

Tests for causality between integrated variables  
using asymptotic and bootstrap distributions

R Scott Hacker and Abdalnasser Hatemi-J

October 2003

2003:2



LUND  
UNIVERSITY

DEPARTMENT OF  
STATISTICS

S-220 07 LUND  
SWEDEN

# Tests for Causality Between Integrated Variables Using Asymptotic and Bootstrap Distributions

R. Scott Hacker

Jönköping International Business School  
P.O. Box 1026, SE-551 11, Jönköping, Sweden

E-mail: [Scott.Hacker@ihh.hj.se](mailto:Scott.Hacker@ihh.hj.se)

Telephone: +46 36 15 75 14, Fax + 46 36 12 18 32  
and

Abdulnasser Hatemi-J

*Lund University and*

*University of Skövde*

*P.O. Box 408, SE-541 28, Skövde, Sweden.*

E-mail: [abdulnasser.hatemi-j@ish.his.se](mailto:abdulnasser.hatemi-j@ish.his.se)

Tel: +46 500 44 87 31, Fax: +46 500 44 87 99

## Abstract

Causality tests in the Granger's sense are increasingly applied in empirical research. Since the unit root revolution in time-series analysis, several modifications of tests for causality have been introduced in the literature. One of the recent developments is the Toda-Yamamoto modified Wald (MWALD) test, which is attractive due to its simple application, its absence of pre-testing distortions, and its basis on a standard asymptotical distribution irrespective of number of unit roots and the cointegrating properties of the data. This study investigates the size properties of the MWALD test and finds that in small sample sizes this test has poor size properties when using its asymptotical distribution, the chi-square. We suggest making use of a bootstrap distribution to lower the size distortions. Our Monte Carlo simulation results show that an MWALD test based on a bootstrap distribution has much smaller size distortions compared to corresponding cases when the asymptotic distribution is used. These results hold for different sample sizes, integration orders, and error term processes (homoscedastic or ARCH).

**Key words:** Granger Causality, Toda-Yamamoto test, VAR model, Stability, ARCH, Leverages

**JEL classification:** C32, C15

**Running title:** Tests for Causality Between Integrated Variables

# **Tests for Causality Between Integrated Variables Using Asymptotic and Bootstrap Distributions**

## **Abstract**

Causality tests in the Granger's sense are increasingly applied in empirical research. Since the unit root revolution in time-series analysis, several modifications of tests for causality have been introduced in the literature. One of the recent developments is the Toda-Yamamoto modified Wald (MWALD) test, which is attractive due to its simple application, its absence of pre-testing distortions, and its basis on a standard asymptotical distribution irrespective of number of unit roots and the cointegrating properties of the data. This study investigates the size properties of the MWALD test and finds that in small sample sizes this test has poor size properties when using its asymptotical distribution, the chi-square. We suggest making use of a bootstrap distribution to lower the size distortions. Our Monte Carlo simulation results show that an MWALD test based on a bootstrap distribution has much smaller size distortions compared to corresponding cases when the asymptotic distribution is used. These results hold for different sample sizes, integration orders, and error term processes (homoscedastic or ARCH).

**Key words:** Granger Causality, Toda-Yamamoto test, VAR model, Stability, ARCH, Leverages

**JEL classification:** C32, C15

**Running title:** Tests for Causality Between Integrated Variables

## 1. Introduction

One of the most important and fundamental issues in science is the causal relationship between variables of interest. What is meant by causality and how it can be measured is an issue that is a focus of debate. In economics, Granger causality (Granger, 1969) is a well-known concept and one of the most applied methods in applied research. The definition of causality in the Granger sense is based on the idea that the past cannot be caused by the present or future. Hence, if an event occurs before another event, causality can only occur from the first event to the second one. Granger formulated a test statistic to test whether movements in one variable systematically precede movements in another variable. Granger causality is determined by discovering whether including the past values of a variable in the information set can improve the forecast of another variable. In a regression context this means running a regression of one variable on the past values of itself and the past values of any potentially casual variable, and testing the significance of coefficient estimates associated with the potentially causal variable.

The Granger test was originally based on asymptotic distribution theory. However, Granger and Newbold (1974) showed through Monte Carlo simulations that if the data generating process (DGP) of the variables is characterised by integration (non-stationarity), the regression analysis based on the asymptotic distribution theory does not work well and the estimated results can be spurious. Phillips (1986) provided an analytical basis for this assertion. To remedy this problem one can transform the data by differencing, but this will result in loss of long run information. Since the development of cointegration by Granger (1981), Engle and Granger (1987), and Johansen (1988) (among others), the vector error correction model (VECM) has become a useful tool for empirical research of integrated variables. Granger causality tests can also be performed by using the VECM (Granger, 1988). This procedure, however, requires pretesting for unit roots and cointegration, which may not be of interest if the primary objective is to determine whether or not there is causality.

Sims et. al. (1990) showed that the asymptotic distribution theory cannot be applied for testing causality of integrated variables in level form using the vector autoregressive (VAR) model even if the variables cointegrate. Based on augmented VAR modelling, Toda and Yamamoto (1995) introduced a Wald test statistic that asymptotically has a chi-

square distribution irrespective of the order of integration or cointegration properties of the variables in the model.<sup>1</sup> The test is very simple to apply and it is one of the most applied tests for causality between integrated variables in applied research.

The objective of this study is to investigate for finite samples the size properties of the Toda-Yamamoto test statistic for Granger causality when the DGP is characterized for both variables by integration of the same order (degree zero, one or two) or when it is characterized by different integration orders for the two variables. Using Monte Carlo simulations, we look at the size properties of this test statistic using the asymptotic chi-square distribution and investigate whether using a bootstrapped distribution could be a better alternative. We also investigate whether the Toda-Yamamoto test statistic is sensitive for the presence of autoregressive conditional heteroscedasticity (ARCH). All of the simulations are conducted by using GAUSS.

This paper continues as follows. Section 2 defines the Toda-Yamamoto test statistic. The simulation design and the bootstrap simulation technique is described in Section 3. The simulation results are presented in section 4 and the conclusion is presented in the last section. Associated tables are provided at the end of the paper.

## 2. The Toda-Yamamoto Test Statistic

The Toda-Yamamoto procedure is based on a lag(s) augmentation of the VAR model. Consider the following VAR( $p$ ) process:

$$y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t, \quad (1)^2$$

where  $y_t$ ,  $v$ , and  $\varepsilon_t$  are  $n$ -dimensional vectors and  $A_r$  is an  $n \times n$  matrix of parameters for lag  $r$ . The error vector,  $\varepsilon_t$ , is a zero-mean independent, identically distributed process with a non-singular covariance matrix  $\Sigma_\varepsilon$  that meets the condition  $E|\varepsilon_{it}|^{2+\lambda} < \infty$  for some positive  $\lambda$ , where  $\varepsilon_{it}$  is the  $i$ th element of  $\varepsilon_t$ . Toda and Yamamoto (1995) suggest the following

---

<sup>1</sup> Dolado and Lütkepohl (1996) provided independently a similar procedure as the Toda-Yamamoto Wald test. However, Dolado and Lütkepohl dealt with only variables that are integrated of degree one.

<sup>2</sup> For a new approach to choose optimal lag order ( $p$ ) in the VAR model see Hatemi-J (2003).

augmented VAR( $p+d$ ) model that can be used for tests of causality between integrated variables:

$$y_t = \hat{\nu} + \hat{A}_1 y_{t-1} + \dots + \hat{A}_p y_{t-p} + \dots + \hat{A}_{p+d} y_{t-p-d} + \hat{\varepsilon}_t, \quad (2)$$

where the circumflex above a variable represents its OLS estimate, a notation that continues throughout this paper. The order  $p$  of the process is assumed to be known and  $d$  is equal to the maximum order of integration of the variables. The  $k$ th element of  $y_t$  does not Granger-cause the  $j$ th element of  $y_t$  if the following hypothesis is not rejected:

$$H_0: \text{the row } j, \text{ column } k \text{ element in } A_r \text{ equals zero for } r = 1, \dots, p. \quad (3)$$

It should be noted that the parameters for the extra lag(s), i.e.  $d$ , are unrestricted in testing for Granger causality. Their function is to guarantee the use of asymptotical distribution theory, according to Toda and Yamamoto (1995). Before defining the test statistic introduced by Toda-Yamamoto for testing the hypotheses of interest, let us define the following denotations for a sample size  $T$ :

$$\begin{aligned} Y &:= (y_1, \dots, y_T) \quad (n \times T) \text{ matrix,} \\ \hat{D} &:= (\hat{\nu}, \hat{A}_1, \dots, \hat{A}_p, \dots, \hat{A}_{p+d}) \quad (n \times (1 + n(p+d))) \text{ matrix,} \\ Z_t &:= \begin{bmatrix} 1 \\ y_t \\ y_{t-1} \\ \vdots \\ y_{t-p-d+1} \end{bmatrix} \quad ((1 + n(p+d)) \times 1) \text{ matrix, for } t = 1, \dots, T, \\ Z &:= (Z_0, \dots, Z_{T-1}) \quad ((1 + n(p+d)) \times T) \text{ matrix, and} \\ \hat{\delta} &:= (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_T) \quad (n \times T) \text{ matrix.} \end{aligned}$$

By using this notation, the estimated VAR( $p+d$ ) model including an estimated constant term ( $\hat{\nu}$ ) can be written compactly as:

$$Y = \hat{D}Z + \hat{\delta}. \quad (6)$$

We proceed by estimating  $\hat{\delta}_U$ , the  $(n \times T)$  matrix of estimated residuals from the *unrestricted* regression (6). We then calculate the matrix of cross-products of these residuals as  $S_U = \hat{\delta}_U' \hat{\delta}_U$ . Let us define  $\beta = \text{vec}(v, A_1, \dots, A_p, 0_{n \times nd})$  and  $\hat{\beta} = \text{vec}(\hat{D})$ , where  $\text{vec}$  denotes the column-stacking operator and  $0_{n \times nd}$  denotes a zero matrix with  $n$  rows and  $n(d)$  columns. The modified Wald (MWALD) test statistic for testing non-Granger causality of one variable in  $y_t$  on another variable in  $y_t$ , suggested by Toda and Yamamoto (1995), can be then written as

$$MWALD = (C\hat{\beta})' [C((Z'Z)^{-1} \otimes S_U)C']^{-1} (C\hat{\beta}), \quad (7)$$

where  $\otimes$  is the Kronecker product, and  $C$  is a  $p \times n(1+n(p+d))$  matrix. Each of the  $p$  rows of  $C$  is associated with the restriction to zero of one parameter in  $\beta$ . The elements in each row of  $C$  take on the value of one if the associated parameter in  $\beta$  is zero under the null hypothesis, and they take on the value of zero if there is no such restriction under the null. None of the rows in  $C$  are associated with restrictions on the last  $n^2(d)$  elements in  $\beta$ , which correspond the  $0_{n \times nd}$  matrix noted above. Using this compact notation, the null hypothesis of non-Granger causality would be given by

$$H_0 : C\beta = 0.$$

The MWALD test statistic is asymptotically  $\chi^2$  distributed with the number of degrees of freedom equal to  $p$ , the number of restrictions to be tested.

