A Bootstrap Test for Causality with Endogenous Lag Length Choice

- theory and application in finance

R. Scott Hacker and Abdulnasser Hatemi-J

April 2010

A Bootstrap Test for Causality with Endogenous Lag Length Choice:

Theory and Application in Finance

R. Scott Hacker

Jönköping International Business School, Jönköping University, Sweden

E-mail: Scott.Hacker@jibs.hj.se

and

Abdulnasser Hatemi-J

Department of Economics and Finance,UAE University, UAE,

and Department of Statistics, Lund University

E-mails: AHatemi@uaeu.ac.ae, and Abdulnasser.Hatemi-J@stat.lu.se

Abstract

Granger causality tests have become among the most popular empirical applications with time series data. Several new tests have been developed in the literature that can deal with different data generating processes. In all existing theoretical papers it is assumed that the lag length is known *a priori*. However, in applied research the lag length has to be selected before testing for causality. This paper suggests that in investigating the effectiveness of various Granger causality testing methodologies, including those using bootstrapping, the lag length choice should be endogenized, by which we mean the data-driven preselection of lag length should be taken into account. We provide and accordingly evaluate a Granger-causality bootstrap test which may be used with data that may or may not be integrated, and compare the performance of this test to that for the analogous asymptotic test. The suggested bootstrap test performs well and appears to be also robust to ARCH effects that usually characterize the financial data. This test is applied to testing the causal impact of the US financial market on the market of the United Arab Emirates.

1. Introduction

Tests for causality in Granger's (1969) sense are increasingly conducted in applied research when time series data are used. This is especially the case for empirical studies in economics and finance. Several new methods have been put forward in the literature for testing causality that can deal with different data generating processes; see among others Granger (1988) and Toda and Yamamoto (1995). Hacker and Hatemi-J (2006) investigated the size (Type-I error) properties of Granger causality testing using a Wald test both with a modification for integrated variables suggested in Toda and Yamamoto (1995) and without that modification, along with bootstrap version the Wald test (again with and without the Toda and Yamamoto modification). That paper showed through simulations that the Toda and Yamamoto (1995) methodology is an important tool in dealing with integrated data. It also showed that the bootstrapping provides improvements with small sample sizes.

A common factor in all existing theoretical papers that we know of on this topic is that the lag length is assumed to be known *a priori*, which is not an appropriate assumption for most applications. It is a common practice in the empirical literature to select the optimal lag order based on the data each time tests for causality are conducted. However, the preselection of the lag order using the data may affect the distribution of the test statistic. This paper suggests we should take into consideration this preselection of the lag order, which we refer to as endogenous lag length choice, in investigating the size and power properties of Granger causality tests, including those using bootstrapping.

We provide a bootstrap test based on the Granger causality Wald test that, when necessary, incorporates the Toda and Yamamoto (1995) methodology if integrated variables are considered to be present. This test was also provided in Hacker and Hatemi-J (2006). In the current paper, simulation results are provided which take into account endogenous lag length choice when comparing the performance of this bootstrap test to an analogous test using a chi-square distribution based on asymptotic theory. This bootstrap methodology appears to have better size properties than its asymptotic-distribution counterpart and is robust to the existence of autoregressive conditional heteroscedasticity (ARCH). This latter property of the test seems to be useful because it is widely agreed in the literature, especially since the pioneering work by Engle (1982), that the volatility of many economic and financial variables is time-varying and ARCH effects usually prevail. We also investigate the power

properties in some situations which show that the bootstrap test has similar or better power properties compared to asymptotic test for the same actual size.

The rest of the paper continues as follows. Section 2 defines Granger causality and discusses the Toda and Yamamoto (1995) methodology for testing it in the presence of integrated data. Section 3 focuses on the issue of choosing the optimal lag length and its implication for inference. Section 4 presents our bootstrap methodology. Sections 5 and 6 discuss the methods and results of the simulations, respectively dealing with the size and power of various Granger-causality test methodologies. An application on financial market integration is illustrated in Section 7, and the conclusions are given in the last section.

2.   Granger Causality Testing

For the purpose of investigating Granger causality, we consider the following vector autoregressive model of order k, VAR(k):

$$y_t = B_0 + B_1 y_{t-1} + \ldots + B_k y_{t-k} + u_t, \tag{1}$$

where y_t, B_0, and u_t are vectors with dimensions $n \times 1$ and B_i, $i \geq 1$ is a parameter matrix with $n \times n$ dimensions. The error vector, u_t, has a zero-expected value and it is assumed to be independent and identically distributed with a non-singular covariance matrix Ω that fulfills the $E|u_{it}|^{2+\lambda} < \infty$ condition for some $\lambda > 0$, with u_{it} being the ith element of u_t (this assumption is needed for appropriate testing conditions). There is non-Granger causality of the rth element of y_t on the jth element of y_t only if the subsequent statement is correct:

$$H_0: \text{the element in } B_i\text{'s row } j, \text{ column } r \text{ is zero for } i = 1, \ldots, k. \tag{2}$$

Toda and Yamamoto (1995) suggest that if one thinks that the variables in equation (1) are integrated of order d, then a legitimate way to proceed is to estimate the equation in (1) with d additional lags in the estimated VAR model (in addition to the k lags), and perform a Wald test with the null hypothesis constraints for no Granger causality applying on the B_i coefficient matrixes for only $i = 1, \ldots, k$. Thus the coefficient matrixes for the last d lags in the estimated VAR model would be unconstrained under the null hypothesis. We refer to the d

additional lags as "augmentation lags" in this paper. The Wald test methodology is described in the appendix.[1] That methodology is followed in our simulations.

