Application of Bootstrap Methods in Investigation of Size of the Granger Causality Test for Integrated VAR Systems

Łukasz Lach

This paper examines the size performance of the Toda-Yamamoto test for Granger causality in the case of trivariate integrated and cointegrated VAR systems. The standard asymptotic distribution theory and the residual-based bootstrap approach are applied. A variety of types of distribution of error term is considered. The impact of misspecification of initial parameters as well as the influence of an increase in sample size and number of bootstrap replications on size performance of Toda-Yamamoto test statistics is also examined. The results of the conducted simulation study confirm that standard asymptotic distribution theory may often cause significant over-rejection. Application of bootstrap methods usually leads to improvement of size performance of the Toda-Yamamoto test. However, in some cases the considered bootstrap method also leads to serious size distortion and performs worse than the traditional approach based on χ^2 distribution.

Key Words: bootstrap methods, simulation, Granger causality, VAR models
JEL Classification: C12, C15

Introduction

The causal relationship (in the Granger sense) between some considered variables is one of the most important issues in modern economics. The existence of this type of dynamic link guarantees that the knowledge of past values of one considered time series is useful in predicting current and future values of another one. Since the development of this concept (Granger 1969) a number of studies examining properties of different testing methods have been published. One of the first approaches was the standard Wald test based on asymptotic distribution theory. The biggest advantage of this method was its simplicity and clarity. However, in case of variables which are integrated of order one ($I(1)$) or cointegrated, the standard asymptotic approach turned out to be an improper

Łukasz Lach is an assistant in the Department of Economics and Econometrics, AGH University of Science and Technology, Poland.

tool for testing the causal effects. These nonstandard asymptotic properties of the Wald test were investigated by Granger and Newbold (1974, empirical findings) and Philips (1986, theoretical framework). As a cure for this problem the idea of the Vector Error Correction Model (see Engle and Granger (1987) and Granger (1988)) was developed. Although theoretically it was a useful tool for testing for causality in integrated-cointegrated VAR systems, the complicated pretesting procedure (estimation of unit roots, analysis of cointegration properties and sensitivity for improper lag establishment) turned out to be a serious difficulty in empirical applications.

Another solution was proposed by Toda and Yamamoto (1995). This approach ensures that asymptotic distribution theory is valid for VAR systems, regardless of the order of integration of considered variables or the dimension of cointegration space. Furthermore, the important advantage of this method is its simplicity since it is just a small modification of the standard Wald test. The absence of pretesting bias made this procedure one of the most widely applied approaches in recent economic research. However, when some standard assumptions do not hold (especially concerning the distribution of error term) the Toda-Yamamoto approach is also likely to fail. Application of the bootstrap approach may often provide better results since bootstrapping does not strictly depend on model specification (for more details on bootstrap see Efron (1979)).

The properties of the augmented Wald test in both the asymptotic and bootstrap variant were examined by a number of authors in recent years. Dolado and Lütkepohl (1996) conducted a simulation exercise to examine the power of the considered testing method in the case of the integrated VAR model (in this paper the error term was independently drawn from identical multivariate normal distribution). Their outcomes show that in high dimensional VARs with a small true lag length the significant reduction of power of the considered causality test may occur, especially for small samples. Mantalos (2000) conducted similar studies of size and power properties of eight versions of the Granger causality test (this time the error term was only $N(0, I_2)$ i. i. d.). His findings indicate that the standard asymptotic approach may often lead to significant size distortion. Application of the residual-based bootstrap technique usually improves the size and power performance of causality tests. Hacker and Hatemi (2006) examined size properties of the TY (Toda-Yamamoto) test for two-dimensional VAR systems. In contrast to previously mentioned authors, they also investigated the simple ARCH(1) case for error term

series, finding that the bootstrap technique performed relatively well in all cases. On the other hand they restricted the research only to models without cointegration.

This paper is a generalization of previous studies concentrated on investigation of size properties of the TY test. The simulation study contained in this article (in both asymptotic and bootstrap variants) examines three-dimensional integrated and cointegrated VAR models. All possible cointegration ranks are also considered. To check the size properties of the investigated test (also in cases where some standard assumptions do not hold) a variety of distributions of error term is applied in DGP (spherical multivariate normal distribution, highly correlated error terms, structural break, mixture of distributions, ARCH(2) effect). The impact of misspecification of initial parameters is also examined in each case. Finally, the impact of increase of sample size (from small to medium) as well as the influence of increase in the number of bootstrap replications on size performance of the TY test is examined in some specific cases. To the knowledge of the author, the results of this kind of study of size performance of the TY test in both asymptotic and bootstrap variant have not been published so far.

This paper is organized as follows. The next section contains the main research hypotheses to be tested by the simulation study. Section 3 provides details on the methodology of the TY test, specification of VAR models used for simulation purposes and the considered bootstrap technique. Section 4 contains results of all conducted simulations. Section 5 concludes the paper.

Main Hypotheses

The main objective of this paper is the investigation of size properties of the Toda-Yamamoto test for Granger causality. The first important point that distinguishes this study from the existing literature is the use of the trivariate VAR model for simulation purposes. Most of the previous papers examine two-dimensional models. In the three-dimensional case the structure of causal links may be more extended. Another important point is the fact that this paper examines all possible dimensions of cointegration space. As already mentioned, former studies concentrating on a similar topic provided evidence of poor performance of the modified Wald procedure in the case of nonstationary variables. Thus, it seems to be reasonable to formulate:

H₁ *The Toda-Yamamoto test (asymptotic variant) often tends to over-*

> *reject the null hypothesis for integrated and cointegrated* VAR *systems (with various cointegration ranks).*

There are some ways to avoid the mentioned problem. One of the possibilities is the application of bootstrap methods. This approach has been commonly used in recent years despite its numerical complexity. Thus, one may be interested in testing the following hypothesis:

H$_2$ *The residual-based bootstrap method usually improves size performance of the* TY *test.*

In practice the proper specification of the VAR model is often difficult to obtain. One of the most common problems is the misspecification of lag parameter. Previous studies (see Hacker and Hatemi (2006) and Mantalos (2000)) show that in this case the size performance of the TY test (asymptotic variant) may significantly worsen. It may be interesting to determine how the bootstrap-based technique performs in this case. Therefore, we should test:

H$_3$ *Misspecification of lag parameter in the* VAR *model leads to considerable aggravation of size performance of* TY *only in the asymptotic variant.*

Despite the fact that bootstrap methods are often a useful tool to overcome the problem of size distortion in the TY test, there are some specific cases where this approach may also fail. One important point that distinguishes this study from the existing literature is the fact that, in order to perform suitable simulation, a variety of types of error term distribution was used (possibilities, where some standard assumptions about the structure of the considered VAR models and TY methodology are unfulfilled, are examined). Therefore, this paper contains verification of the following:

H$_4$ *Residual-based bootstrap is likely to fail in some specific cases and therefore should not be used without second thought.*

One of the main problems with the application of standard asymptotic distribution theory is the sample size. Previous papers provided empirical proof that the increase of sample size may significantly improve size performance of the TY test (see Dolado and Lütkepohl 1996; Hacker and Hatemi 2006; Mantalos 2000). However, this process may strongly depend on model specification (especially the error term structure). Thus, it seems to be interesting to test the following hypothesis:

H$_5$ *When standard assumptions hold, the increase of sample size improves size performance of the* TY *test (asymptotic variant).*

In order to apply the bootstrap technique the researcher must establish the number of bootstrap replications. In previous papers this number varied significantly (from dozens to hundreds). It may be interesting to investigate whether the change of number of bootstrap replications may lead to significant improvement of size performance of the TY test in some specific cases (namely, cases of relatively significant size distortion). This problem may be captured in verification of following:

H$_6$ *There is a relationship between the number of bootstrap replications and size performance of the* TY *test in some specific cases.*

In order to test the above research hypotheses some simulation study must be performed. In the first step, comprehensive analysis of the considered methodology and DGP should be presented. The next section contains some essential information concerning methodology and data.

Methodology and the Data Generating Process

In this article the Toda-Yamamoto approach for testing Granger causality is considered. This method has been commonly applied in recent studies since it is relatively simple to perform and free of complicated pretesting procedures. Another issue worth underlying is the fact that this method is useful for integrated and cointegrated systems. To understand the idea of this type of causality testing consider the following n-dimensional VAR(p) process:

$$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + \varepsilon_t, \tag{1}$$

where $y_t = (y_t^1, \ldots, y_t^n)^{tr}$, $c = (c_1, \ldots, c_n)^{tr}$ and $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})^{tr}$ are n-dimensional vectors, and $\{A_i\}_{i=1}^{p}$ is a set of $n \times n$ matrices of parameters for appropriate lags (in this paper transpose of matrix M is denoted by M^{tr}). The order p of the process is assumed to be known. Furthermore, we shall assume that the error vector is an independent white noise process with nonsingular covariance matrix $\sum_\varepsilon$ (the elements of which are constant over time). In this article cases where these standard assumptions do not hold are also investigated. We also assume that the condition $E|\varepsilon_{k,t}|^{s+2} < \infty$ holds true for all $k = 1, \ldots, n$ and some $s > 0$. The Toda-Yamamoto (1995) idea of testing for causal effects is based on esti-

TABLE 1   Compact notation used to formulate TY test statistics

Object	Description
$Y := (y_1, \ldots, y_T)$	$n \times T$ matrix
$\hat{D} := (\hat{c}, \hat{A}_1, \ldots, \hat{A}_p, \ldots, \hat{A}_{p+d})$	$n \times (1 + n(p + d))$ matrix
$Z_t := \begin{bmatrix} 1 \\ y_t \\ y_{t-1} \\ \ldots \\ y_{t-p-d+1} \end{bmatrix}$	$(1 + n(p + d)) \times 1$ matrix, $t = 1, \ldots, T$
$Z := (Z_0, \ldots, Z_{T-1})$	$(1 + n(p + d)) \times T$ matrix
$\hat{\delta} := (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)$	$n \times T$ matrix

mating the augmented VAR($p + d$) model (circumflex indicates the OLS estimator of a specific parameter):

$$y_t = \hat{c} + \sum_{i=1}^{p+d} \hat{A}_i y_{t-i} + \hat{\varepsilon}_t. \tag{2}$$

The value of parameter d is equal to the maximum order of the integration of considered variables $y^1, \ldots, y^n$. We say that the k-th element of y_t does not Granger-cause the j-th element of $y_t (k, j \in \{1, \ldots, n\})$ if there is no reason for rejection of the following hypothesis:

$$\text{H}_0: a_{jk}^s = 0, \tag{3}$$

for $s = 1, \ldots, p$, where $A_s[a_{pq}^s]_{p,q=1,\ldots,n}$ for $s = 1, \ldots, p$. According to Toda and Yamamoto (1995) the number of extra lags (parameter d) is an unrestricted variable since its role is to guarantee the use of asymptotic theory. In order to present the test statistics we shall make use of the compact notation (T denotes the considered sample size) presented in table 1.

The initial point of the considered procedure is the calculation of $S_U = \hat{\delta}\hat{\delta}^{tr}/T$ – the variance-covariance matrix of residuals from the unrestricted augmented model (i. e. model (2)). Then we can define $\beta := vec(c, A_1, \ldots, A_p, 0_{n \times nd})$ and $\hat{\beta} := vec(\hat{c}, \hat{A}_1, \ldots, \hat{A}_p, \ldots, \hat{A}_{p+d})$ where $vec(\cdot)$ denotes the column stacking operator and $0_{n \times nd}$ stands for the $n \times nd$ matrix filled with zeros. Using this notation one can write the Toda-Yamamoto test statistics for testing for causal effects between variables in y_t in the following form:

$$\text{TY} := (C\hat{\beta})^{tr}(C((ZZ^{tr})^{-1} \otimes S_U)C^{tr})^{-1}(C\hat{\beta}), \tag{4}$$

where $\otimes$ denotes Kronecker product and C is the matrix of suitable linear restrictions. In our case (testing for causality from one variable in y_t to another) C is $p \times (1 + n(p + d))$ matrix, the elements of which take only the value of zero or one. Each of p rows of matrix C corresponds to a restriction of one parameter in β. The value of every element in each row of C is one, if the associated parameter in β is zero under the null hypothesis and it is zero otherwise. There is no association between matrix C and the last $n^2 d$ elements in β. This approach allows us to write the null hypothesis of non-Granger causality in the following form:

$$\text{H}_0: C\beta^{tr} = 0. \tag{5}$$

Finally we shall note that the TY test statistic is asymptotically χ^2 distributed with the number of degrees of freedom equal to the number of restrictions to be tested (in our case this value is equal to p). In other words, the TY test is just a standard Wald test applied for the first p lags obtained from the augmented $\text{VAR}(p + d)$ model.

In order to examine the size properties of the TY test some $\text{I}(1)$ models are considered. Causality tests are conducted in the case of various cointegration ranks. At this place we shall once again consider model (1). This process can be rewritten in the following error correction form:

$$\Delta y_t = c + \prod y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t, \tag{6}$$

where $\prod = -\text{I} + \sum_{i=1}^{p} A_i$ and $\Gamma_i = -\sum_{j=i+1}^{p} A_j$. To ensure that y_t is integrated of order one the following assumptions must hold (these assumptions are sufficient to prove the so-called Johansen-Granger representation theorem, for more details see Johansen 1991; 1996):

- The roots of the characteristic polynomial:

$$\det(\text{I}_n - A_1 z - A_2 z^2 - \cdots - A_p z^p) \tag{7}$$

  are either outside the unit circle or equal to one;
- The matrix $\prod$ has reduced rank $r < n$ and therefore may be expressed as the product $\prod = \alpha\beta^{tr}$, where α and β are $n \times r$ matrices of full column rank r;
- The matrix $\alpha_\perp^{tr}\Gamma\beta_\perp$ has full rank, where $\Gamma = I - \sum_{i=1}^{p}\Gamma_i$ and where $\alpha_\perp$ and $\beta_\perp$ are the orthogonal complements to α and β.

TABLE 2   Specification of trivariate VAR models considered in this paper

Matrix form	Properties	Symbol
$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	No cointegration	A_1
$A = \begin{bmatrix} 1 & 0 & -0,125 \\ 0 & 1 & 0 \\ 0,5 & 0,5 & 0,5 \end{bmatrix}$	Two cointegrating equations	A_2
$A = \begin{bmatrix} 0,25 & 0 & -0,125 \\ 0 & 1 & 0 \\ -0,75 & 0 & 0,875 \end{bmatrix}$	One cointegrating equation	A_3

If the first assumption holds, then the considered process is neither explosive (roots in the unit circle) nor seasonally cointegrated (roots on the boundary of the unit circle different from $z = 1$, for more details on this issue see Hylleberg et al. 1990; Johansen and Schaumburg 1988). The second assumption ensures that there are at least $n - r$ unit roots. Cointegration occurs whenever $r > 0$ and the number of cointegrating vectors is equal to r. To restrict the process from being $I(2)$, we shall assume the last condition, because together with the second one it ensures that the number of unit roots is exactly $n - r$.

In this paper trivariate VAR models are considered. In each the case process described by the model is integrated of order one and the parameter p is equal to one. Therefore, we consider the following VAR(1) model which is used as a DGP:

$$y_t = c + Ay_{t-1} + \varepsilon_t, \tag{8}$$

where $c(0, 01\ 0, 01\ 0, 01)^{tr}$ in all cases and matrix A provide specific cointegration properties (see previously presented assumptions). For details about matrices used in the simulation study explore table 2.

Directly from table 2 we can obtain some essential information. Namely, in A_2 and A_3 models y^3 is a causal variable for y^1. Furthermore, in all considered cases y^2 does not Granger-cause y^1 (this will be our null hypothesis for further analysis of size performance). We should underline that in three-dimensional VAR models the relationship between y^3 and y^1, as well as between y^3 and y^2, may have indirect impacts on links between y^2 and y^1. Beside various schemes of algebraic structure,

TABLE 3 Models used to generate distribution of error term

Distribution of error term	Parameters	Symbol
$N(0_{3\times1}, \sigma^2 I_3)$	$\sigma = 1$	E_1
$N(\mu, \sigma)$	$\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0,9 \\ 0 & 0,9 & 1 \end{bmatrix}$	E_2
$N(0_{3\times1}, \sigma_1^2 I_3)$ *for* $t = 1, \ldots, T/2$ $N(0_{3\times1}, \sigma_2^2 I_3)$ *for* $t = T/2 + 1, \ldots, T$	$\sigma_1 = 1, \sigma_2 = 2$	E_3
$sN_1 + (1 - s)N_2$, where: $N_1 \sim N(0_{3\times1}, \sigma_1^2 I_3)$, $N_2 \sim N(0_{3\times1}, \sigma_2^2 I_3)$, $P(s = 1) = p, \ P(s = 0) = 1 - p$	$\sigma_1 = 1, \sigma_2 = 3, p = 0,7$	E_4
$\varepsilon_{j,t} = w_{j,t} \sqrt{0,5 + 0,1\varepsilon_{j,t-1}^2 + 0,4\varepsilon_{j,t-2}^2}$, $w_{j,t} -$ i. i. d. $N(0,1)$	$j = 1, 2, 3, t = 1, \ldots, T$	E_5

some specific distributions of error vectors are also examined. At this place it should be noted that in previous studies concentrating on similar topics the error term was usually $N(0_{n\times1}, \sigma^2 I_n)$ distributed (n stands for considered dimension) for some positive σ (see Hacker and Hatemi 2006; Dolado and Lütkepohl 1996; these authors also consider the case of nonzero covariance between components of error term); Mantalos 2000). In this paper the size properties of the TY test are examined for variety of types of time structure of the error term. In some considered specifications the standard assumptions for the TY method do not hold. Some fundamental information is contained in table 3 (random draw for error term is always based on i. i. d. variables – normal, discrete uniform).

In this paper, beside the standard three-dimensional spherical multivariate normal distribution (denoted as E_1), the situation where vectors $\varepsilon_{2,t}$ and $\varepsilon_{3,t}$ are highly correlated (E_2) is also investigated. In this case the variance-covariance matrix S_U is 'nearly singular,' which may often lead to problems with application of bootstrap methods (see Horovitz 1995; or Chou and Zhou 2006). Another specification of the distribution of error term series is related to the structural break (E_3). It is a well known fact that in this case huge size distortions may occur while testing for Granger causality. Another question is whether application of the bootstrap approach may significantly improve the investigated size properties. The fourth examined possibility (E_4) is related to the idea of a mixture of distributions. The last considered DGP for error vector (E_5)

is a simple ARCH(2) model with constant unconditional variance (equal to one). A similar type of time dependence structure in the error term series was examined by Hacker and Hatemi (2006) (the authors used ARCH(1) model for VAR(1) and VAR(2) processes).

As a cure for the effect of start-up values, 50 presample observations of y_t are generated for each simulation study. Some of these data points (based on random draw from $N(0,1)$ distribution) are used as the initial observations for VAR models. To make the results of the presented research more comparable, the same random draw from $N(0,1)$ distribution is also used for every type of the error term analyzed. Namely, to create the $E_2 = (E_{2,t})_{t=1,...,T}$ series, the following transformation of $E_1 = (E_{1,t})_{t=1,...,T}$ series is applied:

$$E_{2,t} = ZE_{1,t}, \tag{9}$$

where $t = 1, \ldots, T$ and $ZZ^{tr} = \sum$ (Cholesky decomposition). The values of the E_1 series are also used in the process of generation of E_4 series and E_3 series (for first $T/2$ observations). In order to generate E_5 series, initial observations are once again drawn from $N(0,1)$ distribution and $(w_{1,t} \ w_{2,t} \ w_{3,t})^{tr} = E_{1,t}$ for $t = 1, \ldots, T$.

To examine the size properties of the considered test a set of simulated observations is generated each time (using model (1) with specific A_i and E_j) and the TY test statistics are calculated to test the hypothesis that y^2 does not Granger-cause y^1. Typical significance levels (namely, 1%, 5% and 10%) are considered, and both the asymptotic distribution theory (as noted by Toda and Yamamoto) and a residual-based bootstrap approach are used to get suitable critical values.

Let me now discuss shortly the bootstrap methods used in this paper. All bootstrap simulations conducted for the use of this article are based on resampling leveraged residuals. The application of leverages is the simple modification of regression raw residuals, which helps to stabilize their variance (for more details on this issue see Davison and Hinkley 1999; Hacker and Hatemi 2006). At first the considered augmented VAR model (2) is estimated through OLS methodology with the null hypothesis assumed (that is: y^2 does not Granger-cause y^1). Many authors use OLS methodology in their empirical research, although other estimation methods are more adequate for their data. This paper partly investigates the influence of the mentioned approach on performance of the considered causality tests. In the next step, regression raw residuals are transformed with the use of leverages (modified residuals will be denoted as $\{\hat{\varepsilon}_i^m\}_{i=1,...,T}$). Finally, the following algorithm is conducted:

- draw randomly with replacement (each point has a probability measure equal to $1/T$) from the set $\{\hat{\varepsilon}_i^m\}_{i=1,\ldots,T}$ (as a result we get the set $\{\hat{\varepsilon}_i^{**}\}_{i=1,\ldots,T}$);
- subtract the mean to guarantee that the mean of bootstrap residuals is zero (in this way we create the set $\{\hat{\varepsilon}_i^*\}_{i=1,\ldots,T}$, such that

$$\hat{\varepsilon}_{k,i}^* = \hat{\varepsilon}_{k,i}^{**} - \frac{\sum_{j=1}^{T}\hat{\varepsilon}_{k,j}^{**}}{T},$$

$i = 1,\ldots,T, \ k = 1,2,3$);

- generate the simulated data $\{y_i^*\}_{i=1,\ldots,T}$ through use of the original data $(\{y_i\}_{i=1,\ldots,T})$, coefficient estimates from the regression $(\hat{c}, \{\hat{A}_i\}_{i=1,\ldots,p+d})$ and the bootstrap residuals $\{\hat{\varepsilon}_i^*\}_{i=1,\ldots,T}$;
- calculate the TY test statistics.

After repeating this procedure $N = 250$ times it is possible to create the empirical distribution of TY test statistics and next obtain empirical critical values (bootstrap critical values). The suitable procedure (which allows one to conduct every type of simulation presented in this article) written in Gretl is available from the author upon request.

Empirical Results

In this section, results of the conducted causality tests are presented. The following tables contain the rejection rates obtained while testing the null hypothesis in the TY test with the application of both the standard asymptotic distribution theory and the residual-based bootstrap approach. In recent years the problem of establishing adequate significance levels for diagnostic applications has been intensively discussed. Some researchers recommended relatively large levels (Maddala 1992), while others argue that typical values are the best choice (MacKinnon 1992). As already mentioned in this article, typical significance levels are considered. Thus the results of the presented simulations are more comparable with the similar research conducted by Hacker and Hatemi (2006) and Mantalos (2000). To judge whether empirical rejection rates are significantly different from considered nominal sizes for each significance level, the 95% two-sided confidence intervals were created by the following expression:

$$Ts \pm 2\sqrt{\frac{Ts(1-Ts)}{N_r}}, \tag{10}$$

where *Ts* denotes the considered nominal size (1%, 5%, 10%) and $N_r = 1000$ stands for the number of repetitions. This value ($N_r = 1000$) was also used by Dolado and Lütkepohl (1996), Hacker and Hatemi (2006) and Mantalos (2000). Furthermore, the considered type of confidence intervals was used by Dolado and Lütkepohl (1996) and Mantalos (2000). In this way, the intervals [0.4%; 1.6%], [3.6%; 6.4%], [8.1%; 11.9%] were established for 1%, 5% and 10% significance levels respectively. The considered approach leads to the criteria of bad performance, namely, the actual test size is significantly distorted whenever it lies outside the suitable confidence interval. In the following tables these findings are indicated by bold typeface. In each case the parameter d (maximal order of integration of considered variables) is equal to one (properly specified). For tables 4–9 the considered sample size is $T = 40$ (small sample size).

First we shall focus on cases where parameter p was chosen properly. Suitable results are contained in tables 4–6.

After analyzing the results contained in table 4, one can easily see that the asymptotic distribution theory was found to cause serious size distortions in almost all cases. The largest distortions were indicated in the case of structural change in error term distribution (E_3). Furthermore, it should be noted that whenever critical values were taken from suitable χ^2 distribution the over-rejection was indicated, which seems to prove that Hypothesis 1 is true. The application of the bootstrap method improved the size properties of the TY test for all significance levels in cases of E_1, E_4 and E_5 distribution. These results provided a strong basis for claiming that Hypothesis 2 is also true. Although the significant over-rejection was still found for E_3 error distribution (except for 10% level), size distortions were much smaller than in the non-bootstrap approach. However, one must note that the bootstrap test was found to under-reject the null hypothesis in the case of E_2 distribution, which led to significant size distortions by 5% and 10% significance levels (even worse performance than for χ^2 distribution). The outcomes obtained by Hacker and Hatemi (2006) in corresponding research conducted for similar two-dimensional cases (A_1 model, E_1 and E_5 error term) are in line with the results presented in table 4.

The outcomes contained in table 5 and 6 also lead to some interesting regularities and provide no significant reason for rejection of Hypothesis 1 or Hypothesis 2. Firstly, they confirmed the hypothesis that the TY test based on asymptotic distribution theory tends to over-reject the null hypothesis also when there exists cointegration between considered vari-

TABLE 4 Size of TY test for Granger causality – no-cointegration case

(1)	(2)	(3)	χ^2 distribution			Bootstrap distribution		
			1%	5%	10%	1%	5%	10%
A_1	E_1	1	**1.7%**	6.1%	**13.2%**	0.8%	4.6%	10.9%
	E_2	1	**1.9%**	5.6%	11.6%	0.4%	**2.9%**	**7.7%**
	E_3	1	**7.7%**	**15.3%**	**20.6%**	**2.8%**	**7.4%**	10.8%
	E_4	1	**1.7%**	**7.8%**	**12.4%**	0.6%	4.2%	9.6%
	E_5	1	1.4%	**6.5%**	11.2%	0.8%	5.2%	9.1%

NOTES   Column headings are as follows: (1) algebraic structure, (2) distribution of error term, (3) lag p.