### 3. Monte Carlo Simulation Design and Bootstrapping

In order to check the size properties of MWALD test statistic we make use of the following two-dimensional order-two VAR model:

$$y_t = v + Ay_{t-1} + By_{t-2} + \varepsilon_t, \quad (8)$$

where  $y_t = (y_{1t}, y_{2t})'$  and  $y_{it}$ ,  $i = 1, 2$  is a scalar variable. We constrain  $A$  and  $B$  to be diagonal matrixes so that the individual variables depend only on their past values and not on the past values of the other variable. Therefore, in the “true” model we have non-Granger causality for each variable with respect to the other. For each variable  $y_{it}$  in  $y_t$  we look at the following formulation:

$$(1 - a_{ii}L)(1 - b_{ii}L)y_{it} = v_i + \varepsilon_{it}, \quad \text{for } i = 1, 2 \quad (9)$$

or

$$y_{it} = v_i + (a_{ii} + b_{ii})y_{it-1} - (a_{ii}b_{ii})y_{it-2} + \varepsilon_{it}, \quad \text{for } i = 1, 2 \quad (10)$$

where  $L$  signifies the lag operator.

Therefore, the parameter matrixes are defined as

$$A = \begin{bmatrix} a_{11} + b_{11} & 0 \\ 0 & a_{22} + b_{22} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -a_{11}b_{11} & 0 \\ 0 & -a_{22}b_{22} \end{bmatrix}.$$

The cases that we investigate are divided into two groups. The first group comprises a VAR(1) model when both variables are I(1) (integrated of degree one), both are I(0), or there are mixed orders of integration. The second group consists of VAR(2) when both variables are I(2), both are I(1), both are I(0), or there are mixed orders of integration. If  $b_{ii} = 0$ , the order of integration for  $y_{it}$  is one if  $|a_{ii}| = 1$ , and it is zero if  $|a_{ii}| < 1$ . If  $b_{ii} \neq 0$  then the order of integration for  $y_{it}$  is equal to number of times that either  $|a_{ii}|$  or  $|b_{ii}|$  is equal to one.

For our simulations we draw each error term independently from a standard normal distribution. However, the data generating process for many economic time series can be described by periods of high volatility. To investigate the sensitivity of tests for causality in such cases, we also conduct the simulations when the variance of the error terms can be described by the following simple autoregressive conditional heteroscedasticity (ARCH) process, as introduced by Engle (1982):

$$\varepsilon_{it} = w_{it} \times [(\alpha_i + \gamma_i \varepsilon_{it-1}^2)]^{\frac{1}{2}}, \quad i = 1, 2 \quad (11)$$



where  $\alpha_i$  and  $\gamma_i$  are constants and  $w_{it}$  signifies a standard normal variable. Under such a situation, the conditional mean and conditional variance of  $\varepsilon_{it}$  are respectively

$$E[\varepsilon_{it} | \varepsilon_{it-1}] = 0, \quad (12)$$

$$Var[\varepsilon_{it} | \varepsilon_{it-1}] = E[\varepsilon_{it}^2 | \varepsilon_{it-1}] = E[w_{it}^2] \times [\alpha_i + \gamma_i \varepsilon_{it-1}^2] = \alpha_i + \gamma_i \varepsilon_{it-1}^2. \quad (13)$$

Consequently,  $\varepsilon_{it}$  is heteroscedastic conditional on  $\varepsilon_{it-1}$ . The unconditional variance of  $\varepsilon_{it}$  is

$$Var[\varepsilon_{it}] = E[Var[\varepsilon_{it} | \varepsilon_{it-1}]] = \alpha_i + \gamma_i E[\varepsilon_{it-1}^2] = \alpha_i + \gamma_i Var[\varepsilon_{it-1}], \quad (14)$$

but if the error terms are generated through a process that is variance stationarity (which holds given our DGP for the error terms), then we have a constant unconditional variance:

$$Var[\varepsilon_{it}] = Var[\varepsilon_{it-1}] = \frac{\alpha_i}{1 - \gamma_i}. \quad (15)$$

We set

$$\alpha_i = (1 - \gamma_i). \quad (16)$$

By doing this we make sure that the unconditional variance of  $\varepsilon_{it}$  is also one in the ARCH case, the same as the error variance as in the homoscedastic case, making comparison more transparent.

Combining the last equation with (13) leads to

$$Var[\varepsilon_{it} | \varepsilon_{it-1}] = (1 - \gamma_i) + \gamma_i \varepsilon_{it-1}^2, \quad (17)$$

a formula which is used directly in our simulations for including ARCH effects.

We generate 100 presample observations in order to cancel the effect of start-up values. Some of these presample observations are used as the lagged value for the first  $p+d$  observations so we have the same number of observations in estimating the VAR model irrespective of the number of lags. For comparability and parsimonious use of the random number generator from a  $N(0,1)$  distribution, the random number draw is initially the same for homoscedastic and ARCH versions with the ARCH version subsequently multiplying draws by an appropriate time varying number to modify the variable .

