. These critical values (instead of asymptotic critical values) were later used to study the size and power of the test statistics.

The remaining paper proceeds as follows: Section 2 discusses the hypotheses of volatility spillover and the test procedures used. Section 3 describes the extensive set of Monte Carlo simulations designed for the present study. In Section 4, we summarize the size result of these tests for no spillover. Results for the size adjusted power are summarized in Section 5. Finally, in the last section, general conclusions of the analysis are drawn.

2. Hypotheses of interest and test procedures

Following the work of Cheung and Ng (1996), we let $I_{i,t}, i = 1, 2$, be the information set of the time series $Y_{i,t}$ available at period t, and $I_t = (I_{1,t}, I_{2,t})$. Let

$$\epsilon_{i,t} = Y_{i,t} - \mu_{i,t}, \quad i = 1, 2, \tag{1}$$

where $\mu_{i,t}$ is the mean of $Y_{i,t}$ conditioned on $I_{i,t-1}$ and assume that

$$\epsilon_{i,t} = \xi_{i,t}(h_{i,t})^{1/2}. \tag{2}$$

Here $h_{i,t}$ is a positive time-varying measurable function that is the conditional variance of $\epsilon_{i,t}$ and $\xi_{i,t}$ is an innovation process with

$$E(\xi_{i,t}|I_{i,t-1}) = 0, \text{ and } \quad E(\xi^2_{i,t}|I_{i,t-1}) = 1. \tag{3}$$

The null hypothesis that $y_{2,t}$ does not Granger-cause $y_{1,t}$ in variance is:

$$H_0 : \mathrm{Var}(\xi_{1,t}|I_{1,t-1}) = \mathrm{Var}(\xi_{1,t}|I_{t-1}),$$

versus the alternative

$$H_A : \mathrm{Var}(\xi_{1,t}|I_{1,t-1}) \neq \mathrm{Var}(\xi_{1,t}|I_{t-1}).$$

As the squared innovations, $\xi^2_{i,t}$, are not observable, the squared standardized residual is used as an estimate of $\xi^2_{i,t}$ and the normalized residuals are defined as:

$$\hat{u}_t = \hat{\epsilon}^2_{1,t}/\hat{h}_{1,t} - 1 \quad \text{ and } \quad \hat{v}_t = \hat{\epsilon}^2_{2,t}/\hat{h}_{2,t} - 1. \tag{4}$$

where $\hat{}$ indicates suitable estimates of the corresponding entities.

Cheung and Ng (1996) proposed a test statistic for H_0 by using the sample cross-correlation function between $\hat{u}_t$ and $\hat{v}_t$, defined as

$$\hat{\rho}_{uv}(j) = \{\hat{C}_{uu}(0)\hat{C}_{vv}(0)\}^{-1/2}\hat{C}_{uv}(j), \tag{5}$$

where the sample cross-covariance function is

$$\hat{C}_{uv}(j) = \begin{cases} T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{v}_{t-j}, & j \geq 0 \\ T^{-1} \sum_{t=-j+1}^{T} \hat{u}_{t+j} \hat{v}_t, & j < 0, \end{cases} \tag{6}$$

$\hat{C}_{uu}(0) = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2$ and $\hat{C}_{vv}(0) = T^{-1} \sum_{t=1}^{T} \hat{v}_t^2$. These authors suggest a testing procedure for H_0 using a statistic based on M squared cross-correlations as given by

$$S_1 = T \sum_{j=1}^{M} \hat{\rho}_{uv}^2(j). \tag{7}$$

They claim that under H_0, the statistic S_1 has an asymptotic χ^2_M distribution. Here M is the number of lags from the sample cross-correlation. The squared cross-correlations from lag $-M$ to -1 can be used for testing the reverse hypothesis of $Y_{1,t}$ does not cause $Y_{2,t}$ in variance. When there is no prior information about the direction of causality, it is appropriate to test the two-way hypothesis that neither $Y_{1,t}$ Granger-cause $Y_{2,t}$ nor $Y_{2,t}$ Granger-cause $Y_{1,t}$ in variance with respect to $(I_{1,t}, I_{2,t-1})$ or $(I_{1,t-1}, I_{2,t})$ respectively. The test statistic for testing the two-way hypothesis is

$$S_2 = T \sum_{j=-M}^{M} \hat{\rho}_{uv}^2(j). \tag{8}$$

The same authors have shown that under H_0, the statistic S_2 has an asymptotic χ^2_{2M+1} distribution.

An interesting feature of financial markets is that the volatility of a current asset or market is often more affected by the recent volatility pattern of another asset or market than by the distant past values. Hong (2001) introduced a weighting scheme to incorporate such a situation. According to Hong, when large M is used, the S_1 and S_2 statistic may not efficiently estimate volatility spillover as it gives equal weights to each of the M sample cross-correlations. A more efficient test may hence be obtained by assigning higher weights to more recent information. Hong (2001) proposed a test statistic described as

$$Q_1 = \{T \sum_{j=1}^{T-1} k^2(j/M) \hat{\rho}_{uv}^2(j) - C_{1T}(k)\} / \{2D_{1T}(k)\}^{1/2}\}. \tag{9}$$

where

$$C_{1T}(k) = \sum_{j=1}^{T-1} (1 - j/T) k^2(j/M),$$

and

$$D_{1T}(k) = \sum_{j=1}^{T-1} (1 - j/T)\{1 - (j+1)/T\} k^4(j/M).$$

The constants $C_{1T}(k)$ and $D_{1T}(k)$ can be seen as the mean and variance of the sum in equation (9) and $k(.)$ is a kernel function. Hong (2001) has shown through simulations that the test provides better power performance when non-uniform kernels are used. Similarly, for the bidirectional hypothesis a proposed statistic was given by

$$Q_2 = \{T \sum_{j=1-T}^{T-1} k^2(j/M)\hat{\rho}_{uv}^2(j) - C_{2T}(k)\}/\{2D_{2T}(k)\}^{1/2}\}. \tag{10}$$

where

$$C_{2T}(k) = \sum_{j=1-T}^{T-1} (1 - |j|/T)k^2(j/M),$$

and

$$D_{2T}(k) = \sum_{j=1-T}^{T-1} (1 - |j|/T)\{1 - (|j| + 1)/T\}k^4(j/M).$$

According to Hong (2001), Q_1 and Q_2 have asymptotic $\mathrm{N}(0, 1)$ distributions. We use a non-uniform Bartlett kernel (Priestley, 1981), as a weighting function in our simulation:

$$k(z) = \begin{cases} 1 - |z|, \, |z| \leq 1 \\ 0 \qquad\quad \text{otherwise} \end{cases}$$

We use $M = 1$, 5, 10, 15 and 20 for our analysis.

3. The Monte Carlo design

In this section we describe our data generating processes with a number of GARCH specifications that will be used to study the size and power of the above-mentioned tests. The model we consider is given by:

$$Y_{i,t} = 0.8Y_{i,t-1} + \epsilon_{i,t} \quad \text{and} \quad \epsilon_{i,t} = \xi_{i,t}h_{i,t}^{1/2}, \quad i = 1, 2, \quad t = 1, \ldots, T,$$

where $\xi_{i,t} \sim NID(0, 1)$ and the conditional variance is driven as:

$$h_{i,t} = w_i + \alpha_i\epsilon_{i,t-1}^2 + \beta_i h_{i,t-1} + \phi_i\epsilon_{j,t-1}^2 + \psi_i h_{j,t-1}, \quad i \neq j, \quad i, j = 1, 2,$$

For simplicity, we can rewrite it as:

$$\begin{bmatrix} h_{1,t} \\ h_{2,t} \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 & \phi_1 \\ \phi_2 & \alpha_2 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} \beta_1 & \psi_1 \\ \psi_2 & \beta_2 \end{bmatrix} \begin{bmatrix} h_{1,t-1} \\ h_{2,t-1} \end{bmatrix}$$

or

$$\mathbf{h}_t = \boldsymbol{\omega} + \mathbf{A}\epsilon_{t-1}^{(2)} + \mathbf{B}\mathbf{h}_{t-1} \tag{11}$$

Based on the above conditional variance processes, we design a data generation process. Since the clustering effect and persistence are the two integral features observed in financial data, a GARCH$(1, 1)$ is an obvious for capturing such effects (from now on we will call them a GARCH effect, for simplicity). However, the data generally varies with respect to the strength of clustering and persistence effects it has. A high value of β means that the volatility is *persistent* and will take a longer time to stabilize. Similarly a high value of α means that the volatility is *spiky* and will react to the market movements very quickly. If for cases when they are close to unity ($\alpha + \beta = 1$), it means the persistence in volatility is high. Awareness of these properties is paramount when modelling financial data. And since they can have different effects on volatility, they can have different implications on the spillover effect, also known as causality pattern.The following Monte Carlo design is made by keeping in mind the significance of variation in financial data. Below we present set of models, that varies with respect to their strength of persistence and clustering.

$$H_0(\text{Model 1}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.19, 0.8, 0, 0), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.19, 0.8, 0, 0), \end{cases}$$

$$H_0(\text{Model 2}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.09, 0.9, 0, 0), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.09, 0.9, 0, 0), \end{cases}$$

$$H_0(\text{Model 3}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.05, 0.94, 0, 0), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.05, 0.94, 0, 0), \end{cases}$$

$$H_0(\text{Model 4}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.15, 0.8, 0, 0), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.15, 0.8, 0, 0), \end{cases}$$

$$H_0(\text{Model 5}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.05, 0.9, 0, 0), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.05, 0.9, 0, 0), \end{cases}$$

$$H_0(\text{Model 6}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.2, 0.7, 0, 0), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.2, 0.7, 0, 0), \end{cases}$$

$$H_A(\text{Model 7}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.1, 0.8, 0.09, 0.8), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.19, 0.8, 0, 0), \end{cases}$$

$$H_A(\text{Model 8}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.3, 0.4, 0.15, 0.4), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.45, 0.5, 0, 0), \end{cases}$$

$$H_A(\text{Model 9}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.3, 0.4, 0.20, 0.4), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.3, 0.65, 0, 0), \end{cases}$$

$$H_A(\text{Model 10}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.2, 0.5, 0.14, 0.15), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.2, 0.5, 0.14, 0.15), \end{cases}$$

$$H_A(\text{Model 11}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.2, 0.5, 0.2, 0.15), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.2, 0.5, 0.2, 0.15), \end{cases}$$

$$H_A(\text{Model 12}) : \begin{cases} (\alpha_1, \beta_1, \phi_1, \psi_1) = (0.15, 0.45, 0.15, 0.3), \\ (\alpha_2, \beta_2, \phi_2, \psi_2) = (0.15, 0.45, 0.15, 0.3), \end{cases}$$

In the first three models, the strength of persistence is kept high and they are equal in parametric value ($\alpha + \beta = 1$), but the way they drive individual volatility process is different. Similarly model 4 and 5 are chosen to have slightly lower persistence but behave in a different manner. For the Model 6, we chose to simulate with slightly lower persistence($\alpha+\beta = 0.9$). The whole idea is to investigate how difference in persistence affect the causality pattern evaluated by a single test statistic. Moreover, there is no volatility spillover between $Y_{1,t}$ and $Y_{2,t}$ under H_0 for the first six models. For Model 7 to Model 9 under H_A there is a one-way volatility spillover from $Y_{2,t}$ to $Y_{1,t}$ but not from $Y_{1,t}$ to $Y_{2,t}$. For the last three models there is two-way volatility spillover from $Y_{1,t}$ to $Y_{2,t}$ as well as from $Y_{2,t}$ to $Y_{1,t}$.

As we discussed earlier, different GARCH(1,1) processes can have different implications for the causality pattern. In addition to this, there are some assumptions which must be satisfied for the causality test to be asymptotically distributed as χ^2_M. The asymptotic distribution of the cross-correlation between standardized residuals follows a χ^2_M if the standardized residuals are independent of each other. It also requires that the squared cross-correlations at any lag must be independent of at any other lag. Dependence between the squared cross-correlation vectors at any stage can violate the assumption of χ^2_M. Moreover, unequal variance or strong dependence between the residuals can affect the asymptotic properties of the distribution. It is quite likely that the presence of serial autocorrelation will influence the ability of the test to detect a causality pattern in mean and variance. Moreover, a significant autocorrelation in the standardized residuals can change the distribution shape of the test statistic.

Models such as these considered here, are very common in financial data such as, exchange rates data. Discussing the strength of the test statistic require a separate paper with more focus and it has been done in few other papers by the same authors. In this paper we focus on discussing the discrepancies found in the test statistic though extensive simulations. Therefore, we only report results obtained through simulations in this paper, and interested reader may refer to articles such as, Javed (2013) and Mantalos and Shukur (2010) on detailed analysis and discussion on the causality test when applied to empirical data.

We consider two sample sizes, $T = 300$ and $T = 1000$. All simulations are further replicated 1000 times to study the size and power of the tests. For each series we generate $T + 1000$ observations and then the first 1000 are discarded to reduce the possible effect of the start-up values. An autoregressive model $Y_{i,t}$ with GARCH errors is considered first. To obtain this model, we first simulate the variance process as a GARCH$(1, 1)$ process with $\alpha_i = 0.19$ and $\beta_i = 0.8$, as the GARCH process depends on its lag value. Unconditional variance is used as an initial value for $h_{i,0}$. The simulated results are used to generate $Y_{i,t}$. Then we fit an AR(1)-GARCH(1,1) model to the simulated $Y_{i,t}$ series to obtain the estimates of residuals and conditional variances. Using these estimates, we compute the squared standardized residuals for the two series $Y_{1,t}$ and $Y_{2,t}$. Finally, the correlation structure computed from these standardized residuals is used to calculate the test statistics S_i and Q_i, $i = 1, 2$. The same procedure of simulation and computation is applied to the remaining models.

4. Size of the tests

To examine the effect of these GARCH processes on the proposed tests discussed in section 2 (S_i and Q_i for $i = 1, 2$), we consider a number of squared cross-correlations at lag $M =$ 1, 5 and 15.

The performance of the chosen causality tests is investigated with the help of volatility models described in section 3. We notice a change of behavior with the change of parameter combinations in the GARCH process. Our findings for the size and power of the tests depend on the use of the asymptotic distribution of test statistics (either chi-square, with appropriate degree of freedom, or Gaussian) as basis for choosing critical values. The actual figures obtained from this investigation are not reported here – they will however be provided upon request – instead the general tendencies that were found are summarized below. The critical values from the χ^2_j distribution are reported in Table 1 in the appendix.

For the size investigation, the simulated processes are divided into three classes. In the first class, three processes (Model 1 to Model 3) are considered with different parameter combinations. All these GARCH models indicate a process near integration. First, in Model 1 with $\alpha_i = 0.19$ and $\beta_i = 0.8$, both tests (S_1 and Q_1) for one-way causality are over-rejected at all lags (for $M = 1$ the rejection probability is 20% at the nominal 10% level) and the rate of over-rejection increases with larger M. For the two-way causality tests, the over-rejection rate is 30% at the 10% level, which is higher than what we observed for the one-way tests. For model 2, another process is simulated with a slightly lower value of α_i ($\alpha_i = 0.09$) compared to Model 1 while keeping $\beta_i = 0.8$. A low rate of over-rejection has been noticed comparing with the first model with around 10% at the 5% level of significance. The tests Q_1 and S_1 perform better than their two-way counterparts, Q_2 and S_2. A third combination is considered in Model 3, with $\alpha_i = 0.05$ and $\beta_i = 0.94$. In this model, we notice that the proportion of rejection approaches the nominal size. The rejection probability is around 7% (10%) at the 5% (10%) level of significance. A slightly higher over-rejection is observed in the case of the two-way causality tests. The purpose of using three different combinations of the same kind of model ($\alpha_i + \beta_i = 0.99$) is to study the effect of parameter change on the tests. We noticed an increasing over-rejection rate with an increase in the clustering parameter α_i for processes near integration.

In the second class of models $\alpha_i + \beta_i = 0.95$. Two different parameter combinations are studied here. In Model 4, where $\alpha_i = 0.15$ and $\beta_i = 0.8$, Q_1 and S_1 perform better at initial lags (the over-rejection rate is 10% at the 5% significance level). However, the rejection rate increases with larger lags (say, 10% for $M = 15$). Of the two-way causality tests, S_2 performs better than Q_2. However, they both have high over-rejection rates compared to the one-way tests. In Model 5, where $\alpha_i = 0.05$ and $\beta_i = 0.9$, the results are close to the nominal sizes for S_1 and S_2. Their performance seems consistent with the lags (the rate of rejection is around 6% (10%) at the 5% (10%) levels) while the other two tests (Q_1 and Q_2) showed some over-rejection especially at higher M. Similar to what we observed in the first class of models ($\alpha_i + \beta_i = 0.99$), a low value of α_i yields rejection rate close to the nominal sizes (i.e. 0.10 or 0.05) for these tests.

In the third class we find Model 6 with a slightly higher value of α_i. We consider a model with $\alpha_i = 0.2$ and $\beta_i = 0.7$. This model is even further below unity than in the previous models. Both tests S_1 and S_2 perform much better at lower order M(5% at $M = 1$) but they over-reject at higher M (10% at $M = 15$) at the 5% level of significance. As we expect the model to perform near nominal sizes both at 10% and 5%, for both Q_1 and Q_2, the performance is far from expectation (the probability of rejection is 14% (17%) at the 5% (10%) level). If we compare the same parameter value for α_i, namely $\alpha_i = 0.05$ in two Models (say, Model 3 and Model 5), we can say that the further away the process is from integration in the GARCH process, the less is the impact of over rejection on the tests for causality.

As mentioned earlier, the presence of autocorrelation can affect the ability of the tests to capture the causality pattern. We found serial autocorrelation in our estimated standardized residuals for models in which there was severe size distortion. We therefore estimated the percentiles of the test statistics for each of the data generating processes. Tables B-E (in the appendix) report the findings of these estimations. It is worth noticing that the percentiles estimated from the data generating processes are far away from the critical values given by the asymptotic distribution (see Table A). Due to these large discrepancies, the critical values from the asymptotic distribution do not seem to be suitable for use in such processes. We estimate the 90% and 95% percentiles of the statistics Q_1, Q_2, S_1 and S_2 for all the simulated processes. The distributions of percentiles for Models 3- 6 are very close to each other.

We tried several combinations of model structures, and used the estimated percentiles to calculate the size and power of the tests. The estimated results for $T = 1000$ are presented in Tables 1 and 2 (for $T = 300$, see Tables F and G in the appendix). From these tables, we clearly see that the nominal size has been achieved in all cases. This shows that changing the parameter combination in the model does indeed change the distribution of the process.

Table 1. The size of the tests at 10% and 5% levels for $T = 1000$ using empirical critical values given in Table C. The numbers represent the empirical probability of rejection of the null-hypothesis

				M	
Model	Statistic	Levels	1	5	15
Model 1	Q_1	10%	0.10	0.09	0.10
		5%	0.04	0.04	0.05
	S_1	10%	0.10	0.10	0.10
		5%	0.05	0.05	0.05
	Q_2	10%	0.10	0.11	0.11
		5%	0.06	0.07	0.07
	S_2	10%	0.11	0.10	0.10
		5%	0.05	0.07	0.06
Model 2	Q_1	10%	0.10	0.10	0.09
		5%	0.05	0.05	0.05
	S_1	10%	0.08	0.09	0.10
		5%	0.041	0.04	0.05
	Q_2	10%	0.10	0.11	0.11
		5%	0.04	0.06	0.06
	S_2	10%	0.10	0.09	0.10
		5%	0.06	0.05	0.04
Model 3	Q_1	10%	0.09	0.08	0.08
		5%	0.04	0.04	0.04
	S_1	10%	0.10	0.10	0.08
		5%	0.06	0.04	0.03
	Q_2	10%	0.13	0.13	0.12
		5%	0.08	0.07	0.07
	S_2	10%	0.07	0.07	0.08
		5%	0.04	0.04	0.05

Table 2. The size of the tests at 10% and 5% levels for $T = 1000$ using empirical critical values given in Table C. The numbers represent the empirical probability of rejection of the null-hypothesis