If the null hypothesis in a Wald test is true, then asymptotically the resulting Wald test statistic is χ^2 distributed with the number of restrictions under the null providing the number of degrees of freedom. In the case of Granger causality testing with only two variables in the VAR model, the degrees of freedom is equal to the lag order k, regardless of whether the Toda and Yamamoto augmentation lags are included, since the coefficient estimates for those additional lags are unconstrained under the null hypothesis.

3. Determination of Lag Length, and Its Effect on Inference

Typically applied researchers do not know what the lag order k is and therefore use the data to help determine the "optimal lag order". In this paper we assume for our simulations and for our bootstrapping the optimal lag order is determined by estimating the VAR(k) model in equation (1) for $k = 0, \ldots, K$, where K is the maximum lag length considered, and finding that k which minimizes the multivariate version of the Schwarz Bayesian Criterion, SBC. This information criterion is defined as

$$
\text{SBC} = \ln\left(\det \widehat{\Omega}_j\right) + k\left(\frac{n^2 \ln T}{T}\right),
\tag{3}
$$

where $\det \widehat{\Omega}_k$ is the determinant of the estimated variance-covariance matrix of error term vector u_t when the VAR model is estimated using lag order k, and T is the number of observations (time periods) utilized to estimate the VAR model.

The fact that applied researchers have finite sample sizes, and often reasonably small ones, calls into question whether using an asymptotic distribution for making inference with a test statistic (e.g. using the χ^2 distribution for making inference from the Wald test statistic) is

---

[1] Lütkepohl (1991) describes the methodology for performing a Wald test with linear constraints in a VAR model. He uses the example of Granger causality with none of the additional lags as suggested by Toda and Yamamoto (1995), but the methodology can clearly be generalized to any set of linear constraints on the VAR system.

appropriate. This problem is compounded by the fact that these researchers also often resort to information-criterion minimization or some other data-driven technique to determine optimal lag order. Bootstrapping, the subject of the next section, can deal with these issues.

4. The Bootstrapping Methodology with Endogenous Lag Length

We suggest that with reasonably small sample sizes, bootstrapping the two-stage process of researcher—first finding the optimal lag length and then finding a Wald statistic for testing Granger causality—can provide a decent distribution for considering the significance of a Wald statistic.[2] In this section we describe explicitly the type of bootstrapping we are considering.

The first stage of that bootstrapping process involves the estimation of equation (1) using the optimal lag order k (determined by the methodology discussed in the previous section) without imposing any restriction implied by the null hypothesis of non-causality. Then, we generate the simulated data, denoted by y_t^*, as the following:

$$y_t^* = \hat{B}_0 + \hat{B}_1 y_{t-1} + \ldots + \hat{B}_k y_{t-k} + \hat{u}_t^*$$

for the period $t = 1, \ldots, T$, where the circumflex above a parameter matrix represents the estimate of that parameter matrix in the first step, and $\hat{u}_t^*$ is a vector of bootstrapped error terms. The set of T bootstrapped error-term vectors is found by drawing randomly with replacement from the vectors of modified residuals of the regression (defined below), with equal probability of drawing any of those vectors, and for each element in the error vector, subtracting the mean of the associated drawn modified residuals from each of those modified residuals (to make sure that the mean of the bootstrapped error term vectors is a zero vector).

---

[2] The bootstrap approach, which is a data resampling method used to approximate the distribution of a test parameter, was popularized by Efron (1979).

The regression's modified residuals are the raw residuals modified via *leverages*[3] to have constant

variance (this modification can improve the bootstrapping under circumstances of heteroscedasticity).

Further notation helps clarify the leverage modification. We define Y_{-P} to be $(y_{1-P}, \cdots, y_{T-P})$ and

$Y_{i,-P}$ to be Y_{-P}'s *i*th row. Thus, $Y_{i,-P}$ is defined as a row vector of y_{it}'s lag P values during the

sample period $t = 1, \ldots, T$. Suppose also that $V = (Y'_{-1}, \cdots, Y'_{-k})$ and $V_i = (Y'_{i,-1}, \cdots, Y'_{i,-k})$ for $i = 1, \ldots,$

n. For the equation that generates y_{1t}, the independent variable matrix is given by V_I; this equation has

the null hypothesis restrictions of non-Granger causality. V provides the matrix of independent

variables for the equation generating y_{2t}; this equation is not restricted by the null hypothesis of non-

Granger causality and it includes the lag values of all variables in the VAR model. Now we are in the

position to define the $T \times 1$ *leverages* vectors for y_{1t} and y_{2t} as the following:

$$l_1 = \mathrm{diag}\left(V_1 \left(V_1' V_1 \right)^{-1} V_1' \right), \text{ and } l_2 = \mathrm{diag}\left(V \left(V' V \right)^{-1} V' \right).$$

These leverages are used to modify the residuals in order to take into account the effect of ARCH.