Each time we generate a set of simulated observations, we calculate the MWALD test statistic to test the hypothesis that  $y_{2t}$  does not Granger cause  $y_{1t}$ . We determine whether the hypothesis would be rejected at the  $\alpha$ -level ( $\alpha = 1\%$ ,  $5\%$ , or  $10\%$ ) based upon (i) the asymptotic distribution noted by Toda and Yamamoto and (ii) a bootstrapped distribution. We consider the bootstrapped distribution because using the asymptotic distribution may lead to the test not having the correct size in finite samples, and inference based on it could be misleading. The bootstrap method, developed by Efron (1979), is based on resampling the data to estimate the distribution of a test statistic. Using this distribution may reduce bias in inference by providing more reliable critical values.

For the bootstrap simulations we first estimate regression (1) through OLS with the restriction for the null hypothesis of no Granger causality on one of the included equations. For each bootstrap simulation we generate the simulated data,  $y_t^*$ ,  $t = 1, \dots, T$ , based on the coefficient estimates from this regression,  $\tilde{A}_1, \dots, \tilde{A}_p$ ; the original  $y_{t-1}, \dots, y_{t-p}$  data; and  $\tilde{\varepsilon}_t^*$ , the bootstrapped residuals. These residuals are based on  $T$  random draws with replacement from the regression's modified residuals (to be defined shortly), each with probability mass  $1/T$ . The mean of the resulting set of drawn modified residuals is subtracted from each of the modified residuals in that set, resulting in the set of bootstrapped residuals used to generate  $y_t^*$ . The subtraction of the mean is done to guarantee the mean of the bootstrapped residuals is zero.

The modified residuals are the regression's raw residuals modified to have constant variance, through the use of *leverages*, as noted in Davison and Hinkley (1999). This seems to be important in our case because we allow the model to have conditional heteroscedasticity in some of our simulations. To be explicit about our modification, we

need to introduce some more notations. Let  $Y_{-L} = (y_{1-L}, \dots, y_{T-L})$  and let  $Y_{i,-L}$ ,  $i = 1$  or  $2$ , be the  $i$ th row of  $Y_{-L}$ , i.e. it is a row vector of the lag  $L$  values for variable  $y_{it}$ ,  $t = 1, \dots, T$ . Let also  $X = (Y_{-1}', \dots, Y_{-p}')$  and  $X_i = (Y_{i,-1}', \dots, Y_{i,-p}')$ . For the equation that determines  $y_{1t}$ , the explanatory variable matrix for the regression is  $X_1$ ; this equation is restricted to have no Granger causality. For the equation that determines  $y_{2t}$ , the explanatory variable matrix for the regression is  $X$ ; this equation allows all lags of both variables to be included. The  $T \times 1$  leverage vectors for  $y_{1t}$ , and  $y_{2t}$ ,  $t = 1, \dots, T$  are respectively defined as

$$h_1 = \text{diag} \left( X_1 (X_1' X_1)^{-1} X_1' \right), \text{ and}$$

$$h_2 = \text{diag} \left( X (X' X)^{-1} X' \right).$$

Let  $h_{it}$  be the  $t$ th element of  $h_i$ , and let  $\tilde{\mathcal{E}}_{it}$  be the raw residual from the regression for  $y_{it}$ ,  $i = 1, 2$ . The modified residual for  $y_{it}$  is defined as

$$\tilde{\mathcal{E}}_{it}^m = \frac{\tilde{\mathcal{E}}_{it}}{\sqrt{1 - h_{it}}}.$$

By performing the bootstrap simulation 800 times and subsequently producing the MWALD test statistic each time, we can produce an approximate distribution for the MWALD test statistic. After these 800 estimations we find the  $(\alpha)$ th upper quantile of the distribution of bootstrapped MWALD statistics and obtain the  $\alpha$ -level ‘‘bootstrap critical values’’ ( $c_\alpha^*$ ). With the MWALD statistic calculated from the original simulated data (not the bootstrapped simulated data), we reject the null hypothesis based on bootstrapping if the actual MWALD is greater than  $c_\alpha^*$ .

#### 4. Case Specifications and Results

For our simulations the cases that are considered comprise of all variables being I(1), I(2), or a combination of I(0), I(1), and I(2). Two different sample sizes are considered ( $T = 40$ ,

and  $T = 100$ ) and two different error-generating processes are considered (homoscedastic and ARCH). We will refer to these four combinations of sample sizes and error generating processes as *scenarios*. The size properties of the test are investigated for the significance levels 1%, 5% and 10% respectively. Many combinations of the parameters in the underlying model are used to make the results more representative. We choose parameters for our simulations from combinations arising from the following sets:

$$\begin{aligned} a_{11} &= \{-1.0 \quad -0.2 \quad 0.4 \quad 1.0\}, \\ a_{22} &= \{-1.0 \quad -0.6 \quad 0.8 \quad 1.0\}, \\ b_{11} &= \{-1.0 \quad -0.2 \quad 0.0 \quad 0.4 \quad 1.0\}, \\ b_{22} &= \{-1.0 \quad -0.2 \quad 0.0 \quad 0.4 \quad 1.0\}, \end{aligned}$$

where  $a_{ii}$  and  $b_{ii}$ ,  $i = 1, 2$ , are the parameters given in equation (10). The chosen value for  $\gamma_i$  (from equation 17) for each equation is 0 in our homoscedastic cases and 0.5 in our ARCH cases. For each parameter combination we perform 1000 simulations for a given scenario. Each of those 1000 simulations has an associated 800 bootstrap simulations.

The simulation results are presented in Tables 1-4. These tables are organized in the columns according to eleven categories based on three characteristics: (i) the order of the VAR model indicated by the population parameters, (ii) the degree of integration indicated by the population parameters, and (iii) the number of extra lags used. For the VAR(1) model there are 16 parameter combinations, 4 of which have both variables as I(1), another 4 of which have both variables as I(0), and 8 of which have one variable as I(0) and the other as I(1). For these three situations we look at the size properties when one extra lag is used to calculate the MWALD statistic. For the VAR(2) model there are 384 parameter combinations, 16 of which have both variables as I(2), 96 of which have both variables as I(1), 32 of which have both variables as I(0), and 240 of which have integration orders I(0), I(1), or I(2) which are different (mixed) for the two variables. For these four situations we look at the size properties when one extra lag is used and when two extra lags are used, leaving us with eight more categories based on VAR order, integration order, and extra lags, in addition to the three situations noted with VAR(1), so we are considering eleven categories overall.

Three of the categories have the number of extra lags matching exactly the degree of integration for both variables (VAR(1), both I(1), one extra lag; VAR(2), both I(1), one extra lag; and VAR(2), both I(2), two extra lags). Four of the categories have the number of extra lags greater than the degree of integration for both variables (VAR(1), both I(0), one extra lag; VAR(2), both I(0), one extra lag; VAR(2), both I(1), two extra lags; and VAR(2), both I(0), two extra lags). One category has the number of extra lags less than the degree of integration (VAR(2), both I(2), one extra lag). The other three categories have mixed integration orders.

For each of these eleven categories we consider size properties for three different nominal sizes, 1%, 5%, and 10%. We will refer to each category and nominal size combination as a *case*. With three nominal sizes and eleven categories, there are 33 cases considered on each table. Each table deals with a different scenario on sample size (40 or 100) and error variance (homoscedasticity or ARCH with  $\gamma_i = 0.5$ ).

In these tables we have three ways of looking at the size performance of the MWALD test using its asymptotic distribution ( $\chi^2$ ) and using a bootstrapped distribution. First we look at the percent rejecting the null hypothesis of no Granger causality over all the simulations for all the parameter combinations in the category (1000  $\times$  the number of parameter combinations in the category). Second, we look at the mean absolute deviation of the empirical (simulation-based) size from the nominal size over all the simulations for all the parameter combinations in the category. Third, we look at the percent of parameter combinations within the category for which the performance using one distribution (bootstrap or  $\chi^2$ ) is better than that of using the other. By “better” we mean a lower mean absolute deviation of the empirical size from the nominal size over the 1000 simulations for a particular parameter combination. We provide shading on the tables to indicate which distribution—asymptotic or bootstrapped—is performing relatively best in each case for each of the three ways of considering size performance.

Based on these results we can see that the test statistic tends to over-reject the null hypothesis in almost all cases in all scenarios using the asymptotic distribution. However, when the bootstrap distribution is used, the performance of the test improves in almost all cases in all scenarios when looking at the percent rejecting the null hypothesis. In many

cases the bootstrapped version of the test is very close to the correct size. In some cases there is some notable size distortion when the bootstrap distribution is used, but the size distortion of a bootstrap  $p$ -value for the causality test is much smaller than that of a corresponding asymptotic  $p$ -value using the chi-square distribution. When considering the mean absolute deviation, the bootstrap distribution indicates better (lower) values in almost all cases in all scenarios. When considering the percent of parameter combinations in which either distribution is better, bootstrapping again is indicated to be superior to using the chi-square distribution, with a few exceptions in the VAR(1) categories.

For example, let us consider the category in Table 1 (sample size of 40 without ARCH) of two variables that are I(1) following a VAR(2) process with one extra lag and focus on the success of using the bootstrap distribution rather than the chi-square distribution at the 5% significance level. The simulated actual size of the MWALD test is 7.8% if the  $\chi^2$  distribution is used. However, the simulated actual size is 5.5% if the bootstrap distribution is used. This implies that using the bootstrap distribution improves the size properties by 2.3 percentage points in this particular case. The mean absolute deviation of the simulated actual size from the nominal size in this case is 2.9 percentage points using the  $\chi^2$  distribution and it is 0.8 percentage points using the bootstrap distribution. Also for this case, in 89.6% of the parameter combinations there is a lower mean absolute deviation using bootstrapping and in 10.4% of the parameter combinations there is a lower mean absolute deviation using the chi-square distribution.

Another interesting piece of information that we obtain from the simulations is that having more extra lags than the integration order results in less size distortion than having less extra lags than the integration order. As an example let us consider two categories in Table 1, VAR(2) with both variables being I(2) with one extra lag, and VAR(2) with both variables being I(1) with two extra lags. In the first case, where the integration order is higher than the number of extra lags, at the 5% nominal significance level the actual size using the  $\chi^2$  distribution is 15.4% and it is 7.2% using the bootstrap distribution. In the second case, however, where the integration order is lower than the number of extra lags, at the 5% nominal significance level the actual size using the  $\chi^2$  distribution is 7.4% and it is 5.4% using the bootstrap distribution.

Unsurprisingly, a comparison of Table 2 with Table 1, and Table 4 with Table 3, indicates that the MWALD test using the asymptotic distribution has better size properties for larger sample sizes in the vast majority of cases. This improving effect of larger sample size is not generally apparent when using the bootstrap distribution. Notable also is that the size properties worsen with the larger sample sizes in the category of VAR(2)/Both I(2)/one extra lag—the only category with the number of extra lags less than the degree of integration—regardless of whether we are using the asymptotic distribution or the bootstrapped one and regardless of which nominal size we considering.