				M	
Model	Statistic	Levels	1	5	15
Model 4	Q_1	10%	0.08	0.08	0.08
		5%	0.04	0.05	0.04
	S_1	10%	0.11	0.10	0.10
		5%	0.05	0.06	0.05
	Q_2	10%	0.11	0.10	0.10
		5%	0.06	0.05	0.04
	S_2	10%	0.11	0.11	0.12
		5%	0.07	0.06	0.07
Model 5	Q_1	10%	0.09	0.10	0.10
		5%	0.06	0.05	0.05
	S_1	10%	0.09	0.08	0.10
		5%	0.05	0.05	0.05
	Q_2	10%	0.10	0.10	0.10
		5%	0.05	0.04	0.05
	S_2	10%	0.10	0.10	0.10
		5%	0.05	0.05	0.05
Model 6	Q_1	10%	0.13	0.13	0.10
		5%	0.05	0.06	0.05
	S_1	10%	0.10	0.12	0.11
		5%	0.06	0.07	0.07
	Q_2	10%	0.107	0.09	0.10
		5%	0.03	0.04	0.05
	S_2	10%	0.09	0.10	0.11
		5%	0.05	0.05	0.06

5. Power of the size adjusted tests

The powers of the tests were initially investigated on the basis of the asymptotic distribution. The results obtained in this analysis were far away from the expectations. For brevity reasons, we do not report the results here, they will, however, be provided upon request. Instead the power of the size adjusted tests, i.e., these using the empirical critical values discussed in previous sections, is investigated, using several GARCH processes. The results are reported in Tables 3-8.

Table 3. The size adjusted power of the tests for one-way spillover using Model 7. The numbers represent the empirical probability of rejection of the null-hypothesis

				M	
Sample Size	Statistic	Levels	1	5	15
T=300	Q_1	10%	0.08	0.07	0.07
		5%	0.05	0.04	0.04
	S_1	10%	0.11	0.09	0.10
		5%	0.04	0.05	0.05
	Q_2	10%	0.07	0.10	0.10
		5%	0.05	0.05	0.05
	S_2	10%	0.09	0.09	0.10
		5%	0.04	0.06	0.06
T=1000	Q_1	10%	0.07	0.07	0.09
		5%	0.05	0.05	0.04
	S_1	10%	0.12	0.10	0.09
		5%	0.06	0.04	0.04
	Q_2	10%	0.10	0.09	0.09
		5%	0.04	0.05	0.05
	S_2	10%	0.10	0.10	0.09
		5%	0.04	0.05	0.05

Table 4. The size adjusted power of the tests for one-way spillover using Model 8. The numbers represent the empirical probability of rejection of the null-hypothesis

				M	
Sample Size	Statistic	Levels	1	5	15
T=300	Q_1	10%	0.12	0.19	0.15
		5%	0.07	0.07	0.11
	S_1	10%	0.13	0.15	0.18
		5%	0.06	0.11	0.13
	Q_2	10%	0.11	0.14	0.16
		5%	0.06	0.09	0.11
	S_2	10%	0.13	0.18	0.19
		5%	0.09	0.11	0.15
T=1000	Q_1	10%	0.15	0.12	0.15
		5%	0.08	0.08	0.10
	S_1	10%	0.13	0.15	0.21
		5%	0.06	0.09	0.12
	Q_2	10%	0.09	0.15	0.21
		5%	0.07	0.11	0.14
	S_2	10%	0.14	0.18	0.22
		5%	0.08	0.12	0.16

For the process generated from Model 7, we simulated a series with high persistence and low volatility clustering. We observed a small rate of proportion defining power of one-way spillover which increases with the lag length. Adding more sample observation do increase the power but still seems unsatisfactory. Both of the causality tests hardly captured the one-way spillover effect at all. The tests suggest that there is no spillover in the underlying process. Similar observation has been noticed for Models 8 and 9. It can be seen that the results are still not satisfactory and probably large lags are required to capture the one-way spillover effect. According to Nakatani and Teräsvirta (2008), the GARCH process in Eqn. (11) is said to be weekly stationary if and only if the module of the largest eigenvalue of $(\mathbf{A}+\mathbf{B})$ is less than 1. In order to identify the effect of such processes on the causality tests, we simulate processes with different eigenvalues and evaluate the size adjusted power performance. The results are summarized below.

Table 5. The size adjusted power of the tests for one-way spillover using Model 9. The numbers represent the empirical probability of rejection of the null-hypothesis

				M	
Sample Size	Statistic	Levels	1	5	15
T=300	Q_1	10%	0.13	0.12	0.13
		5%	0.06	0.05	0.08
	S_1	10%	0.11	0.15	0.21
		5%	0.07	0.10	0.14
	Q_2	10%	0.10	0.13	0.17
		5%	0.06	0.08	0.10
	S_2	10%	0.11	0.13	0.18
		5%	0.08	0.08	0.13
T=1000	Q_1	10%	0.14	0.11	0.16
		5%	0.06	0.07	0.11
	S_1	10%	0.09	0.14	0.19
		5%	0.06	0.08	0.10
	Q_2	10%	0.10	0.14	0.19
		5%	0.05	0.09	0.10
	S_2	10%	0.14	0.17	0.22
		5%	0.07	0.11	0.14

Table 6. The size adjusted power of the tests for two-way spillover using Model 10. The numbers represent the empirical probability of rejection of the null-hypothesis

				M	
Sample Size	Statistic	Levels	1	5	15
T=300	Q_1	10%	0.14	0.12	0.15
		5%	0.08	0.08	0.11
	S_1	10%	0.11	0.15	0.24
		5%	0.06	0.11	0.17
	Q_2	10%	0.11	0.15	0.20
		5%	0.07	0.10	0.14
	S_2	10%	0.14	0.18	0.21
		5%	0.09	0.13	0.16
T=1000	Q_1	10%	0.19	0.20	0.22
		5%	0.09	0.10	0.15
	S_1	10%	0.15	0.21	0.22
		5%	0.09	0.14	0.16
	Q_2	10%	0.13	0.21	0.27
		5%	0.09	0.13	0.20
	S_2	10%	0.22	0.28	0.34
		5%	0.13	0.21	0.25

For the size adjusted power analysis, we chose three different processes, Models 10 to 12 in section 3. For Model 10, the largest eigenvalue is 0.99, whereas for Models 11 and 12, it is slightly greater than 1 (1.05 each). We deliberately choose such models which are non-stationary in order to analyze the power property. Rodrigues and Rubia (2007) showed severe finite sample size distortions for nonstationary volatility processes. They argue that the size departures are not mainly due to the small sample effect but will remain asymptotically because of the failure to consistently estimate cross-correlation in this context. We observe similar results for power analysis because if the test is over-rejected for some processes, the power of the test for such processes will be high. For processes away from nonstationarity as in Model 10, we find less power and only S_2 has managed to get almost 35% rejection rate for $n = 1000$. But for nonstationary processes as in Model 11 and 12, we observe an increase in power which is up to 65% and 85% percent for Model 11 and 12 respectively. All three models gain more power with increased sample size and lag values.

Table 7. The size adjusted power of the test for two-way spillover using Model 11. The numbers represent the empirical probability of rejection of the null-hypotheses

				M	
Sample Size	Statistic	Levels	1	5	15
T=300	Q_1	10%	0.17	0.28	0.38
		5%	0.12	0.22	0.30
	S_1	10%	0.23	0.35	0.42
		5%	0.16	0.29	0.34
	Q_2	10%	0.27	0.39	0.47
		5%	0.19	0.30	0.40
	S_2	10%	0.35	0.44	0.49
		5%	0.28	0.36	0.44
T=1000	Q_1	10%	0.33	0.45	0.54
		5%	0.24	0.38	0.47
	S_1	10%	0.41	0.51	0.55
		5%	0.32	0.44	0.48
	Q_2	10%	0.38	0.50	0.57
		5%	0.31	0.42	0.52
	S_2	10%	0.52	0.57	0.63
		5%	0.44	0.52	0.59

Table 8. The size adjusted power of the test for two-way spillover using Model 12. The numbers represent the empirical probability of rejection of the null-hypothesis

				M	
Sample Size	Statistic	Levels	1	5	15
T=300	Q_1	10%	0.51	0.54	0.61
		5%	0.40	0.47	0.57
	S_1	10%	0.42	0.60	0.65
		5%	0.36	0.55	0.62
	Q_2	10%	0.50	0.61	0.68
		5%	0.42	0.55	0.63
	S_2	10%	0.57	0.66	0.68
		5%	0.50	0.59	0.65
T=1000	Q_1	10%	0.78	0.79	0.82
		5%	0.65	0.75	0.79
	S_1	10%	0.42	0.60	0.65
		5%	0.36	0.55	0.62
	Q_2	10%	0.79	0.83	0.85
		5%	0.77	0.81	0.83
	S_2	10%	0.83	0.86	0.88
		5%	0.79	0.85	0.86

6. Conclusion

The causality in variance tests are sensitive to the presence of various volatile data. We analyzed the size and power properties of the causality tests using asymptotic null distribution. Our findings show severe distortion in the size, especially for processes near integration. Moreover, we noticed serial correlation between the standardized residuals for cases in which there is size distortion. Due to these findings, we calculate the empirical percentiles of the tests for several volatility structures. When these percentiles were used as critical values for the causality tests, we obtained a significant improvement in the results. The tests attained the nominal size for one- and two-way spillover hypotheses. We conclude that the asymptotic distribution should not be used without a proper understanding of the underlying processes. Without such understanding, the risk is immediate that the analyst is inferring a wrong causation pattern.

These findings contributes to the existing literature by (1) analyzing the sensitivity of causality in variance tests to several GARCH processes and (2) investigating the performance of the tests using the asymptotic null distribution and the true percentile of the test statistics for each of the underlying processes. Future extension of this study can be seen as a thorough investigation of GARCH processes with different parameterizations.

These findings will help to overcome the problems, such as one mentioned in this paper, and build more robust statistical methods based on such process.

7. Acknowledgments

The authors wish to thank Professor Björn Holmquist for all his guidance in helping improving this paper.