The modified residual for y_{it} is produced as

$$\hat{u}_{it}^{m} = \frac{\hat{u}_{it}}{\sqrt{1 - l_{it}}}.$$

where the *t*th element of l_i is given by l_{it} and the raw residual from the regression with y_{it} as the

dependent variable is given by $\hat{u}_{it}$.

---

[3] *Leverages* are suggested by Davison and Hinkley (1999) for univariate cases. Hacker and Hatemi-J (2006)
generalized *leverages* to multivariate cases.

The bootstrap simulation is repeated *M* times and each time a Wald statistic is produced based on the Toda and Yamamoto (1995) methodology using the previously determined optimal lag order for the original data and an assumed order of integration. The resulting set of bootstrapped Wald statistics is referred to as the bootstrapped Wald distribution. Subsequently determined is the "bootstrap critical value" for the α-level of significance (c_α^*), which is the lower boundary of the top α quantile of the bootstrapped Wald-statistic distribution. If the calculated Wald statistic is higher than the bootstrap critical value c_α^* then the null hypothesis of non-causality is rejected based on bootstrapping at the α level of significance.

5. Simulations for Considering Size

In this paper we present the results of two types of simulations—those for considering size properties of the various methodologies in testing Granger causality and those for considering the power properties of those same methodologies. This section discusses the simulation design for considering how well various methodologies perform in having actual size being close to nominal size, and presents the results from such simulations. The next section deals with the power properties of the various methodologies.

The two-dimensional VAR(2) model below is used to consider the size properties:

$$y_t = v + \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \varepsilon_t \qquad (4)$$

where $y_t = (y_{1t}, y_{2t})'$ with y_{it}, $i = 1, 2$ being a scalar, v, and Π_i are parameter matrixes, and ε_t is the error term vector, the generation of which is described later. In the simulated "true" model for considering size there is non-Granger causality in both directions between the variables, so the Π_1 and Π_2 matrixes are diagonal.

In order to be clear about what order of integration is present for any particular parameter combination, we use the following equation for each variable y_{it} in y_t as the fundamental equation in which the parameters are expressed:

$$(1 - a_i L)(1 - c_i L) y_{it} = v_i + \varepsilon_{it}, \qquad \text{for } i = 1, 2 \qquad (9)$$

or equivalently

$$y_{it} = v_i + (a_i + c_i) y_{it-1} - (a_i c_i) y_{it-2} + \varepsilon_{it}, \qquad \text{for } i = 1, 2 \qquad (5)$$

with L denoting the lag operator. How many times either $|a_i|$ or $|c_i|$ is equal to one indicates y_{it}'s order of integration.

Based on the fundamental equations above, the parameter matrixes in (4) are based on the parameters in (5) as shown below:

$$\Pi_1 = \begin{bmatrix} a_1 + c_1 & 0 \\ 0 & a_2 + c_2 \end{bmatrix} \text{ and } \Pi_2 = \begin{bmatrix} -a_1 c_1 & 0 \\ 0 & -a_2 c_2 \end{bmatrix}.$$

In our simulations the error terms are drawn from a normal distribution, but two types of error processes are considered—those which are homoscedastic and those which follow an ARCH(1) process. More specifically, the variance of each error term is generated by the following data generating process:

$$Var[\varepsilon_{it} \mid \varepsilon_{it-1}] = (1 - \gamma_i) + \gamma_i \varepsilon_{it-1}^2, \qquad (6)$$

in which γ_i takes on two possible values in our simulations: zero for the homoscedastic cases and 0.5 for the ARCH cases. Due to this formulation, in both cases the unconditional variance is 1, so the homoscedastic and ARCH simulations are comparable.

In our simulations we consider various parameter combinations for a_1, a_2, c_1, and c_2 (the specific ones are discussed later), with their values varying between -1 and 1 to capture various situations of orders of integration. The parameter set for evaluating the size properties consists of the following:

$$a_1 = \{-1.0 \quad -0.2 \qquad 0.4 \quad 1.0\},$$
$$a_2 = \{-1.0 \quad -0.6 \qquad 0.8 \quad 1.0\},$$
$$c_1 = \{-1.0 \quad -0.2 \quad 0.0 \quad 0.4 \quad 1.0\},$$
$$c_2 = \{-1.0 \quad -0.2 \quad 0.0 \quad 0.4 \quad 1.0\}.$$

One thousand simulations are performed with a sample size of 40 for each of these parameter combinations, once with homoscedastic errors and once with ARCH.[4] There are 800 bootstrap simulations associated with each of the 1000 Monte Carlo simulations. The test's characteristics in terms of size accuracy are assessed for the conventional levels of significance (1%, 5%, and 10%).

Tables 1 and 2 present the simulation results for considering size properties. The columns of the tables are organized based first on how many augmentation lags are used and second on the order of integration of the two variables in the VAR model, with the following various combinations based upon the data-generating parameters (all of which do not show up in every table): "both I(2)" (both integrated of order 2), "both I(1)", and "both I(0)". It is important when reading these columns to consider if the number of augmentation lags is greater than, equal to, or less than the order of integration of the variables (due to the Toda and Yamamoto (1995) recommendations). Tables 1 and 2 both present homoscedasticity and ARCH situations and differ primarily in the order of VAR model for the data generating process: VAR(1) for Table 1 and VAR(2) for Table 2.