TABLE 5 Size of TY test for Granger causality – case of two cointegrating vectors

(1)	(2)	(3)	χ^2 distribution			Bootstrap distribution		
			1%	5%	10%	1%	5%	10%
A_2	E_1	1	0.8%	**3.5%**	10.8%	0.9%	4.8%	9.9%
	E_2	1	1.2%	5.5%	**14%**	1%	4.9%	11%
	E_3	1	**5%**	**14%**	**25%**	**3.6%**	**8.9%**	**18%**
	E_4	1	**1.9%**	**6.7%**	**14%**	1.1%	5.3%	**12%**
	E_5	1	1.5%	**6.8%**	11%	1.1%	4.7%	10.5%

NOTES   Column headings are as follows: (1) algebraic structure, (2) distribution of error term, (3) lag p.

TABLE 6 Size of TY test for Granger causality – case of one cointegrating vector

(1)	(2)	(3)	χ^2 distribution			Bootstrap distribution		
			1%	5%	10%	1%	5%	10%
A_3	E_1	1	1.2%	**7.4%**	11.5%	0.9%	6.1%	10.6%
	E_2	1	**2.6%**	5.8%	**14.7%**	**0.2%**	**2.1%**	**5.2%**
	E_3	1	**6.7%**	**11.6%**	**26%**	**2.4%**	5.9%	11.4%
	E_4	1	**2.5%**	**8%**	**15.6%**	0.8%	4.7%	10.6%
	E_5	1	1.5%	5.9%	**12.6%**	0.7%	4.2%	9%

NOTES   Column headings are as follows: (1) algebraic structure, (2) distribution of error term, (3) lag p.

ables (Dolado and Lütkepohl (1996) and Mantalos (2000) examine cointegration ranks which are no greater than one). Secondly, they provided a basis for claiming that the application of bootstrap methods leads to reduction of actual test size in comparison to the asymptotic method.

However, this reduction is still insufficient for the A_2 algebraic structure and E_3 error distribution scheme (still over-rejection) and too intensive for the A_3 and E_2 case (under-rejection, worse performance in comparison to χ^2 distribution on 5% and 10% significance levels).

In practice it is often difficult to establish the lag parameter properly before estimating the var model. Despite the variety of econometric methods (aic, bic, fpe information criteria, more recently Hatemi's (2003) criterion) many researchers are still struggling to decide what value of lag length to choose for further analysis. In the context of our investigation this problem was examined by the repetition of all causality tests in case of a misspecified value of parameter p (set at the level of 2). For clarity it should be mentioned that true dgp was unchanged. The results are shown in tables 7–9.

It seems to be obvious that the results contained in tables 7–9 should be analyzed together with corresponding outcomes from previously presented cases (contained in tables 4–6 respectively). After analyzing the results contained in table 7 (no-cointegration case) one can easily see that the standard approach (based on χ^2 distribution) causes even stronger over-rejection (higher rejection rates) than in the corresponding case (table 4). On the other hand, the results obtained with application of the bootstrap method belong to suitable confidence intervals in all except for one case (in comparison to the corresponding case). For the model with two cointegrating vectors (A_2) the actual test size (case of χ^2 distribution) is too high in all except for 3 cases. This means that misspecification of parameter p considerably worsens size performance of the ty test. Furthermore, the actual size of the bootstrap test was found to lie outside the confidence interval for exactly the same combination of considered significance levels and error term schemes, like in the corresponding case (table 5). The standard asymptotic approach was also found to cause serious over-rejection for the A_3 structure in almost all cases. On the other hand, actual test size based on the bootstrap method was distorted only for the E_2 (under-rejection) and E_3 (over-rejection) case. In general, size performance of the ty test worsened significantly only for the asymptotic variant, which allows us to claim that Hypothesis 3 is true. Furthermore, the results contained in tables 4–6 as well as in tables 7–9 strongly indicate that Hypothesis 4 is also true (see results obtained for the E_2 and E_3 cases).

Additionally, to examine the size performance of the ty test in both considered variants, causality tests were conducted for a longer sample.

TABLE 7 Size of TY test for Granger causality – no-cointegration case, misspecified parameter p

(1)	(2)	(3)	χ^2 distribution			Bootstrap distribution		
			1%	5%	10%	1%	5%	10%
A_1	E_1	2	**2.1%**	**10.6%**	**16%**	0.9%	4.5%	10.2%
	E_2	2	**1.8%**	**6.5%**	**13.5%**	0.8%	**3.1%**	**7.1%**
	E_3	2	**9%**	**19%**	**33%**	**4.5%**	**9.1%**	**18.5%**
	E_4	2	**1.8%**	**9%**	**15.5%**	0.9%	4.6%	9.5%
	E_5	2	1.4%	4.6%	**14%**	0.7%	4.1%	9.3%

NOTES Column headings are as follows: (1) algebraic structure, (2) distribution of error term, (3) lag p.