References

Asgharian, H., Bengtsson, C., 2006. Jump spillover in international equity markets. Journal of Financial Econometrics, 4, 167-203.

Bollio, M., Pelizzon, L., 2003. Volatility and shocks spillover before and after EMU in European stock markets. Journal of Multinational Financial Management, 13, 323-340.

Cheung, Y. W., Ng, L. K., 1996. A causality-in-variance test and its application to financial market prices. Journal of Econometrics, 72, 33-48.

Forbes, K. J., Rigobon, R., 2002. No contagion, only interdependence: measuring stock market comovements. Journal of Finance, 57, 2223-2261.

Granger, C. W. J., 1969. Investigation causal relations by econometric models and cross-spectral methods. Econometrica, 37, 424-438.

Granger, C. W. J., 1980. Testing for causality: a personal view. Journal of Economic Dynamics and Control, 2, 329-352.

Hecq, A., 1996. IGARCH effect on autoregressive lag length selection and causality tests, Applied Economics Letters, 3, 317-323.

Hong, Y., 2001. A test for volatility spillover with application to exchange rates. Journal of Econometrics, 103, 183-224.

Javed, F., 2013. Effects of Jumps on the Causation Pattern: An International Evidence. International Journal of Computational Economics and Econometrics, 3, 187-204.

Javed, F., Mantalos, P., 2012. GARCH-Type Models and Performance of Information Criteria. Communications in Statistics - Simulation and Computation, 42, 1917-1933.

Kanas, A., 2000. Volatility spillovers between stock returns and exchange rate changes: international evidence. Journal of Business Finance & Accounting, 27, 447-467.

Li, G., Refalo, F. J., Wu, L., 2008. Causality-in-variance and causality-in-mean among European government bond markets. Applied Financial Economics, 18, 1709-1720.

Mantalos, P., Shukur, G., 2005. The Effect of GARCH(1,1) on Autocorrelation Tests in Dynamic Systems of Equations. Applied Economics, 37, 1907-1903.

Mantalos, P., Shukur, G., 2010. The effect of spillover on the Granger causality test, Journal of Applied Statistics, 37, 1473-1486.

Nakatani, T., Teräsvirta, T., 2008. Positivity constraints on the conditional variances in the family of conditional correlation GARCH models. Finance Research Letters, 5, 88-95.

Priestley, M. B., 1981. Spectral Analysis and Time Series. Academic Press, London.

Rigobon, R., Sack, B., 2003. Spillover across US financial markets. MIT Sloan Working Paper No 4304-03.

Rodrigues, P. M. M., Rubia, A., 2007. Testing for causality in variance under nonstationarity in variance. Economics Letters, 97, 133-137.

Ross, S. A., 1989. Information and volatility: the no-arbitrage martingale approach to timing and resolution irrelevancy. Journal of Finance, 44, 1-17.

van Dijk, D., Osborn, D. R., Sensier, M., 2005. Testing for causality in variance in the presence of breaks. Economic Letters, 89, 193-199.

8. Appendix

Table A. Critical values for the χ_j^2 distribution at 10% and 5% levels

Statistic	Levels	M: 1	5	10	15	20
χ_M^2	10%	2.71	9.24	15.99	22.31	28.41
	5%	3.84	11.07	18.37	24.99	31.41
χ_{2M+1}^2	10%	6.25	17.28	29.62	41.42	52.95
	5%	7.81	19.67	32.67	44.99	56.94

Table B. The percentiles of the one-way spillover tests for T=300. The numbers represent the critical values of the test

Model	Statistic	Levels	M: 1	5	15
Model 1	Q_1	10%	1.86	2.57	3.51
		5%	3.43	4.52	6.48
	S_1	10%	3.23	14.21	38.6
		5%	5.16	19.71	52.7
	Q_2	10%	2.55	3.30	4.47
		5%	4.33	5.32	7.86
	S_2	10%	9.22	29.7	75.2
		5%	15.3	40.5	97.2
Model 2	Q_1	10%	1.25	1.77	2.14
		5%	2.02	2.54	3.33
	S_1	10%	2.81	10.9	27.5
		5%	4.12	15.1	34.9
	Q_2	10%	1.81	2.08	2.59
		5%	3.06	3.51	4.15
	S_2	10%	6.93	20.7	50.1
		5%	8.63	25.9	63.3
Model 3-Model 6	Q_1	10%	1.04	1.27	1.55
		5%	1.83	1.98	2.38
	S_1	10%	2.67	9.76	23.8
		5%	3.83	11.9	28.4
	Q_2	10%	1.34	1.61	1.78
		5%	2.48	2.56	2.67
	S_2	10%	6.85	18.8	45.7
		5%	9.25	23.3	51.9

Table C. The percentiles of the one-way spillover tests for T=1000. The numbers represent the critical values of the test