---

[4] With each of the 1000 Monte Carlo simulations there is 100 presample observations generated to remove the influence of start-up values. For the homoscedastic and ARCH simulations, the random number generator draws are the same from a standard normal distribution, but the ARCH simulations multiply the resulting random draws by a time varying numbers based on equation (6). By having the initial draws the same, the homoscedastic and ARCH simulations are more comparable and are more efficiently produced.

At the bottom of each of these tables is given the number of parameter combinations involved in the calculations (in the row titled "# Cases") and the percent of times one, two, or three lags is chosen. As is the case in all the results shown in the cells of these tables, the first number is that for the homoscedasticity situation and the second number, shown in parentheses, is that for the ARCH situation. The other rows present five categories of results for each of three nominal sizes: 1%, 5%, and 10%. The first and second categories show the percent (frequency) of rejection of the null hypothesis, using the asymptotic χ^2 distribution and the bootstrap distribution respectively. The closer the simulated size is to the nominal size, the better the size performance of the test. The third and fourth categories show the simulated size's mean absolute deviation from the nominal size, using the asymptotic χ^2 distribution and the bootstrap distribution respectively. The last category is the percent of the cases (parameter combinations) in which bootstrapping performs better than the asymptotic χ^2 distribution in having the simulated actual size closer to nominal size.

We see from these tables that the most common problem is having actual size greater than the nominal size. The worst results in terms of deviation of the actual size from nominal size occurs when the augmentation lag is less than the order of integration, i.e. when both the variables are I(1) and there is no augmentation lag or when both the variables are I(2) and there is no augmentation lag or only one augmentation lag. This is true whether we are using the asymptotic χ^2 distribution or the bootstrap distribution. Such results correspond to what we expect (at least with the asymptotic χ^2 distribution) given the results from Toda and Yamamoto (1995).

We see also that the bootstrapping we have simulated tends to improve the size properties further. In comparing the numbers in parentheses and outside parentheses on Tables 1 and 2 and, it appears that ARCH tends to inflate the actual sizes, but that effect is smaller when using bootstrapping rather than the asymptotic χ^2 distribution. In these tables the simulated size's mean absolute deviation from the nominal size is lower when using bootstrapping rather than the asymptotic χ^2 distribution. Also in almost every situation – with respect to VAR order, ARCH existence or not, number of augmentation

lags, order of integration, and nominal significance level – using bootstrapping outperforms using the χ^2 distribution in the vast majority of parameter cases considered.

Table 1: The size properties of the test statistics, T= 40, VAR(1)

Nominal Significance Level ↓	No augmentation lag		One augmentation lag		Two augmentation lags	
	Both I(1)	Both I(0)	Both I(1)	Both I(0)	Both I(1)	Both I(0)
	% rejecting the null hypothesis using χ^2 distribution					
1%	5.1 (6.8)	1.9 (2.0)	1.6 (1.7)	1.5 (1.7)	1.9 (1.6)	1.4 (1.6)
5%	14.2 (16.3)	6.7 (6.9)	6.1 (6.4)	5.7 (6.3)	6.1 (5.9)	5.2 (6.3)
10%	23.1 (24.1)	12.2 (12.5)	12.3 (11.3)	10.9 (11.2)	11.7 (11.1)	10.9 (11.3)
	% rejecting the null hypothesis using bootstrap distribution					
1%	1.8 (2.6)	1.3 (1.3)	1.0 (1.0)	1.2 (1.2)	1.2 (1.0)	1.2 (1.0)
5%	7.6 (8.5)	5.8 (5.9)	5.2 (5.2)	4.7 (5.2)	5.1 (5.2)	4.3 (5.2)
10%	14.0 (15.6)	10.9 (11.7)	10.9 (10.9)	9.7 (10.4)	10.2 (9.8)	9.7 (10.4)
	Mean absolute deviation using χ^2 distribution					
1%	4.1 (5.7)	1.0 (1.0)	0.6 (0.7)	0.6 (0.7)	0.9 (0.6)	0.6 (0.6)
5%	9.2 (11.3)	2.0 (1.8)	1.1 (1.6)	0.7 (1.2)	1.1 (1.2)	0.5 (1.3)
10%	13.1 (14.1)	2.4 (2.5)	2.3 (1.6)	1.2 (1.2)	1.7 (1.3)	0.9 (1.2)
	Mean absolute deviation using bootstrap distribution					
1%	0.9 (1.7)	0.5 (0.3)	0.3 (0.2)	0.5 (0.2)	0.3 (0.3)	0.5 (0.3)
5%	2.6 (3.5)	1.2 (0.9)	0.7 (0.9)	0.5 (0.4)	0.5 (0.9)	0.7 (0.5)
10%	4.0 (5.6)	1.2 (1.6)	1.2 (1.3)	0.4 (0.5)	0.8 (0.9)	0.8 (0.4)
	% of cases in which bootstrapping performs better than χ^2					
1%	100 (100)	75 (75)	75 (100)	75 (100)	100 (75)	75 (75)
5%	100 (100)	75 (75)	75 (75)	50 (100)	75 (75)	25 (25)
10%	100 (100)	75 (100)	75 (50)	75 (100)	75 (50)	50 (100)
# Cases	4(4)	4(4)	4(4)	4(4)	4(4)	4(4)
% 1 lag chosen	98.8 (94.9)	98.8 (96.7)	98.8 (94.9)	98.8 (96.7)	98.8 (94.9)	98.8 (96.7)
% 2 lags chosen	1.1 (4.7)	1.2 (3.3)	1.1 (4.7)	1.2 (3.3)	1.1 (4.7)	1.2 (3.3)
% 3 lags chosen	0.2 (0.4)	0.1 (0.1)	0.2 (0.4)	0.1 (0.1)	0.2 (0.4)	0.1 (0.1)

Results shown outside parentheses are for the homoscedastic situations and those shown in parentheses are for ARCH situations.

Table 2: The size properties of the test statistics, T= 40, VAR(2)

Nominal Significance Level ↓	No augmentation lag			One augmentation lag			Two augmentation lags		
	Both I(2)	Both I(1)	Both I(0)	Both I(2)	Both I(1)	Both I(0)	Both I(2)	Both I(1)	Both I(0)
% rejecting the null hypothesis using χ^2 distribution									
1%	12.2 (13.7)	7.3 (7.6)	2.9 (2.2)	5.6 (6.1)	2.6 (2.3)	1.9 (1.8)	2.7 (3.0)	2.1 (2.2)	1.7 (1.8)
5%	27.0 (28.4)	16.8 (17.4)	8.9 (7.3)	15.0 (15.6)	8.2 (7.4)	6.7 (6.4)	8.6 (8.7)	7.2 (7.0)	6.5 (6.3)
10%	37.2 (38.5)	24.6 (25.3)	15.2 (12.5)	23.1 (23.5)	14.1 (12.9)	12.3 (11.6)	14.7 (14.5)	12.6 (12.2)	11.6 (11.3)
% rejecting the null hypothesis using bootstrap distribution									
1%	2.2 (3.1)	2.9 (3.2)	1.7 (1.6)	1.6 (1.9)	1.2 (1.3)	1.2 (1.3)	1.4 (1.6)	1.1 (1.3)	1.1 (1.2)
5%	9.2 (10.5)	9.1 (10.0)	6.7 (6.0)	7.4 (7.8)	5.6 (5.6)	5.3 (5.4)	5.9 (6.0)	5.3 (5.5)	5.2 (5.3)
10%	16.4 (18.1)	15.5 (16.8)	12.6 (11.1)	13.9 (14.1)	10.8 (10.8)	10.5 (10.5)	11.5 (11.5)	10.5 (10.5)	10.1 (10.2)
Mean absolute deviation using χ^2 distribution									
1%	11.2 (12.7)	6.3 (6.6)	1.9 (1.2)	4.6 (5.1)	1.6 (1.3)	0.5 (0.9)	1.7 (2.0)	1.1 (1.2)	0.8 (0.8)
5%	22.0 (23.5)	11.9 (12.6)	3.9 (2.3)	10.0 (10.6)	3.2 (2.4)	1.2 (1.5)	3.6 (3.8)	2.3 (2.0)	1.5 (1.4)
10%	27.3 (28.5)	14.8 (15.7)	5.4 (2.7)	13.1 (13.5)	4.2 (3.0)	1.3 (1.7)	4.7 (4.5)	2.7 (2.3)	1.7 (1.5)
Mean absolute deviation using bootstrap distribution									
1%	1.3 (2.1)	1.9 (2.2)	0.8 (0.6)	0.7 (0.9)	0.4 (0.4)	0.4 (0.4)	0.4 (0.6)	0.3 (0.4)	0.3 (0.3)
5%	4.2 (5.5)	4.2 (5.2)	1.8 (1.3)	2.5 (2.8)	1.0 (1.0)	0.8 (0.8)	1.2 (1.1)	0.8 (0.8)	0.8 (0.8)
10%	6.4 (8.1)	5.7 (7.4)	2.9 (1.6)	4.1 (4.3)	1.5 (1.4)	1.2 (1.1)	1.6 (1.6)	1.2 (1.0)	1.2 (0.9)
% of cases in which bootstrapping performs better than χ^2									
1%	100 (93.8)	95.8 (95.8)	87.5 (87.5)	93.8 (100)	85.4 (92.7)	84.4 (81.3)	100 (100)	87.5 (92.7)	75.0 (87.5)
5%	93.8 (100)	93.8 (87.5)	93.8 (81.3)	100 (100)	90.6 (81.3)	81.3 (84.4)	93.8 (100)	86.5 (90.6)	87.5 (90.6)
10%	93.8 (87.5)	92.7 (88.5)	93.8 (81.3)	100 (100)	86.5 (83.3)	78.1 (81.3)	100 (100)	84.4 (85.4)	81.3 (78.1)
# Cases	16 (16)	96 (96)	32 (32)	16 (16)	96 (96)	32 (32)	16 (16)	96 (96)	32 (32)
% 1 lag chosen	0.1 (0.02)	22.3 (55.2)	59.2 (91.4)	0.1 (0.02)	22.3 (55.2)	59.2 (91.4)	0.1 (0.02)	22.3 (55.2)	59.2 (91.4)
% 2 lags chosen	96.6 (92.6)	75.2 (42.1)	39.9 (8.2)	96.6 (92.6)	75.2 (42.1)	39.9 (8.2)	96.6 (92.6)	75.2 (42.1)	39.9 (8.2)
% 3 lags chosen	3.3 (7.2)	2.5 (2.7)	0.9 (0.4)	3.3 (7.2)	2.5 (2.7)	0.9 (0.4)	3.3 (7.2)	2.5 (2.7)	0.9 (0.4)

