# Application of bootstrap methods in investigation of size of the Granger causality test for integrated VAR systems

Lach, Łukasz

AGH University of Science and Technology, Department of Applications of Mathematics in Economics

2010

**Łukasz Lach***

# Application of bootstrap methods in investigation of size of the Granger causality test for integrated VAR systems

***Abstract***

*This paper examines the size performance of Toda-Yamamoto test for Granger causality in case of trivariate integrated-cointegrated VAR systems and relatively small sample size. 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 increase of sample size and number of bootstrap replications on size performance of Toda-Yamamoto test statistics is also examined.*

*The results of 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 Toda-Yamamoto test. However, in some cases considered bootstrap method also leads to serious size distortion and performs worse than the traditional approach based on $\chi^2$ distribution.*

## 1. Introduction

The causal relationship (in 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 (see [7]) 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 tool for testing the causal effects. These nonstandard asymptotic properties of Wald test were investigated by Granger and Newbold (see [8] for some empirical findings) and Philips ([21] - theoretical framework). As a cure for this problem the idea of Vector Error Correction Model (see [6] and [9]) 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, sensitivity for improper lag establishment) turned out to be a serious difficulty in empirical applications.

Another solution was proposed by Toda and Yamamoto ([22]). This approach ensures that asymptotic distribution theory is valid for VAR systems, regardless 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

* AGH University of Science and Technology, Department of Applications of Mathematics in Economics.

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. The application of bootstrap[1] approach may often provide better results since bootstrapping does not strictly depend on model specification.

The properties of augmented Wald test in both the asymptotic and bootstrap variant were examined by a number of authors in recent years. Dolado and Lütkepohl [4] conducted a simulation exercise to examine the power of considered testing method in case of integrated VAR model[2]. Their outcomes show that in high dimensional VARs with a small true lag length the significant reduction of power of considered causality test may occur, especially for small samples. Mantalos [20] conducted similar studies of size and power properties of eight versions of the Granger causality test[3]. His findings indicate that standard asymptotic approach may often lead to significant size distortion. The application of residual-based bootstrap technique usually improves size and power performance of causality tests. Hacker and Hatemi [10] examined size properties of 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 bootstrap technique performed relatively well in all cases. On the other hand they restricted the research only to models without cointegration.

This paper is the generalization of previous studies concentrated on investigation of size properties of 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 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 of number of bootstrap replications on size performance of 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 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 TY test, specification of VAR models used for simulation purposes and considered bootstrap technique. Section 4 contains results of all conducted simulations. Section 5 concludes the paper.

## 2. Main hypotheses

The main objective of this paper is the investigation of size properties of Toda-Yamamoto test for Granger causality. First important point that distinguishes this study from the existing literature is the use of trivariate VAR model for simulation purposes[4]. Another important point is the fact that this paper examines all possible dimensions of cointegration space. As it

[1] For more details on bootstrap see [5].

[2] In [4] the error term is independently drawn from identical multivariate normal distribution.

[3] In [20] the error term was only $N(0,I_2)$ i.i.d..

[4] