Linear and non-linear causality tests in a LSTAR model: wavelet decomposition in a non-linear environment

Yushu Li[1] and Ghazi Shukur[2]

[1] Center for Labor Market Policy Research (CAFO) Department of Economic and Statistics, Växjö University, Sweden

[2] Jönköping International Business School, and Center for Labor Market Policy Research (CAFO) Department of Economic and Statistics, Växjö University, Sweden

Abstract

In this paper, we use simulated data to investigate the power of different causality tests in a two-dimensional vector autoregressive (VAR) model. The data are presented in a non-linear environment that is modelled using a logistic smooth transition autoregressive (LSTAR) function. We use both linear and non-linear causality tests to investigate the unidirection causality relationship and compare the power of these tests. The linear test is the commonly used Granger causality F test. The non-linear test is a non-parametric test based on Baek and Brock (1992) and Hiemstra and Jones (1994). When implementing the non-linear test, we use separately the original data, the linear VAR filtered residuals, and the wavelet decomposed series based on wavelet multiresolution analysis (MRA). The VAR filtered residuals and the wavelet decomposition series are used to extract the non-linear structure of the original data. The simulation results show that the non-parametric test based on the wavelet decomposition series (which is a model free approach) has the highest power to explore the causality relationship in the non-linear models.

Keywords: Granger causality; LSTAR model; Wavelet multiresolution; Monte Carlo simulation

2000 Mathematics Subject Classifications:

I. Introduction

In vector auto-regression (VAR) models, the causality or feedback relationship between subsets of the variables is always an attractive aspect for further analysis. The causality test is widely applied to explore this kind of relationship, and the Granger-type test has been most frequently used in previous studies. The main idea of the Granger causality test is to measure whether the past information of a set of variables contains information on changes in another set of variables and helps to predict them. It is carried out by checking if the variance of the prediction error of one set of variables at the present time is reduced by the incorporation of the past values from the other set of variables. The mean square error (MSE) is commonly used as a measurement for the prediction error and Granger causality is concerned with the minimized MSE predictor, which is the unbiased conditional mean. However, in Granger causality, the conditional mean is always set as a linear function of variables. Thus when we discuss Granger causality, it always implies that the past values of a set of variables help 'linear predict' another set of variables. Therefore, although the Granger causality test performs well in linear models, when the data show non-linear properties, then factors such as the structural change or heteroscedasticity will affect the forecast error variance based on the linear model that influences the Granger causality test. In other words, the linear Granger causality test tests the significance of the linear coefficients of set of past values of variables in the model. If the underlying variables in the VAR system contain non-linear relationships and we use the test based on the linear model, the linear coefficients of the model may be insignificant, and as a result, the test can not explore any causality relationship and will loose power in the corresponding non-linear environment.

Numerous empirical studies have showed that many economic variables display non-linear features and can only be modelled with non-linear models, such as the business cycle, structural swings, abrupt breaks or time varying coefficients. In order to investigate the causality relationship in non-linear dynamic models, a large number of studies have appeared in the field. Chen *et al.* [1] proposed a method that identifies non-linear dependence according to locally linear approximations and phase reconstructions, with a linear regression predictor employed for the local neighbourhood. The causality test checks if the time index of the neighbourhood points in the reconstructed space helps in predicting the future dynamics. Later, Ancona *et al.* (2004) pointed out that Chen *et al.* [1] required adequately high neighbourhood points in the local linear fitting. Instead, Ancona *et al.* (2004) first proposed a statistically independent condition that should be satisfied for their extended non-linear Granger causality method. They also described a class of non-linear models that satisfies this condition and proposed radial basis functions methods to choose those models. In addition to the above-mentioned parametric methods, Bell *et al.* (1996) proposed a non-parametric method that, using a back-fitting algorithm, first estimates underlying smooth functions to describe the relationship of the response and explanatory variables. Then the authors used F statistics based on the residual sum of squares from the restrictive non-causal and the alternative causal related equations. Another type of non-parametric test was first proposed by Baek and Brock (1992) and later modified by Hiemstra and Jones (1994). This non-parametric test can be viewed as a test of non-causality in density. It is based on the correlation integral that is the estimator of spatial probabilities over time. This non-parametric test has the advantage of simple implementation with good size and power properties, and it is robust to the series being tested as it does not require a specified a priori model. It is therefore widely applied to exploit the non-linear causality relationships in non-linear vector time series , see e.g. Abhyankar (1998) and Huh (2002). Here in this paper, we will use this test to test the causality relationship in our non-linear VAR model.

Among the many varieties of non-linear models, the logistic smooth transition autoregressive (LSTAR) model allows non-linear structures between the data regimes to be described by a logistic smooth regime transition function. This is of particular interest in fields that contain mass of units, where even if the decisions leading unit structure break are made discretely, the aggregated behaviour shows smooth regime changes (see Teräsvirta, 1994). It is natural to extend the univariate LSTAR model to a VAR system when the causality relationship among the variables varies with time smoothly, where the LSTAR model can be used to capture the time dependent characteristic, and the series in the VAR system are non-linear. It is also obvious that in this case, the traditional linear Granger causality test may loose power in exploiting time dependent varying dynamics. Thus, in this paper, we test the time varying causality relationships that can be modelled by a LSTAR type of model in a two variable VAR system, and we use both linear Granger test and the non-parametric non-linear test proposed by Baek and Brock (1992), and Hiemstra and Jones (1994) in order to compare their performances. The investigation of the finite samples properties of the tests has been done by means of Monte Carlo experiment where we use 5,000 replications for the size and power estimations, while for producing the critical values tables we use 10,000 replications.

This paper is organized as follows. Section II presents the non-linear bivariate VAR model for the causality test. Section III introduces the simulated model and the data generating process (DGP) together with the size and power property result of the traditional linear causality test. In section IV, we present and apply the non-parametric non-linear test, while in section V the non-parametric test based on the wavelet decomposition series is introduced. The last section contains a short conclusion.

II. The bivariate VAR model with LSTAR causality

In this paper, we are interested in causality tests in the bivariate VAR system where the causality relationship from series $\{Y_t\}$ to series $\{X_t\}$ is time varying, and this time dependent character is captured by a LSTAR model. Thus the resulting VAR model is as follows:

$$\begin{aligned} x_t &= \mu_1 + \alpha_1 x_{t-1} + F(t,\gamma,c)(\beta_1 + \beta_2 y_{t-1}) + \varepsilon_{1t} \\ y_t &= \mu_2 + \alpha_2 y_{t-1} + \varepsilon_{2t} \end{aligned}, \text{ where} \quad (1)$$

$$F(t,\gamma,c) = \frac{1}{(1+\exp\{-\gamma(t-c)\})} - \frac{1}{2}, \quad \varepsilon_{1t}, \varepsilon_{2t} \sim i.i.d\ N(0,1)$$

The transition function $F(t,\gamma,c)$ implies that the causality relationship from $\{Y_t\}$ to $\{X_t\}$ changes smoothly as time evolves under the restriction of $\beta_2 \neq 0, \gamma \neq 0$. In $F(t,\gamma,c)$, γ determines the speed of transition from one extreme regime to another at time c, and the larger is γ, the steeper the transition function will be, leading to a faster transition speed. In Figure 1, we set c fixed at halfway with γ = 20, 10, 5. Then the smooth transition function Y is a bounded continuous non-decreasing transition function with t from 1 to 44.

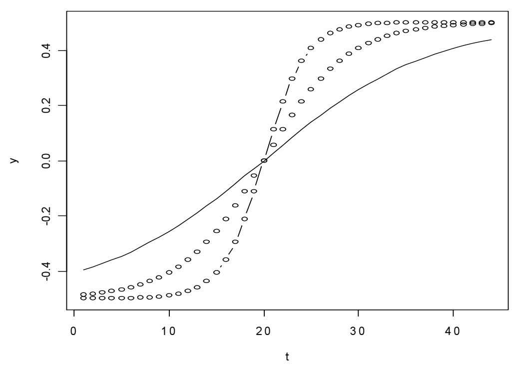

Figure 1. Logistic smooth transition functions: γ = 20 (dashed-dotted line), γ = 10 (dotted line), γ = 5 (solid line).

It is obvious that the past information of $\{Y_t\}$ will help to predict $\{X_t\}$ under the restriction of $\beta_2 \neq 0, \gamma \neq 0$. Moreover, with the causality relationship changing over time according to the logistic smooth transition function, $\{X_t\}$ is also non-linear with smooth structural change. Thus in this two variables VAR model, we have a non-linear series $\{X_t\}$ and a linear series $\{Y_t\}$, with the non-linearity of $\{X_t\}$ caused by the time varying causality dependence on $\{Y_t\}$. In the following sections, we use both linear and non-linear tests to explore the causality relationship in equation (1).

III. The DGP process and the linear causality test

Based on the procedure of the Granger linear test to test causality between $\{X_t\}$ and $\{Y_t\}$, we first regress $\{Y_t\}$ on its own past values and lag values of $\{X_t\}$ with lag length Lx, then regress $\{X_t\}$ on its own past value and lag values of $\{Y_t\}$ with lag length Ly. Then F test statistics is used to determine whether the coefficients of the past values of X_t and Y_t are zero, and insignificant coefficients imply non-causality or feedback relationship. Here for the testing procedure based on data from equation (1), as we already know the data generating process, we set Ly = Lx = 1. To test the unidirectional causality relationship from $\{Y_t\}$ to $\{X_t\}$, the test procedure is based on the following linear VAR model:

$$\begin{array}{l} x_t = u_1 + \varphi_1 x_{t-1} + \gamma_1 y_{t-1} + \varepsilon_{1t} \\ y_t = u_2 + \varphi_2 y_{t-1} + \gamma_2 x_{t-1} + \varepsilon_{2t} \end{array}, \quad \varepsilon_{1t}, \varepsilon_{2t} \sim i.i.d\ N(0,1) \tag{2}$$

Furthermore, the null hypothesis of non-causality is $H_0 : \gamma_1 = 0$. The test statistic for the linear causality test is:

$F_{linear} = \dfrac{(SSR_R - SSR_u)/m}{SSR_u/(T-K)}$, where SSR_R and SSR_U are the sum of squared errors from the restricted and unrestricted models respectively, m is the number of restrictions, T is the number of observations, and K is the number of parameters in the unrestricted model. Under H_0 of non-causality, the test statistics F_{linear} will follow a $F(m, T-k)$ distribution.

The simulated data from equation (1) are used to investigate the size and power properties of the causality test based on different parameter restrictions on the DGP. For both cases, we set μ_1 = 0.02, μ_2 = 0.03, α_1 = 0.5, α_2 = 0.5 in the linear part of the model. In order to investigate the size property of the F test statistic, we generate the DGP under the null hypothesis by setting β_1 = β_2 = 0, thus the simulated observations $\{X_t\}$ and $\{Y_t\}$ are mainly two independent series with non-causality. Here to produce the critical values for the size table, we set the number of Monte Carlo replication to 5,000. This means that we simulate $\{X_t\}$ and $\{Y_t\}$ under null hypothesis 5,000 times and carry out the Granger causality F test to find out the proportion of the rejection times. Table 1 presents the size property of the linear test for the finite sample sizes: T=25, 50, 100, 250, 500 at the 5% significant level.

Table 1. Size property for the linear Granger causality test.

T	25	50	100	250	500
Size (T)	0.0495	0.0525	0.0546	0.0530	0.0512

To judge the validity of the results, the estimated size of the test should lay between the approximate 95% confidence intervals of the nominal size 5%. With 5,000 replications, the confidence interval for the estimated size is: $0.05 \pm 1.96\sqrt{0.05(1-0.05)/5000} = (0.0441, 0.0559)$. Thus from Table 1 we can see that at the 95% confidence level, the linear Granger causality F test statistic has an unbiased size. However, the result is expected when β_1 = β_2 = 0, because the DGP is a pure linear independent modelled VAR system. What we are more interested in here is the power property of the linear test if the DGP is in a non-linear unicausality relation. In this situation, we need a DGP that satisfies the unicausality relationship. Thus in the linear part of equation (1), we still set μ_1 = 0.02, μ_2 = 0.03, α_1 = 0.5, α_2 = 0.5, but in the nonlinear part we set $\gamma = 1$, $c = T/2$, β_1 = 3 and β_2 = 3. With these fixed parameters set, there exist a unicausality relationship from $\{Y_t\}$ to $\{X_t\}$ and Figure 2 shows the structure of the simulated data $\{X_t\}$ and $\{Y_t\}$ when $T = 200$.

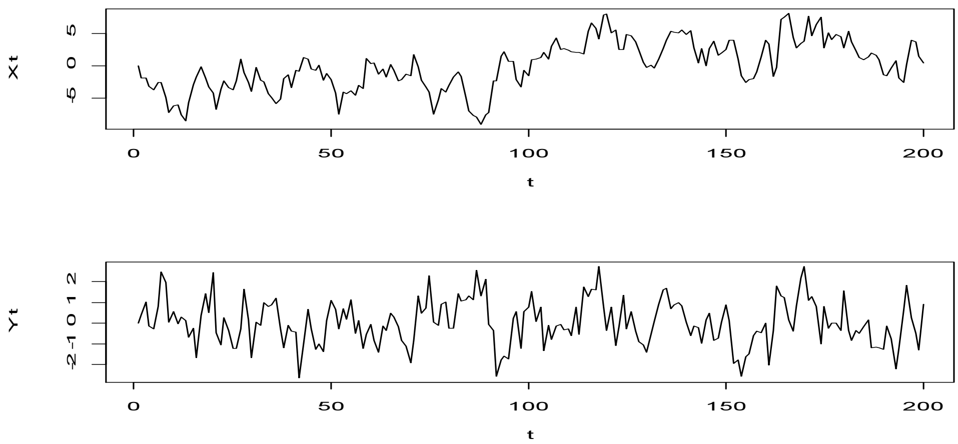

Figure 2. VAR(1) system with LSTAR non-linear structure.

Figure 2 shows that $\{X_t\}$ has a structural break at the midpoint of the sample making the whole series show non-linearity, and this is due to the structural change point c in the transition function is $T/2$. Thus the causality relationship from $\{Y_t\}$ to $\{X_t\}$ shows different patterns in the two regimes before and after time point c. The other variable $\{Y_t\}$ is still a linear time series over the entire sample and it does not include any past information of $\{X_t\}$, which means the causality relationship is just from $\{Y_t\}$ to $\{X_t\}$. In this section, we will use the above linear Granger F test to investigate its power properties for the finite sample sizes. Here, with m = 1, K = 6 and different sample sizes T, at the 5% significant level, we replicate the test procedure 5,000 times based on the DGP under restriction of μ_1 = 0.02, μ_2 = 0.03, $\alpha_1 = \alpha_2$ = 0.5; $\gamma = 1$, $c = T/2$, $\beta_1 = \beta_2$ = 3, and we get the power table, Table 2 below, for the linear Granger test.

Table 2. Power properties for the linear Granger causality test.

T	25	50	100	250	500
Power (T)	0.116	0.109	0.102	0.097	0.095

Table 2 shows that the linear Granger test has very low power in testing the causality relationship when the lag value of $\{Y_t\}$ influences $\{X_t\}$ according to a non-linear structure. This is not surprising as the non-linear structure will result in an insignificant parameter γ_1 in equation (2), which reduces the power of the linear Granger F test. It proves that although the linear Granger causality test may perform well in testing linear prediction among the variables, if there are some factors that cannot be modelled by a linear function, the causality relationship may not be exposed by the linear causality test, because the linear coefficients of the lag value is not significant in the linear model. As we can see in our case, when there exists a non-linear causality relationship, the linear test has very low power in exploring it, although one of the series $\{Y_t\}$ still shows linearity.

IV. The non-parametric non-linear test

For the STAR model, there is a parametric test for causality proposed by Dimitris and Miguel (2008) where they used the Taylor expansion method to test the time varying causality relationship between variables. It first tests the linearity of the variables, and if linearity is rejected, a F test statistic is proposed after the estimation of the non-linear parameters. Moreover, Li (2006) proposed a heteroscedasticity-robust Wald test. This joint test of the threshold and causality leads to a non-standard asymptotic distribution because of the nuisance parameters. These two test methods basically follow the logic of the minimal variance error of prediction, as in the linear Granger test. Here, we switch to the non-parametric test that was first proposed by Baek and Brock (1964), and later modified by Hiemstra and Jones (1994). The main idea of the test is illustrated as follows:

Consider two strictly stationary and weakly dependent time series $\{X_t\}$ and $\{Y_t\}$ that satisfy the ergodicity condition. Denote the m-length lead vector of $\{X_t\}$ as X_t^m, the Lx-lag length vector of $\{X_t\}$ as X_{t-Lx}^{Lx}, and the Ly-length of $\{Y_t\}$ as Y_{t-Ly}^{Ly}. For given values of m, Lx, $Ly \geq 1$, and $e > 0$, $\{Y_t\}$ does not strictly Granger cause $\{X_t\}$ will be expressed as:

$$\begin{aligned}&\Pr(\left\|X_t^m - X_s^m\right\| < e \mid \left\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\right\| < e, \left\|Y_{t-Ly}^{Ly} - Y_{s-Ly}^{Ly}\right\| < e)\\ &= \Pr(\left\|X_t^m - X_s^m\right\| < e \mid \left\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\right\| < e)\end{aligned} \tag{3}$$

Where $\Pr(\cdot)$ measures the probability and $\|\cdot\|$ measures the vector distance, which is the maximum norm of the two vectors. Thus $\Pr(\left\|X_t^m - X_s^m\right\| < e \mid \left\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\right\| < e, \left\|Y_{t-Ly}^{Ly} - Y_{s-Ly}^{Ly}\right\| < e)$ in equation (3) is the conditional probability that the distance of two arbitrary m-length lead vectors X_t^m, X_s^m is within the distance e, given that the corresponding Lx and Ly-length lag vectors of $\{X_t\}$ and $\{Y_t\}$ are within distance of e. $\Pr(\|X_t^m - X_s^m\| < e \mid \|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\| < e)$ is the conditional probability without taking the lag vector of $\{Y_t\}$ into consideration. The test based on equation (3) is a non-causality test that implies for the given criteria measurement which is represented by the vector distance e, the Ly-length of the number of lag values of $\{Y_t\}$ does not help predict the future period values of $\{X_t\}$, given the Lx-length of the number of lag values of $\{X_t\}$. Generally speaking, this non-parametric causality test tests for non-causality in density which is different from the Granger linear test, which tests non-causality in mean. Non-causality in mean depends on the properties of the prior model, which needs stronger hypothesis. Thus, this non-parametric test is more robust to the data, and can be used to test the causality relationship when the data structure shows non-linearity.

For further construction of the test statistic we can rewrite equation (3) into the following equation:

$$\frac{\Pr_1(\|X_t^m - X_s^m\| < e, \|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\| < e, \|Y_{t-Ly}^{Ly} - Y_{s-Ly}^{Ly}\| < e)}{\Pr_2(\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\| < e, \|Y_{t-Ly}^{Ly} - Y_{s-Ly}^{Ly}\| < e)} = \frac{\Pr_3(\|X_t^m - X_s^m\| < e, \|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\| < e))}{\Pr_4(\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\| < e)} \tag{4}$$

Moreover, we can use the correlation integral estimators C_1, C_2, C_3, C_4 as the estimators of probabilities $\Pr_1$, $\Pr_2$, $\Pr_3$, $\Pr_4$, where when we set $I(Z_1, Z_2, e)$ as the index function of vectors Z_1, Z_2 within maximum norm distance e, we have:

$$\begin{aligned} C_1 &= \frac{2}{n(n-2)} \sum_{t<s}\sum I(x_{t-L_x}^{m+L_x}, x_{s-L_x}^{m+L_x}, e) * I(y_{t-L_y}^{L_y}, y_{s-L_y}^{L_y}, e), \quad C_3 = \frac{2}{n(n-2)} \sum_{t<s}\sum I(x_{t-L_x}^{m+L_x}, x_{s-L_x}^{m+L_x}, e) \\ C_2 &= \frac{2}{n(n-2)} \sum_{t<s}\sum I(x_{t-L_x}^{L_x}, x_{s-L_x}^{L_x}, e) * I(y_{t-L_y}^{L_y}, y_{s-L_y}^{L_y}, e), \quad C_4 = \frac{2}{n(n-2)} \sum_{t<s}\sum I(x_{t-L_x}^{L_x}, x_{s-L_x}^{L_x}, e) \end{aligned} \tag{5}$$

and equation (4) turn into :

$$\frac{C_1(m+Lx, Ly, e)}{C_2(Lx, Ly, e)} = \frac{C_3(m+Lx, e)}{C_2(Lx, e)}. \tag{6}$$

Thus the non-causality relationship implies that $\{Y_t\}$ and $\{X_t\}$ should satisfy equation (6), and a natural test statistic is:

$$t = \frac{C_1(m+Lx, Ly, e)}{C_2(Lx, Ly, e)} - \frac{C_3(m+Lx, e)}{C_2(Lx, e)}.$$

Under H_0 of non-causality between $\{Y_t\}$ and $\{X_t\}$, the asymptotic distribution of the test statistic is: $\sqrt{T}t \to N(0, \sigma^2(m, Lx, Ly, e))$. A detail derivation of the asymptotic distribution of the test statistic and estimator of $\sigma^2(m, Lx, Ly, e)$ can be found in Baek and Brock (1992) and Hiemstra and Jones (1994).

Note that this non-parametric test is always implemented using the linear VAR model with filtered residuals of the original data, while what we are testing is the pure 'non-linear predictive' power of the causality. The logic is that the linear predictive element can be extracted by the pre-chosen linear VAR model and the residuals will retain the non-linear structure of the original series. However, this procedure requires caution as the filter processes depend heavily on the right choice of the linear VAR model. The reason is that a misspecified VAR model may withdraw 'too much' non-linear information or 'too little' linear information of the dependence relationship in the original data, and leave the residual not informative enough to illustrate the non-linear causality relationship, which might lead to lost power for this non-parametric test. Hiemstra and Jones (1994) showed that the asymptotic distribution of the test statistics is the same when we use the original data or the VAR filtered residuals. As the asymptotic distribution needs further strict assumptions to ensure good power and size properties, here we only discuss the test properties for the finite samples. We now apply the test to both the original data and the linear VAR filtered residuals, and then compare the power properties. To carry out the tests in small samples, we first use the Monte Carlo experiment to generate the critical value table for the test in both situations using the DGP system satisfies the null hypothesis as follows:

$$\begin{aligned} x_t &= 0.03 + 0.5x_{t-1} + \varepsilon_{1t} \\ y_t &= 0.02 + 0.5y_{t-1} + \varepsilon_{2t} \end{aligned} \quad \text{where } \varepsilon_{1t}, \varepsilon_{2t} \sim i.i.d\ N(0,1). \qquad (7)$$

The above DGP system simulates two independent series $\{Y_t\}$ and $\{X_t\}$. Thus the critical values of the test statistic $t = \dfrac{C_1(m+Lx, Ly, e)}{C_2(Lx, Ly, e)} - \dfrac{C_3(m+Lx, e)}{C_2(Lx, e)}$ are constructed from the data generated by this DGP system which satisfies the null hypothesis of non-causality. Using a 10,000 Monte Carlo replications, we get the following critical value table for the finite sample sizes T=25, 50, 100, 250, 500:

Table 3. Critical values for the non-parametric test based on original data.

T	99%	97.5%	95%	90%	10%	5%	2.5%	1%
25	–0.9354	–0.7559	–0.6156	–0.4585	0.4249	0.5642	0.7002	0.8680
50	–0.7793	–0.6168	–0.5058	–0.3816	0.3324	0.4347	0.5342	0.6437
100	–0.6254	–0.5150	–0.4309	–0.3296	0.2945	0.3880	0.4750	0.5820
250	–0.5416	–0.4639	–0.3868	–0.2944	0.2600	0.3375	0.4108	0.4805
500	–0.5129	–0.4251	–0.3569	0.2751	0.2552	0.3310	0.3987	0.4770

Table 3 shows that under H_0, the test statistic in finite samples is distributed symmetrically around 0, which is the expected value based on the relationship from equation (4). Moreover, the QQ plot (not included here to save space) shows that the distribution of the test approaches to the normal distribution as the sample size grows. Since the size property of the test here is unbiased, we only need to examine the power property. When we perform the power investigation for the non-parametric test, we use the same DGP as in the power investigation procedure in section III, that is the data is from the non-linear LSTAR models in equation (1) under the restrictions $\gamma = 1$, $c = T/2$, β_1 = 3, β_2 = 3 which leads to the unidirection causality from $\{Y_t\}$ to $\{X_t\}$. A significant positive test statistic suggests that $\{Y_t\}$ helps in predicting $\{X_t\}$ while a significant negative value implies that $\{Y_t\}$ confounds the prediction of $\{X_t\}$, and hence we only use the right tailed critical values. Based on the 5,000 simulated values from the DGP with unidirection Granger causality from $\{Y_t\}$ to $\{X_t\}$ and the critical values at the 5% significance level from Table 3, we obtain the power of the non-parametric test based on the original data.

Table 4. Power properties for the non-parametric causality test based on original data.

T	25	50	100	250	500
Power (T)	0.142	0.228	0.358	0.6325	0.836

Table 4 shows that for T=25, 50 and 100, the non-parametric test has low power. However when T increases to 250 and 500, the power increases significantly. When comparing Table 4 with Table 2, the non-parametric test shows an obvious power improvement, which proves that the non-parametric test performs better than the traditional linear Granger test when the series has a non-linear structure.

We next apply the non-parametric test to the VAR model filtered residuals. A VAR model with one lag is considered as a reasonable pre-chosen model to extract the linear structure of the original data, although the same procedure can be applied to higher order VAR models. The filter procedure is as follows: we first use a linear VAR model with one lag to model the simulated data from the independent model based on equation (7), and each equation in (7) generates a series of residuals. Then we rescale the two series of residuals and use them to construct the test statistic. In the later step when we carry out the test to investigate the power, we use the linear filtered residuals as well. Then, based on the VAR model filtered residuals of data from equation (7), we get the following critical values table by Monte Carlo simulations with 10,000 simulations:

Table 5. Critical values for the non-parametric test based on residuals.

T	99%	97.5%	95%	90%	10%	5%	2.5%	1%
25	–0.7676	–0.6150	–0.5057	–0.3799	0.3702	0.4896	0.5948	0.7189
50	–0.5341	–0.4422	–0.3733	–0.2886	0.2793	0.3507	0.4244	0.5032
100	–0.4233	–0.3592	–0.3039	–0.2361	0.2317	0.2950	0.3548	0.4256
250	–0.3726	–0.3090	–0.2630	–0.2049	0.2013	0.2590	0.3117	0.3805
500	–0.3575	–0.3051	–0.2514	–0.1975	0.1892	0.2424	0.2898	0.3504

Table 5 shows the same distribution characteristics as Table 3 with the approach to normal distribution with expected value of zero and shorter interval for larger sample sizes. Furthermore, to investigate the power property, again we use the DGP from equation (1), under the restrictions of $\gamma = 1$, $c = T/2$, β_1 = 3, and β_2 = 3 which leads to causality relationship from $\{Y_t\}$ to $\{X_t\}$, and using the rescaled residual after the linear VAR filtering, we get the following power table based on 5,000 Monte Carlo simulations :

Table 6. Power property for the non-parametric causality test based on residuals.

T	25	50	100	250	500
Power (T)	0.0645	0.1045	0.167	0.35	0.593

For small samples T=25 and 50, the test based on the residuals has even less power than the linear Granger test. Although the power increases for larger sample sizes and it exceed the power in linear test, it is still lower than the test based on the original data. This outcome may be due to that when the sample is too small, the non-linear structure in the model is not obvious and a pre-chosen linear VAR model extracts all the dependency structure in the data, which leads to a low power if we use the residuals for the test. It shows that when we use the residuals in the test after the linear filter process, the pre-used linear VAR filter model may extract too much information about the dynamics between the variables and destroy their dependency, thus when we run the test based on the residuals, we get very weak power.

V. The non-parametric test based on the wavelet multiresolution

Section IV presents the non-parametric non-linear test based on the original data and the linear VAR model filtered residuals, with the first one performing much better (in terms of higher power) than the latter one for the larger sample size, which remind us that the residual based methods are sensitive to the specification of the pre-chosen linear filter model. However, if the residual is used to maintain the non-linear relationship of the original data, we can consider another alternative method of achieving this. This alternate method captures the non-linear characteristics of the original data using the wavelet multiresolution analysis (MRA). This frequency decomposition method has been widely used after its theoretical development in the 1980s (see e.g. Mallat 1989). In signal smoothing and spectrum analysis Chiann and Morettin (1998) showed how wavelet captures signals in different scales by wavelet spectrum decomposition. In economics, Schleicher (2002) also mentioned that the wavelet method can catch macroscopic behaviour as well as the microscopic detail in economic areas. In perspective of nonlinear models, the wavelet decomposition is an efficient method due to its ability of localizing the non-stationary structure which depending on time. Moreover, it is of particular use in the identification of non-linear models, see Coca and Billings (2001) who use wavelet multiresolution to process the nonlinear system in NARMAX models. Chang and Shi (2009) also use this methodology to identify time-varying properties of hysteretic structures, and a comprehensive elaborate of MRA can be found in Carl (2005). In our circumstance, we expect the low frequency wavelet smooth based on MRA to capture the main trend of the data, which is the non-linear LSTAR structure in the VAR system, leading to a better performance of the test. We use the maximal overlap discrete wavelet transform (MODWT) as it has no restriction on the sample size. We first start with a brief introduction of the MODWT transform and wavelet multiresolution, and more details are to be found in the cited references.

For an N dimensional vector $\mathrm{X}=\{X_t, t=0,...,N-1\}$, the level J MODWT transform of X contains $J+1$ vectors $\mathbf{W}_1,...,\mathbf{W}_J,\mathbf{V}_J$ with the wavelet coefficients $\mathbf{W}_j$ corresponding to changes on scale $\tau_j = 2^{j-1}$, while the wavelet scaling coefficients $\mathbf{V}_J$ corresponds to average on scale $\lambda_J = 2^j$. The N dimension vectors $\mathbf{W}_j$ and $\mathbf{V}_J$ are computed by $\mathbf{W}_j = \mathcal{W}_j\mathrm{X}$, $\mathbf{V}_J = \mathcal{V}_J\mathrm{X}$ where $\mathcal{W}_j$ and $\mathcal{V}_J$ are $N\times N$ matrixes. Then the MODWT based MRA of X is defied as: $\mathrm{X}=\sum_{j=1}^{J}\mathcal{W}_j^T\mathbf{W}_j+\mathcal{V}_J^T\mathbf{V}_J=\sum_{j=1}^{J}D_j+S_J$, where D_j is the j^{th} level MODWT details containing the microscopic detail of X and S_J is the J^{th} level MODWT smooth containing landscape characteristics of X. Thus the MODWT transformation and multiresolution can be viewed as a band-pass filter process of X, and based on different transformation matrices $\mathcal{W}_j$ and $\mathcal{V}_J$, we have different choices of filters. For more information about the MODWT method and how to choose the suitable filter, we refer to Vidakovic (1999), Percival and Walden (2000), and Gençay *et al.* (2001). An important issue now is how to choose the wavelet filter. A central factor to use a particular wavelet is to match the characteristics of the series under consideration. The number in the name of the wavelet indicates the width of the filter. In general, the wavelets with small L are narrower and less smooth, while wavelets with large L, are relatively wide and smooth. However, in this paper we use the wavelet Least Asymmetric with L = 8, i.e. LA(8), since it has better band pass characteristics.

Thus by MRA, the wavelet smooth S_J can capture the low frequency trend of the original data, which we use to capture the nonlinear and linear structures of $\{X_t\}$ and $\{Y_t\}$ for the nonparametric test, instead of using the VAR filtered residuals. Here, and to simplify the procedure, we simply let $J=1$ and get the low frequency wavelet smoothes $\{S_{1t}\}$, $\{S_{2t}\}$ of the original time series $\{X_t\}$, $\{Y_t\}$:

$S_{1t} = \sum_{l=0}^{L-1} g_l^{N} V_{1t+l \bmod T}$, $S_{2t} = \sum_{l=0}^{L-1} g_l^{N} V_{2t+l \bmod T}$ with the MODWT scale coefficients $V_{1t} = \sum_{l=0}^{L-1} g_l X_{t-l \bmod T}$, $V_{2t} = \sum_{l=0}^{L-1} g_l Y_{t-l \bmod T}$, while $\{g_l\}$ is the scaling filter satisfying $\sum_l g_l = 1$, $\sum_l g_l^2 = 1/2$, $\sum_l g_l g_{l+2*n} = 0$ and $\{g_l^{N}\}$ is the periodization of $\{g_l\}$ to circular filters of length N. Since $\{g_l\}$ and $\{g_l^{N}\}$ are the low band pass filters, the resulting wavelet smooth maintains the low frequency structure of the original data which is the non-linear structure of $\{X_t\}$ and the linear character of $\{Y_t\}$ in our case. Thus, the corresponding wavelet smooth of the data presented in Figure 2 is now shown in Figure 3.

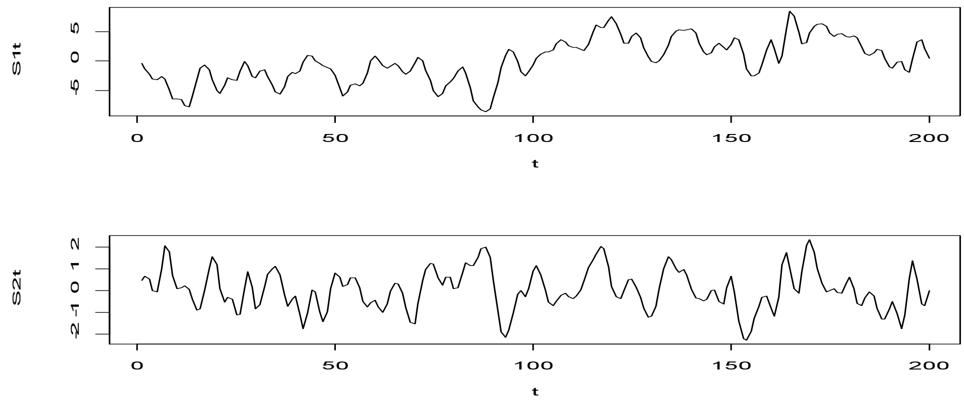

Figure 3. Wavelet smooth of the VAR(1) system with LSTAR non-linear structure.

Figure 3 shows that the wavelet smooth (which is the reconstructed signal from the scale coefficients) is smoother compared with original data in Figure 2 with a clearer non-linear structural in $\{X_t\}$. Then if we use the wavelet smooth, we can keep the main structure of the original data, regardless of whether it is linear or non-linear. Especially if the original data have a non-linear structure, compared with the linear VAR filtered residuals, the wavelet smooth will maintain the non-linear information as well as the dependence relationship between $\{X_t\}$ and $\{Y_t\}$. Now to carry out the non-parametric causality test, when creating the critical value table, we first simulate the same DGP process as in the last sections (i.e. equation 7) which satisfies the null hypothesis that is μ_1 = 0.02, μ_2 = 0.03, $\alpha_1 = \alpha_2$ = 0.5 and $\beta_1 = \beta_2$ = 0, but instead of using the original data or the linear filtered residuals, we use the wavelet smoothes S_{1t} and S_{2t} in the test.

The test statistic is now: $t_s = \dfrac{C_1(m+Ls_1, Ls_2, e)}{C_2(Ls_1, Ls_2, e)} - \dfrac{C_3(m+Ls_1, e)}{C_2(Ls_1, e)}$,

Where, Ls_1 and Ls_1 are the lag indices of S_{1t} and S_{2t}, respectively, corresponding to Lx and Ly as the lag indices of $\{X_t\}$ and $\{Y_t\}$ in the original test statistic in section IV. Then based on 10,000 Monte Carlo replications, we get the critical values table for the test based on S_{1t} and S_{2t} as follows:

Table 7. Critical values for the non-parametric test based on wavelet smooth.

T	99%	97.5%	95%	90%	10%	5%	2.5%	1%
25	-0.6478	-0.5313	-0.4226	-0.3103	0.3657	0.4815	0.5803	0.7019
50	-0.5425	-0.4473	-0.3593	-0.2696	0.2876	0.3767	0.4558	0.5396
100	-0.4574	-0.3734	-0.3115	-0.2378	0.2352	0.3099	0.3889	0.4648
250	-0.3844	-0.3087	-0.2590	-0.1983	0.2017	0.2599	0.3126	0.3894
500	-0.3501	-0.2964	-0.2476	-0.1952	0.1855	0.2460	0.2914	0.3599

Table 7 also shows symmetric, zero centred distribution which approaches the normal distribution as the sample size grows. This implies that the asymptotic distribution of the test statistic based on the wavelet smooth is still normal distribution. This is obvious because here in the wavelet environment we use the same form of the non-parametric test statistic $t = C_1/C_2 - C_3/C_4$ with the only difference is that we, instead of the original $\{X_t\}$ and $\{Y_t\}$, use $\{S_{1t}\}$ and $\{S_{2t}\}$ in C_i. Thus the test statistic is:

$$t_s = \frac{C_1(m+Ls_1, Ls_2, e)}{C_2(Ls_1, Ls_2, e)} - \frac{C_3(m+Ls_1, e)}{C_2(Ls_1, e)},$$

where its asymptotic distribution satisfies $\sqrt{T} t_s \rightarrow N(0, \sigma^2(m, Ls_1, Ls_2, e))$. A detailed explanation of $\sigma^2(m, Ls_1, Ls_2, e)$ can be found in Baek and Brock (1992) and Hiemstra and Jones (1994).

For the finite sample size, and to investigate the power property, we still use the same DGP from equation (1) under the restrictions of $\gamma = 1$, $c = T/2$, β_1 = 3, and β_2 = 3 to generate $\{X_t\}$ and $\{Y_t\}$ which satisfies nonlinear causality relationship, and using their wavelet smooth S_{1t} and S_{2t} to carry out the test. Based on the critical value at 5% significant level in Table 7 and 2,000 Monte Carlo replications, we get the power property of the non-parametric test in wavelet environment, see Table 8 below.

Table 8. Power properties for the non-parametric causality test based on a wavelet smooth.

T	25	50	100	250	500
Power	0.166	0.358	0.657	0.975	1

Although the power is still low when the sample size $T = 25$, the power for other sample sizes are much higher comparing with Tables 4 and 6. When T goes to 250, the power approaches almost one. Thus based on the power table, the wavelet smooth performs best in extracting the non-linear characteristics and the dependency relationship of the original data. Moreover, the wavelet method has the advantage of being easy to perform as it saves us the effort of choosing the correct specification of the VAR linear model, compared with the residual-based test. We can also asses its asymptotic distribution as we only need to put the wavelet smooth, instead of original data, into the test statistic and we still retain the asymptotic normal distribution with zero mean value but with different variance. Further more, as wavelet smooth is essentially the out coming of low pass filter which is actually in terms of weighted moving average, it deserves more exploration of using other smoothing or moving average methods for further comparison. Thus we carry out more simulations to invest power properties for more smoothing methodologies which are LOWESS(Local weighted scatter plot smoothing) and HP (Hodrick-Prescott) filter for their robustness to the nonlinear structure of the LSTAR series. Unfortunately when applying these two methods, we come across two main problems, first is that the corresponding simulation results show that power properties are very sensitive to the smooth restriction criterion, such as the span parameter in LOWESS and the multiplier λ in HP filter, and our simulation outcomes reveal that "mis-specified" smooth parameters will lead to very low power. However, in empirical implications, as we may know scarce information of the original data, it is always not easy to choose a reasonable smooth parameter at the first sight and the result maybe serious distorted. The second problem is that as it is not easy to get a explicate representation of the smoothed series from the LOWESS and HP filter, we can not derive further asymptotic distributions of the test statistics. However when turning back to the wavelet smooth method, we find it is quite straightforward to implicate and here just choosing the first level smooth, we capture the main nonlinear structure as well as maintain the causality information of the data, and we also get a power property which actually overmatch all the results from LOWESS and HP filter (We do not show the simulation tables here as it will take too much space if we present all the tables based on different smooth parameters). What's more, the explicit representation of the wavelet smooth S_t leads to the possibility for derivation of the asymptotic distributions of the smoothed series.

VI. **Conclusion**

In this paper, we investigated the power properties of causality tests when the causality relationship is characterized by a LSTAR model. We compared power values from the linear causality test, the non-parametric test based on the original data, the VAR filtered residuals and the wavelet smooth based on MRA for which the best power property is obtained. The low power property in the residual-based test might be due to that the linear model we used for filtering extract too much information and thus destroy the dependency relationship that is supposed to be retained in the residuals. However, all the tests here use the data for the whole period for testing, although there are data breaks and the causality relationship may vary with time. Li (2006) proposed a test that can test causality before and after breaks, but with more complicated procedure. Thus for the VAR model with LSTAR non-linear characteristics, the non-parametric test can be used for a first examination of the causality relationship between the variables. Furthermore, when applying the non-parametric test, the simulation results show that when we try to identify the non-linear characteristics, the VAR filtered residuals should be used carefully, while on the other hand the non-parametric test based on wavelet multiresolution can capture the non-linear relationship well which leads to the best power property.

References

Abhyankar, A. (1998) "Linear and Non-linear Granger Causality: Evidence from the FT–SE 100 Stock Index Futures and Cash Markets". *Journal of Futures Markets*. Vol.18, pp.519–540.

Ancona N., Marinazzo D. and Stramaglia S. (2004) "Radial Basis Function Approach to Non-linear Granger Causality of Time Series". *Physical Review* E 70.

Baek E.G. and Brock W.A. (1992) "A general Test for Non-linear Granger Causality: Bivariate Model". *Technical report. Iowa State University and University of Wisconsin, Madison.*

Bell D., Key J., Malley J. (1996) "A Non–parametric Approach to Non–linear Causality Testing". *Economics Letter*. Vol.51, pp.7–18.
Carl R. (2005) *Nonlinear Smoothing And Multiresolution Analysis*. Birkhäuser Basel
Chang C.C and Shi Y.F (2009) "Identification of time varing hysteretic structure using wavelet multiresolution analysis". *International Journal of Non-Linear Mechanics.* In press.
Chen Y., Rangaraja G., Feng J. and Ding M. (2004) "Analyzing multiple non-linear time series with extended Granger causality". *Physics Letters A*, 324, 26–35.
Chiann C. and Morettin P.A. (1998) "A Wavelet Analysis for Time Series". *Journal of Nonparametric Statistics,* Vol. 10, pp.1-46.
Coca D and Billings S. A. (2001) "Non-linear system identification using wavelet multiresolution models". *International Journal of Control,* Vol. 74, Issue 18 , pp. 1718 – 1736.
Dimitris K.C. and Miguel A.L. (2008) "Testing for Granger (Non–) Causality in A Time Varying Coefficient VAR model". *Journal of Forecasting.* Vol.27, Number 4.
Gençay R., Selçuk F. and Whitcher, B. (2001) *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press, San Diego.
Hiemstra C. and Jones J. (1994) "Testing for Linear and Non-linear Granger Causality in the stock price–volume relation" *Journal of Finance.* Vol.49, number5, pp.1639–1664.
Huh H. (2002) "GDP Growth and the Composite Leading Index: a Non-linear Causality Analysis for Eleven Countries". *Economics Letters.* Vol. 77, pp.93–99.
Li J. (2006) "Testing Granger Causality in the Presence of Threshold Effects". *International Journal of Forecasting.* Vol.22, pp.771–780.
Mallat S.G.. (1989) "A theory for multiresolution signal decomposition: the wavelet representation".*Pattern Analysis and Machine Intelligence, IEEE Transactions*. Vol. 11, Issue. 7, pp. 674–693.
Percival D.B. and Walden A.T. (2000) *Wavelet Methods for Time Series Analysis.* Cambridge Press.
Schleicher C. (2002). "An Introduction to Wavelets for Economists". *Working paper,* Bank of Canada , series: *2002–3.*
Teräsvirta T. (1994) "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models". *Journal of the American Statistical Association,*Vol. 89, pp. 208–218.
Vidakovic B. (1999) *Statistical* Modeling *by Wavelets*. New York: John Wiley and Sons.