Results shown outside parentheses are for the homoscedastic situations and those shown in parentheses are for ARCH situations.

6. Simulations for Considering Power

To investigate some power properties of the discussed methodologies for testing Granger causality, we first focus on the following VAR(1) model with various values for $B_{1,12}$:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.5 & B_{1,12} \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

The parameter $B_{1,12}$ is chosen from the set {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1.0} and the error terms are drawn from a standard normal distribution (power is not investigated when ARCH effects are present). In these simulations both variables are I(0). For each parameter combination we perform 1000 Monte Carlo simulations, and for each of those 1000 simulations there is an associated 800 bootstrap simulations. From these simulations we calculate the frequency in which the null hypothesis of no Granger causality of y_2 on y_1 is rejected. When $B_{1,12} = 0$, then this null hypothesis is true, otherwise it is false.

The power table in Table 3 presents the results for the nominal size of 5% for the bootstrap. For chi-square the nominal size was adjusted so that its actual size matches the actual size of the corresponding bootstrap. This makes the powers for chi-square and the bootstrap comparable. As can be seen the powers for the two methodologies are very similar for chi-square and the bootstrap. In the situations considered here, the process is stationary, so no augmentation lags are needed according to the Toda and Yamamoto (1995) recommendations. When moving from no augmentation lags to one augmentation lag and then two augmentation lags, the power is decreasing, as any augmentation lags greater than zero are unnecessary.

Table 3: Power Table, VAR(1), I(0), No ARCH

$B_{1,12}$	0	0.1	0.2	0.3	0.4	0.5	0.8	1
Chi-Square no aug lag[a]	0.056	0.107	0.251	0.496	0.706	0.883	0.999	0.999
Bootstrap no aug lag	0.056	0.104	0.253	0.503	0.707	0.880	0.998	0.999
Chi-Square one aug lag	0.064	0.087	0.193	0.398	0.588	0.790	0.985	0.990
Bootstrap one aug lag	0.064	0.090	0.192	0.398	0.598	0.797	0.985	0.998
Chi-Square two aug lags	0.059	0.087	0.173	0.356	0.560	0.768	0.983	0.998
Bootstrap two aug lags	0.058	0.088	0.173	0.351	0.556	0.761	0.983	0.998

[a]aug ≡ augmentation

To investigate power properties when both of the variables are I(1), we focus on the following VAR(1) model with various values for $B_{1,12}$:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & B_{1,12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

The simulations are performed analogously to those which produced Table 3, and the results are shown in Table 4 and Figure 1. Here we see that when there is no augmentation lag (less than recommended), the bootstrap power is notably better than the power using the asymptotic distribution. When one augmentation lag or two augmentation lags are used the bootstrap power is slightly better than that with the asymptotic distribution.

Table 4: Power Table, VAR(1), I(1), No ARCH

$B_{1,12}$	0	0.1	0.2	0.3	0.4	0.5	0.8	1
Chi-Square no aug lag[a]	0.143	0.106	0.352	0.601	0.832	0.927	0.996	0.999
Bootstrap no aug lag	0.143	0.224	0.536	0.753	0.92	0.972	0.997	0.999
Chi-Square one aug lag	0.06	0.086	0.199	0.327	0.541	0.742	0.979	0.997
Bootstrap one aug lag	0.06	0.092	0.214	0.351	0.559	0.755	0.98	0.997
Chi-Square two aug lags	0.06	0.082	0.205	0.297	0.511	0.717	0.967	0.994
Bootstrap two aug lags	0.06	0.085	0.21	0.312	0.523	0.717	0.969	0.994

[a]aug ≡ augmentation
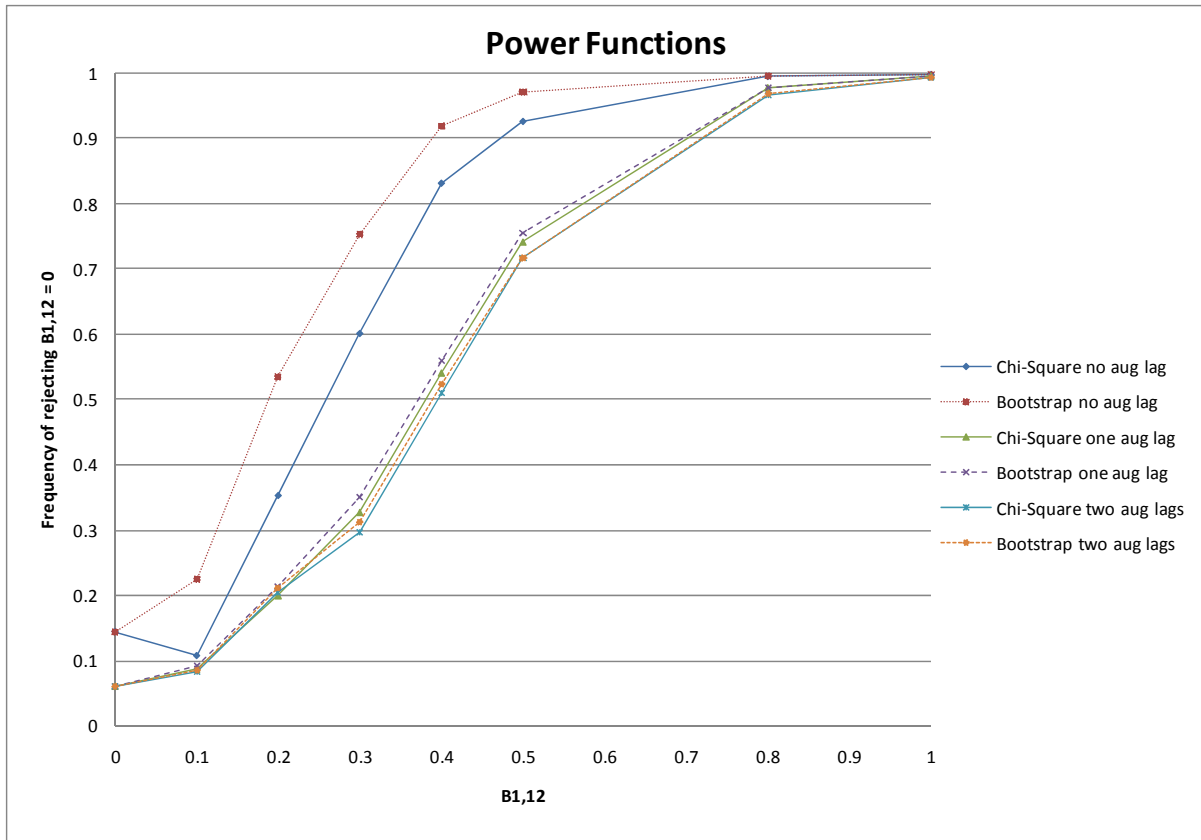
Power Functions

Figure 1. Power Functions for VAR(1), I(1), No ARCH

7.   An Application in Finance

The method suggested in this paper is applied to testing the causal effect of the US financial market on the UAE financial market using daily data during the period January 2, 2008 to April 8, 2009. We use the UAE Abu Dhabi Securities Market Index (AMF) and the Standard & Poors 500 Composite Index. Notably, the UAE dirham is fixed to the US dollar so there is no exchange rate risk with regard to the dollar in this particular case. The data are collected from the data-base *Ecowin*. The optimal lag order was selected to be two days based on minimizing the SBC with a maximum considered lag order of 40. An additional unrestricted lag was included in the VAR model in order to allow for one unit root. The residuals in the VAR model were tested for multivariate ARCH effects by using the test outlined in Hacker and Hatemi-J (2005). The results, not presented, showed that the null of no multivariate ARCH effects could be strongly rejected. Thus, applying the bootstrap-corrected test seems to be required in order to obtain more precise empirical results regarding the potential causal relationship between the underlying returns. The issue is important from a financial economic

perspective for two reasons. To find out whether the variables cause each other or not provides knowledge about: (i) The financial markets' degree of integration, and (ii) The existence or non-existence of international portfolio diversification benefits between the markets.

The results of the causality test are presented in Table 5. These results reveal that there is a strong causal impact of the US market on the UAE market. The implication of this empirical finding is that the UAE financial market is substantially integrated with the US market. Thus, there may not be a large amount of diversification benefits for the investors across the two markets.[5] It should be mentioned that we did not test the opposite hypothesis because we do not believe that the UAE market would cause the US market.

Table 5: The Results of Test for Causality Using the Bootstrap-Corrected test.

THE ESTIMATED TEST VALUE (*WALD*)	1% BOOTSTRAP CRITICAL VALUE	5% BOOTSTRAP CRITICAL VALUE	10% BOOTSTRAP CRITICAL VALUE
20.085	9.478	6.495	4.805

Notes: The null hypothesis is that the US financial market does not Granger cause the UAE financial market.

8. Conclusions

The objective of many empirical studies is to investigate the causal relationship between the underlying variables. Significance tests are used to test whether a variable's past values have explanatory power for another variable in a VAR model framework. This is basically testing for causality in Granger's (1969) sense. There are several test methods available in the literature for this purpose. See among others Granger (1969, 1988), Toda and Yamamoto (1995) and Hacker and Hatemi-J (2006). In the existing literature it is assumed that the lag order in the VAR model is known. However, in empirical studies the lag order is not known and it has to be selected first and then test for causality can be conducted using the selected lag order. In this paper we investigate situations where the lag order is determined endogenously when the test for causality is conducted.

---

[5] The sum of the estimated causal coefficients is positive.

We conduct a battery of simulations to evaluate the size and power properties of the Wald test for testing the null hypothesis of non-Granger-causality for different data generating processes in small sample sizes. Our simulations indicate that bootstrapping with endogenous lag length choice works well, often outperforming the Wald test using the asymptotic chi-square distribution and having an actual size closely matching the nominal size when the number of augmentation lags (as suggested by Toda and Yamamoto, 1995) is equal to or greater than the integration order. The bootstrapping technique also appears rather robust to ARCH effects. The power of the bootstrapping technique appears to be at least as good, if not better, than the power using the asymptotic distribution for I(0) and I(1) variables in the homoscedastic cases we examined.

An application for testing the causal impact of the US financial market on the UAE financial market is provided. It is found that the US market strongly Granger causes the UAE market. This implies that the UAE market has substantial integration with the US market, which in turn implies that international portfolio diversification benefits might not be large for the investors across these two markets. In order to check the generality of the empirical findings in this paper, there is a need for more studies in the future applying the suggested method.[6]

Acknowledgements

A version of this paper was presented at the UAE University. We thank the participants for their participation. The usual disclaimer applies.

---

[6] A consumer-friendly code written in GAUSS for applying the bootstrap test is available from the authors on request.

References

Davison, A. C., and Hinkley, D.V. (1999) *Bootstrap Methods and Their Application.* Cambridge University Press. Cambridge, UK.

Efron, B. (1979) Bootstrap Methods: Another Look at the Jackknife, *Annals of Statistics 7,* 1-26.

Engle, R. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 50(4), 987-1007

Granger, C. W. J, (1969) Investigating Causal Relations by Econometric Models and Cross Spectral Methods. *Econometrica*, 37, 424-438.

Granger, C. W. J. (1988) Some recent developments in the concept of causality, Journal of Econometrics, 39,199–211.

Hacker R. S. and Hatemi-J A. (2005) A test for multivariate ARCH effects. *Applied Economics Letters*, 12(7), 411-417.

Hacker R. S. and Hatemi-J A. (2006) Tests for causality between integrated variables using asymptotic and bootstrap distributions: theory and application. *Applied Economics*, 38(13), 1489-1500.

Hatemi-J A. (2004) Multivariate tests for autocorrelation in the stable and unstable VAR models. *Economic Modelling*, 21(4), 661-683.

Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis.* Springer-Verlag: Heidelberg, Germany.

Toda, H. Y. and Yamamoto, T. (1995) Statistical inference in vector autoregressions with possibly integrated processes, *Journal of Econometrics*, 66, 225–50.

Appendix

In order to express a Wald test statistic that can be used to test the null hypothesis defined by (2) in a compact way, we assume for each of the variables $y_{-k+1}, \ldots, y_0$ there are available k presample values and define the following matrixes:

$Y := (y_1, \cdots, y_T) \quad (n \times T)$ matrix,

$D := (B_0, B_1, \cdots, B_k) \quad (n \times (1 + nk))$ matrix,

$$Z_t := \begin{bmatrix} 1 \\ y_t \\ y_{t-1} \\ \vdots \\ y_{t-k+1} \end{bmatrix} \quad ((1 + nk) \times 1) \text{ matrix, for } t = 1, \ldots, T,$$

$Z := (Z_0, \cdots, Z_{T-1}) \quad ((1 + nk) \times T)$ matrix, and

$\hat{\delta} := (\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_T) \quad (n \times T)$ matrix.

By using these notations, we are in a position to define the estimated VAR(k) model, written compactly as:

$$Y = DZ + \delta. \tag{A1}$$

The next step is to estimate δ_U, the ($n \times T$) matrix of residuals from the *unrestricted* regression (A1) when the null hypothesis is not imposed. Then the variance-covariance of these residuals is calculated as $S_u = \dfrac{\delta_U \delta_U'}{T - (1 + nk)}$. Let us define $\beta = vec(B_0, B_1, \cdots, B_k)$ or $\beta = vec(D)$, where *vec* signifies the column-stacking operator, and consider $\hat{\beta}$ as the ordinary least squares estimate of β. The Wald (W) test statistic for testing the null hypothesis that one of y_t's variables (a specific variable) is not Granger causing another of y_t's varables (another specific variable) is then written as

$$W = \left(Q\hat{\beta}\right)' \left[Q\left((Z'Z)^{-1} \otimes S_U\right)Q'\right]^{-1} \left(Q\hat{\beta}\right), \tag{A2}$$

where $\otimes$ is the Kronecker product that represents elements by all elements multiplication of matrixes, and Q is a $k \times n(1+nk)$ matrix. Each of Q's k rows is associated with a zero restriction on one of β's parameters. Each element in each row of Q is given the value of one

if the associated parameter within the vector β is zero given the null hypothesis of non-causality is true, and it is given a zero value otherwise. By utilizing these compact notations, the null hypothesis of non-Granger causality can also be defined as

$$H_0 : Q\beta = 0 .$$ (A3)

Assuming the data is generated according to equation (1) with the null hypothesis being true; the Wald test statistic presented in equation (A2) is distributed asymptotically as χ^2 with the degrees of freedom equivalent to the number of restrictions under the null, which is equal to the lag order k in this particular case.