TABLE 8 Size of TY test for Granger causality – case of two cointegrating vectors, misspecified parameter p

(1)	(2)	(3)	χ^2 distribution			Bootstrap distribution		
			1%	5%	10%	1%	5%	10%
A_2	E_1	2	1.3%	6.1%	**12.8%**	1.2%	4.6%	9.4%
	E_2	2	1.4%	**7.2%**	**13.6%**	0.8%	4.8%	9.6%
	E_3	2	**8.5%**	**20%**	**27%**	**6.1%**	**14%**	**19.7%**
	E_4	2	**2.8%**	**6.8%**	**17.1%**	0.8%	4.8%	**13.4%**
	E_5	2	**2.1%**	**8.4%**	**12.7%**	1.1%	5.6%	9.7%

NOTES Column headings are as follows: (1) algebraic structure, (2) distribution of error term, (3) lag p.

TABLE 9 Size of TY test for Granger causality – case of one cointegrating vector, misspecified parameter p

(1)	(2)	(3)	χ^2 distribution			Bootstrap distribution		
			1%	5%	10%	1%	5%	10%
A_3	E_1	2	1.4%	6.3%	**14.1%**	0.9%	4.8%	10.3%
	E_2	2	**3.9%**	**8.2%**	**14.8%**	**0.1%**	**1.9%**	**5.1%**
	E_3	2	**7.6%**	**13.6%**	**29%**	**3.9%**	**8.3%**	**14.7%**
	E_4	2	**2.8%**	**9.2%**	**17.6%**	1.1%	4.4%	11.3%
	E_5	2	**2.2%**	**8.5%**	**13.9%**	0.8%	4.6%	9.5%

NOTES Column headings are as follows: (1) algebraic structure, (2) distribution of error term, (3) lag p.

One should expect the standard asymptotic approach to perform relatively better in this case. Suitable tests were conducted for the sample size

TABLE 10   Impact of increase of sample size on size properties of TY test for Granger causality – no-cointegration case

(1)	(2)	(3)	χ^2 distribution			Bootstrap distribution		
			1%	5%	10%	1%	5%	10%
A_1	E_1	1	1.1%	6.2%	**12%**	0.9%	4.2%	9.5%
			(1.7%)	(6.1%)	**(13.2%)**	(0.8%)	(4.6%)	(10.9%)
	E_1	2	1.3%	5.6%	**13.5%**	1.1%	4.9%	10.3%
			(2.1%)	(10.6%)	**(16%)**	(0.9%)	(4.5%)	(10.7%)

NOTES   Column headings are as follows: (1) algebraic structure, (2) distribution of error term, (3) lag p.

TABLE 11   Size of TY test for Granger causality – different number of bootstrap replications in specific cointegrated systems

(1)	(2)	(3)	χ^2 distribution			Bootstrap distribution			N
			1%	5%	10%	1%	5%	10%	
A_2	E_3	2	**8.5%**	**20%**	**27%**	**9.1%**	**19.6%**	**24%**	100
						5.2%	**16.3%**	**22.1%**	200
						6.1%	**13.5%**	**20.1%**	300
A_3	E_2	2	**3.9%**	**8.2%**	**14.8%**	**0%**	**3%**	**3.4%**	100
						0.5%	2.5%	**5.5%**	200
						0.6%	1.2%	**3.5%**	300

NOTES   Column headings are as follows: (1) algebraic structure, (2) distribution of error term, (3) lag p.

$T = 100$ and no-cointegration model with parameter $p = 1$ and $p = 2$ (Hacker and Hatemi (2006) also considered a sample size equal to $T = 40$ (small sample) and $T = 100$ (medium sample)). For comparability with previous results (obtained for $T = 40$), the first 40 data points were exactly the same. Once again the true value of parameter d was assumed to be known. The results are presented in table 10. For clarity it should be noted that values in parentheses denote the rejection rates obtained in a similar investigation conducted for a small sample ($T = 40$).

The analysis of the above table confirmed the hypothesis that size properties of the TY test for Granger causality are improving with the increase of sample size. Although for a 10% significance level the actual size of tests still lies outside the 95% confidence interval, the increase of sample size moved actual size closer to the nominal one. Furthermore, the actual size of bootstrap tests was again found to lie with in suitable confidence intervals in all cases. On the other hand, it should be noted

that for other the considered distributions of error term (E_2, E_3, E_4, E_5) such significant improvement of size performance was not found in considered algebraic specification (A_1). All these facts confirm that there is no significant reason for the rejection of Hypothesis 5.

One of the initial arbitrary decisions in every bootstrap application is the establishment of the number of replications. In previous research concentrated on similar investigation this value varied significantly. Horovitz (1994) used 100 replications, Mantalos (2000) – 200, Hacker and Hatemi (2006) – 800, while Davidson and MacKinnon (1996) used 1000 replications to create bootstrap distribution each time. Increase of the number of replications may often have an important impact on improvement of performance of the TY test size. However in some situations bootstrap methods are likely to fail, regardless of the number of replications used (Horovitz 1995). This paper takes part in the discussion of the mentioned problem, as it contains results of some simulations based on different numbers of bootstrap replications. The investigation covers two specific cases in which the size distortion of bootstrap distribution was relatively largest and far away from 95% confidence intervals (namely, high correlation and structural change cases). It should be noted that for comparability with the previously presented outcomes (conducted for 250 bootstrap replications) the same series of random numbers were used to generate the data. Therefore, the actual size of the TY test conducted with application of χ^2 distribution was unchanged. Parameter d was again assumed to be known ($d = 1$). The examined number of bootstrap replications was denoted by N. Table 11 contains the results of suitable simulations.