Model	Statistic	Levels	M: 1	5	15
Model 1	Q_1	10%	2.71	3.87	5.48
		5%	4.32	6.87	9.25
	S_1	10%	4.26	18.4	52.3
		5%	6.26	26.3	78.5
	Q_2	10%	2.96	4.58	6.97
		5%	4.92	6.95	10.7
	S_2	10%	11.3	37.7	104
		5%	16.9	51.9	142
Model 2	Q_1	10%	1.84	2.26	3.24
		5%	2.82	3.87	5.05
	S_1	10%	4.07	13.7	35.9
		5%	5.53	19.5	47.5
	Q_2	10%	2.44	3.25	4.30
		5%	4.21	5.03	6.55
	S_2	10%	9.36	30.1	75.7
		5%	13.0	40.5	110
Model 3-Model 6	Q_1	10%	1.54	1.72	1.88
		5%	2.48	2.49	2.87
	S_1	10%	3.05	10.8	27.3
		5%	4.40	13.5	34.5
	Q_2	10%	1.50	1.78	1.98
		5%	2.35	2.75	3.17
	S_2	10%	7.19	19.8	47.9
		5%	9.01	23.7	56.7

Table D. The percentiles of the two-way spillover tests for T=300. The numbers represent the critical values of the test

Model	Statistic	Levels	M: 1	5	15
Model 7	Q_1	10%	1.86	2.57	3.51
		5%	3.43	4.52	6.48
	S_1	10%	3.23	14.2	38.6
		5%	5.16	19.7	52.7
	Q_2	10%	2.55	3.30	4.47
		5%	4.33	5.32	7.86
	S_2	10%	9.22	29.7	75.2
		5%	15.3	40.5	97.2
Model 8-Model 12	Q_1	10%	1.04	1.27	1.55
		5%	1.83	1.98	2.38
	S_1	10%	2.67	9.76	23.8
		5%	3.83	11.9	28.4
	Q_2	10%	1.34	1.61	1.78
		5%	2.48	2.56	2.67
	S_2	10%	6.85	18.8	45.7
		5%	9.25	23.33	51.9

Table E. The percentiles of the two-way spillover tests for T=1000. The numbers represent the critical values of the test

Model	Statistic	Levels	M: 1	5	15
Model 7	Q_1	10%	2.71	3.87	5.48
		5%	4.32	6.87	9.25
	S_1	10%	4.26	18.4	52.3
		5%	6.26	26.3	78.5
	Q_2	10%	2.96	4.58	6.97
		5%	4.92	6.95	10.7
	S_2	10%	11.3	37.7	104
		5%	16.9	51.9	143
Model 8-Model 12	Q_1	10%	1.54	1.72	1.88
		5%	2.48	2.49	2.87
	S_1	10%	3.05	10.8	27.3
		5%	4.40	13.5	34.5
	Q_2	10%	1.50	1.78	1.98
		5%	2.35	2.75	3.17
	S_2	10%	7.19	19.8	47.9
		5%	9.01	23.6	56.7

Table F. The size of the test for T=300. The numbers represent the empirical probability of rejection of the null-hypothesis

Model	Statistic	Levels	M: 1	5	15
Model 1	Q_1	10%	0.08	0.08	0.10
		5%	0.04	0.04	0.05
	S_1	10%	0.09	0.08	0.10
		5%	0.04	0.04	0.05
	Q_2	10%	0.10	0.11	0.11
		5%	0.06	0.07	0.07
	S_2	10%	0.11	0.10	0.10
		5%	0.05	0.06	0.06
Model 2	Q_1	10%	0.11	0.09	0.09
		5%	0.06	0.05	0.04
	S_1	10%	0.09	0.07	0.10
		5%	0.04	0.03	0.05
	Q_2	10%	0.09	0.09	0.08
		5%	0.04	0.04	0.04
	S_2	10%	0.12	0.12	0.12
		5%	0.07	0.06	0.06
Model 3	Q_1	10%	0.10	0.11	0.10
		5%	0.06	0.07	0.06
	S_1	10%	0.08	0.11	0.10
		5%	0.04	0.07	0.05
	Q_2	10%	0.10	0.11	0.12
		5%	0.05	0.07	0.06
	S_2	10%	0.09	0.10	0.10
		5%	0.04	0.05	0.05

Table G. The size of the test for T=300. The numbers represent the empirical probability of rejection the null-hypothesis

Model	Statistic	Levels	M = 1	5	15
Model 4	Q_1	10%	0.12	0.10	0.10
		5%	0.05	0.05	0.05
	S_1	10%	0.07	0.07	0.07
		5%	0.04	0.04	0.04
	Q_2	10%	0.10	0.10	0.09
		5%	0.05	0.04	0.04
	S_2	10%	0.09	0.10	0.09
		5%	0.04	0.04	0.06
Model 5	Q_1	10%	0.09	0.10	0.10
		5%	0.07	0.07	0.05
	S_1	10%	0.09	0.08	0.10
		5%	0.05	0.05	0.04
	Q_2	10%	0.11	0.10	0.11
		5%	0.04	0.05	0.06
	S_2	10%	0.11	0.11	0.10
		5%	0.05	0.04	0.05
Model 6	Q_1	10%	0.11	0.12	0.09
		5%	0.05	0.06	0.05
	S_1	10%	0.08	0.10	0.10
		5%	0.05	0.06	0.05
	Q_2	10%	0.07	0.09	0.11
		5%	0.05	0.06	0.06
	S_2	10%	0.11	0.11	0.11
		5%	0.05	0.08	0.05