Another result obtained from the simulations (comparing Tables 3 and 4 with Tables 1 and 2 respectively on the percent rejecting the null hypothesis) is that the existence of ARCH effects renders greater size distortions in all cases when the  $\chi^2$  distribution is used with a sample size of 40, and in most cases when the  $\chi^2$  distribution is used with a sample size of 100. Likewise, with either sample size, the existence of ARCH effects results in greater size distortions when the bootstrap distribution is used in most cases, but not as commonly as when the  $\chi^2$  distribution is used.

## 5. Conclusion

When the primary objective of an empirical investigation is to explore causality in the Granger sense between integrated variables, the Toda-Yamamoto modified Wald test might be an appealing alternative. This procedure, in which an empirical VAR model is estimated with additional lag(s) (the number of which is equal to the integration order of the variables), ensures that the  $\chi^2$  distribution with the degrees of freedom equal to the number of restrictions for non-Granger causality is the appropriate asymptotic distribution for testing. However, our simulations show that the asymptotic distribution can be a poor approximation, especially for the small samples that are common in empirical studies. To improve on the size properties of the modified Wald test we look at the properties of this test based on empirical distributions generated through bootstrap resampling. The results show that the bootstrapped empirical size for the modified Wald test is close to the correct size in the different cases investigated in this study when the extra lags are greater than or equal to the integration order of both variables, and it is generally closer to the correct size than the asymptotic distribution empirical size.

Our results also indicate that including more extra lags than the integration order of the variables does not distort the size notably, but having less extra lags than the integration order of the variables results in serious size distortion. This serious size distortion persists (although lessened) when the bootstrap distribution is used rather than the asymptotic one, and can worsen with larger samples.

Finally, we have found that having an ARCH error process rather than a homoscedastic one usually results in greater size distortions, especially when using the asymptotic distribution. Nevertheless, increased size distortions due to ARCH still do not appear significant when using the bootstrap distribution.



## References

- Davison, A. C., and Hinkley, D.V. (1999) *Bootstrap Methods and Their Application*. Cambridge University Press. Cambridge, UK.
- Dolado, J. and Lütkepohl, H. (1996) Making Wald Tests Work for Cointegrated VAR Systems. *Econometric Reviews*, 15, 369-86.
- 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
- Engle, R. F., and Granger, C. W. J (1987) Cointegration and Error Correction: Representation, Estimation and Testing, *Econometrica*, 55(2), 251-581.
- Granger, C. W. J, (1969) Investigating Causal Relations by Econometric Models and Cross Spectral Methods. *Econometrica*, 37, 424-438.
- Granger, C. W. J. (1981) Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics*, 16(1), 121-30.
- Granger, C. W. J. (1988) Some recent developments in the concept of causality. *Journal of Econometrics*, 39, 199-211.
- Granger, C. W. J. and Newbold, P. (1974) Spurious Regression in Econometrics, *Journal of Econometrics*, 2, 111-120.
- Hatemi-J A. (2003) A New Method to Choose Optimal Lag Order in Stable and Unstable VAR Models. *Applied Economics Letters*, 10(3), 135-137.
- Johansen, S. (1988) Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* 12, 231-254. Reprinted in *Long-run Economic Relationships, Readings in Cointegration*, R.F. Engle and C.W.J. Granger (eds.), Oxford University Press.
- Phillips, C. P. (1986) Understanding the Spurious Regression in Econometrics, *Journal of Econometrics*, 33, 311-340.
- Sims, C. A., Stock, J. H., and Watson. M. W. (1990) Inference in Linear Time Series Models with Some Unit Roots. *Econometrica*, 58, 133-44.
- Toda, H. Y. and Yamamoto, T. (1995) Statistical Inference in Vector Autoregressions with Possibly Integrated Processes. *Journal of Econometrics*, 66, 225-250.

Table 1: The size properties of Toda-Yamamoto test statistics, T= 40 without ARCH