The results contained in table 11 confirmed that the increase in the number of bootstrap replications caused a decrease of actual test size for the A_2 model at 5% and 10% significance levels. However, the intensity of this process turned out to be insufficient and the actual size still lay outside confidence intervals in all cases. A similar effect (decrease of actual size) was found for the A_3 model at 5% significance level, but this time the size performance had worsened while N increased. Finally, it should be noted that for the A_3 model the actual size was found to grow with an increase of N at 1% significance level (relatively good performance was found for $N = 200$ and $N = 300$ replications). Summarizing, these outcomes provided no clear evidence of whether Hypothesis 6 is true or false. However, they did provide a strong basis for claiming that Hypothesis 4 is indeed true.

Concluding Remarks

The aim of this paper was to examine the size properties of the Toda-Yamamoto test for Granger causality in the case of a relatively small sample size. The simulation study was conducted for integrated order-1 trivariate VAR models, and a variety of distribution of error vector was also considered during computation. In order to perform suitable research, both the standard asymptotic distribution theory as well as the residual-based bootstrap technique were used.

The results of the conducted simulation study in the case of properly specified lag parameters indicate that the standard asymptotic approach causes significant over-rejection in almost all considered cases. The application of the residual-based bootstrap method improved the size performance of the TY test, however, in the case of structural break and high correlation the actual size was still far away from the nominal one.

The misspecification of the lag parameter caused much worse performance of the TY test when asymptotic theory was applied. In general, the performance of the bootstrap method has not worsened in such a significant way.

The results contained in this paper support the hypothesis that asymptotic distribution theory performs better for longer time series. However, except for the case of spherical multivariate normal distribution of error term, this type of significant improvement has not been observed. Furthermore, test results obtained in cases of high size distortion of the bootstrap-based technique brought no clear suggestion about the relationship between the number of bootstrap replications and the actual size of the test.

The outcomes contained in this article should be useful tips for other researchers using considered variants of the Toda-Yamamoto test in their practical applications. The presented results ensure that bootstrap based on leveraged residuals is often an effective tool for Granger causality testing, which allows avoidance of the problem of over-rejection of the considered null hypothesis. However, the conducted simulation study confirms that this method cannot be used without a second thought, since it is likely to fail for specific models.

References

Chou, P. H., and G. Zhou. 2006. Using bootstrap to test portfolio efficiency. *Annals of Economics and Finance* 2:217–49.

Davidson, R., and J. G. MacKinnon. 1996. The size distortion of bootstrap tests. Working Paper 936, Queen's University.

Davison, A. C., and D. V. Hinkley. 1999. *Bootstrap methods and their application.* Cambridge: Cambridge University Press.

Dolado, J. J., and H. Lütkepohl. 1996. Making Wald tests work for cointegrated var systems. *Econometrics Reviews* 15 (4): 369–86.

Efron, B. 1979. Bootstrap methods: Another look at the Jacknife. *Annals of Statistics* 7 (1): 1–26.

Engle, R. F., and C. W. J. Granger. 1987. Cointegration and error correction: representation, estimation and testing. *Econometrica* 55 (2): 251–81.

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37 (3): 424–38.

———. 1988. Some recent developments in the concept of causality. *Journal of Econometrics* 39 (1–2): 199–211.

Granger, C. W. J., and P. Newbold. 1974. Spurious regression in econometrics. *Journal of Econometrics* 2 (2): 111-20.

Hacker, R. S., and A. Hatemi-J. 2006. Tests for causality between integrated variables using asymptotic and bootstrap distributions: theory and application. *Applied Economics* 38 (13): 1489–500.

Hatemi-J, A. 2003. A new method to choose optimal lag order in stable and unstable var models. *Applied Economics Letters* 10 (3): 135–7.

Horowitz, J. L. 1994. Bootstrap-based critical values for the information matrix test. *Journal of Econometrics* 61 (2): 395–411.

———. 1995. Bootstrap methods in econometrics: Theory and numerical performance. In *Advances in economics and econometrics: Theory and applications; Seventh World Congress,* ed. D. Kreps and K. Wallis, iii:189–222. Cambridge: Cambridge University Press.

Hylleberg, S., R. F. Engle, C. W. J. Granger, and S. Yoo. 1990. Seasonal integration and cointegration. *Journal of Econometrics* 44 (2): 215–38.

Johansen, S. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59 (6): 1551–80.

———. 1996. *Likelihood based inference in cointegrated vector autoregressive models.* 2nd ed. Oxford: Oxford University Press.

Johansen, S., and E. Schaumburg. 1988. Likelihood analysis of seasonal cointegration. *Journal of Econometrics* 88 (2): 301–39.

MacKinnon, J. G. 1992. Model specification tests and artificial regressions. *Journal of Economic Literature* 30 (1): 102–46.

Maddala, G. S. 1992. *Introduction to econometrics.* 2nd ed. New York: Maxwell Macmillan.

Mantalos, P. 2000. A graphical investigation of the size and power of the

Granger-causality tests in integrated-cointegrated va r systems. Studies in *Nonlinear Dynamics & Econometrics* 4 (1), article 2.

Phillips, P. C. B. 1986. Understanding the spurious regression in econometrics. *Journal of Econometrics* 33 (3): 311–40.

Toda, H. Y., and T. Yamamoto. 1995. Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* 66 (1–2): 225–50.