Nominal Significance Level ↓	VAR(1)			VAR(2)				VAR(2)			
	Both I(1)	Both I(0)	Mixed I(d)	Both I(2)	Both I(1)	Both I(0)	Mixed I(d)	Both I(2)	Both I(1)	Both I(0)	Mixed I(d)
	One extra lag			One extra lag				Two extra lags			
	% rejecting the null hypothesis using $\chi^2$ distribution										
1%	1.40	1.40	1.20	5.50	2.40	1.70	2.70	2.50	2.30	1.90	2.20
5%	6.30	5.50	5.60	15.40	7.80	6.50	8.50	8.10	7.40	6.50	7.30
10%	11.80	10.80	11.10	23.50	13.40	11.70	14.40	14.30	13.00	11.90	12.60
	% rejecting the null hypothesis using bootstrap distribution										
1%	1.10	1.00	0.80	1.50	1.20	1.10	1.20	1.20	1.20	1.00	1.20
5%	5.40	4.80	4.90	7.20	5.50	5.00	5.60	5.50	5.40	4.80	5.30
10%	10.60	9.90	10.20	13.90	10.70	10.00	11.00	11.10	10.60	9.70	10.40
	Mean absolute deviation using $\chi^2$ distribution										
1%	0.40	0.40	0.30	4.50	1.40	0.70	1.70	1.50	1.30	0.90	1.20
5%	1.30	0.90	1.10	10.50	2.90	1.90	3.60	3.60	2.50	1.90	2.30
10%	1.80	1.00	1.20	1.35	3.40	1.70	4.40	4.30	3.00	1.90	2.60
	Mean absolute deviation using bootstrap distribution										
1%	0.20	0.10	0.30	0.50	0.40	0.20	0.40	0.30	0.30	0.30	0.30
5%	0.60	0.70	0.50	2.30	0.80	0.50	0.90	1.60	0.70	0.50	0.70
10%	0.10	0.80	0.50	3.90	1.10	0.70	1.40	1.30	0.10	0.80	0.90
	% of parameter combinations in which $\chi^2$ distribution is better than bootstrap distribution										
1%	25.00	25.00	37.50	0.00	7.30	15.60	4.20	0.00	4.20	9.40	4.20
5%	25.00	25.00	75.00	0.00	10.40	9.40	6.70	0.00	5.20	18.80	5.40
10%	25.00	50.00	50.00	0.00	8.30	21.90	8.70	0.30	7.30	18.80	8.70
	% of parameter combinations in which bootstrap distribution is better than $\chi^2$ distribution										
1%	75.0	75.0	62.50	100.0	92.70	84.40	94.60	100.0	93.80	87.50	94.60
5%	75.00	75.0	87.50	100.0	89.60	90.60	92.90	100.0	93.80	90.60	93.30
10%	75.0	50.00	87.50	100.0	91.70	78.10	91.20	100.0	91.70	81.30	92.00
	Number of parameter combinations										
	4	4	8	16	96	32	240	16	96	32	240

The shaded areas indicate best relative performance.

Table 2: The size properties of Toda-Yamamoto test statistics, T = 100 without ARCH

Nominal Significance Level ↓	VAR(1)			VAR(2)				VAR(2)			
	Both I(1)	Both I(0)	Mixed I(d)	Both I(2)	Both I(1)	Both I(0)	Mixed I(d)	Both I(2)	Both I(1)	Both I(0)	Mixed I(d)
	One extra lag			One extra lag				Two extra lags			
	% rejecting the null hypothesis using $\chi^2$ distribution										
1%	1.20	0.90	1.40	6.20	1.40	1.30	1.90	1.40	1.30	1.20	1.40
5%	5.60	5.10	5.20	16.90	6.10	5.50	7.30	6.50	5.90	5.70	5.70
10%	11.00	10.70	10.20	25.20	11.40	10.80	13.00	11.70	11.20	10.90	10.90
	% rejecting the null hypothesis using bootstrap distribution										
1%	1.20	0.80	1.40	2.00	1.10	1.10	1.10	1.20	1.10	1.00	1.10
5%	5.40	4.70	5.00	8.70	5.30	5.30	5.00	5.60	5.40	5.10	5.20
10%	10.60	10.50	10.00	15.90	10.40	10.40	10.10	10.90	10.40	10.30	10.20
	Mean absolute deviation using $\chi^2$ distribution										
1%	0.30	0.10	0.40	5.20	0.50	0.30	1.00	0.40	0.40	0.30	0.40
5%	0.50	0.60	0.70	11.90	1.20	0.60	2.40	1.50	1.10	0.80	0.90
10%	1.00	1.20	0.80	15.20	1.60	0.90	3.20	1.80	1.40	0.90	0.12
	Mean absolute deviation using bootstrap distribution										
1%	0.30	0.20	0.40	1.00	0.30	0.20	0.40	0.20	0.40	0.30	0.30
5%	0.40	0.40	0.70	3.70	0.70	0.40	1.00	0.80	1.10	0.50	0.60
10%	0.60	1.00	0.70	5.90	1.00	0.60	1.50	1.00	1.00	0.60	0.90
	% of parameter combinations in which $\chi^2$ distribution is better than bootstrap distribution										
1%	25.0	50.0	25.0	0.0	19.8	18.8	22.1	0.0	17.7	40.6	27.5
5%	50.0	25.0	25.50	0.0	27.1	37.5	25.0	6.3	21.9	28.1	27.9
10%	0.0	25.0	25.50	0.0	27.1	9.4	26.3	6.3	4.20	18.8	32.5
	% of parameter combinations in which bootstrap distribution is better than $\chi^2$ distribution										
1%	25.0	25.0	25.0	93.8	70.8	59.4	67.1	100.0	70.8	56.3	61.3
5%	50.0	50.0	37.5	100.0	97.9	56.3	72.9	100.0	71.9	40.6	72.1
10%	100.0	50.0	75.0	100.0	93.8	62.5	70.4	100.0	69.8	65.6	67.9
	Number of parameter combinations										
	4	4	8	16	96	32	240	16	96	32	240

The shaded areas indicate best relative performance.

Table 3: The size properties of Toda-Yamamoto test statistics, T = 40 with ARCH

Nominal Significance Level ↓	VAR(1)			VAR(2)				VAR(2)			
	Both I(1)	Both I(0)	Mixed I(d)	Both I(2)	Both I(1)	Both I(0)	Mixed I(d)	Both I(2)	Both I(1)	Both I(0)	Mixed I(d)
	One extra lag			One extra lag				Two extra lags			
	% rejecting the null hypothesis using $\chi^2$ distribution										
1%	1.70	1.50	1.30	5.90	2.50	2.10	2.80	2.90	2.40	2.10	2.40
5%	6.40	5.90	6.10	15.50	7.90	6.90	8.60	8.60	7.50	6.90	7.40
10%	11.40	11.20	11.30	24.00	13.60	12.00	14.50	14.60	13.10	12.20	12.70
	% rejecting the null hypothesis using bootstrap distribution										
1%	1.10	1.00	0.80	1.80	1.30	1.20	1.40	1.50	1.30	1.10	1.30
5%	5.40	5.20	5.20	7.60	5.70	5.30	5.80	6.20	5.50	5.30	5.50
10%	10.20	10.00	10.20	14.60	10.90	10.20	11.10	11.40	10.10	10.20	10.50
	Mean absolute deviation using $\chi^2$ distribution										
1%	0.70	0.50	0.30	4.90	1.50	1.10	1.08	1.90	1.40	1.10	1.40
5%	1.30	0.90	1.10	10.50	2.90	1.90	3.60	3.60	2.50	1.90	2.40
10%	1.50	1.20	1.30	1.40	3.60	2.10	4.50	4.60	3.00	2.20	2.70
	Mean absolute deviation using bootstrap distribution										
1%	0.30	0.10	0.30	0.90	0.40	0.40	0.40	0.50	0.40	0.20	0.40
5%	0.70	0.50	0.60	2.60	0.90	0.50	0.10	1.20	0.80	0.60	0.80
10%	0.90	0.20	0.60	4.60	1.20	0.70	1.50	1.70	0.10	0.70	0.90
	% of parameter combinations in which $\chi^2$ distribution is better than bootstrap distribution										
1%	25.0	0.00	50.0	0.00	2.10	6.30	3.70	0.00	2.10	3.10	1.70
5%	0.00	0.00	12.50	0.00	2.10	0.00	5.80	0.00	5.20	9.40	3.70
10%	50.0	0.00	12.50	0.00	6.30	9.40	7.10	6.30	4.20	15.60	7.10
	% of parameter combinations in which bootstrap distribution is better than $\chi^2$ distribution										
1%	75.0	100.0	50.0	100.0	96.9	90.6	95.8	100.0	95.8	90.6	95.8
5%	100.0	100.0	87.5	100.0	97.9	100.0	94.2	100.0	95.8	100.0	97.1
10%	50.0	100.0	87.5	100.0	93.8	90.6	92.9	100.0	95.8	84.4	92.5
	Number of parameter combinations										
	4	4	8	16	96	32	240	16	96	32	240

The shaded areas indicate best relative performance.

Table 4: The size properties of Toda-Yamamoto test statistics, T = 100 with ARCH

Nominal Significance Level ↓	VAR(1)			VAR(2)				VAR(2)			
	Both I(1)	Both I(0)	Mixed I(d)	Both I(2)	Both I(1)	Both I(0)	Mixed I(d)	Both I(2)	Both I(1)	Both I(0)	Mixed I(d)
	One extra lag			One extra lag				Two extra lags			
	% rejecting the null hypothesis using $\chi^2$ distribution										
1%	1.40	1.20	1.40	6.50	1.60	1.40	2.10	1.60	1.50	1.40	1.50
5%	5.50	5.10	5.10	16.90	6.30	5.80	7.40	6.70	6.00	5.80	5.90
10%	10.50	10.20	11.10	25.50	11.60	10.80	13.00	12.00	11.20	10.80	11.00
	% rejecting the null hypothesis using bootstrap distribution										
1%	1.30	1.10	1.30	2.30	1.30	1.20	1.30	1.20	1.30	1.20	1.30
5%	5.30	4.90	5.00	9.10	5.50	5.30	5.70	5.80	5.40	5.20	5.30
10%	10.30	10.00	9.80	15.90	10.50	10.10	11.00	11.10	10.50	10.30	10.30
	Mean absolute deviation using $\chi^2$ distribution										
1%	0.40	0.20	0.50	5.50	0.70	0.40	1.10	0.60	0.50	0.50	0.50
5%	0.50	0.50	0.40	11.90	1.70	0.90	2.50	1.70	1.10	0.90	0.10
10%	0.50	0.50	0.60	15.50	1.30	0.80	3.20	2.00	1.30	1.10	1.20
	Mean absolute deviation using bootstrap distribution										
1%	0.40	0.10	0.30	1.30	0.40	0.30	0.40	0.40	0.40	0.30	0.30
5%	0.30	0.40	0.40	4.10	0.80	0.60	1.10	1.00	0.70	0.60	0.60
10%	0.40	0.60	0.60	6.10	1.10	0.60	1.50	1.70	0.10	0.70	0.90
	% of parameter combinations in which $\chi^2$ distribution is better than bootstrap distribution										
1%	25.00	0.00	25.0	0.00	5.20	18.80	15.40	12.50	22.90	18.80	15.40
5%	25.00	25.0	37.50	0.00	25.00	21.90	19.60	6.30	19.80	21.90	18.80
10%	25.00	50.00	62.50	12.50	21.90	31.30	25.40	0.00	27.10	21.90	27.90
	% of parameter combinations in which bootstrap distribution is better than $\chi^2$ distribution										
1%	0.00	75.0	75.0	100.0	80.2	65.6	79.9	93.8	74.0	71.9	70.4
5%	75.0	25.0	50.0	100.0	75.0	71.9	79.9	100.0	76.0	62.5	74.6
10%	50.0	25.0	25.0	100.0	77.1	65.6	73.8	87.5	74.0	59.4	69.6
	Number of parameter combinations										
	4	4	8	16	96	32	240	16	96	32	240

The shaded areas indicate best relative performance.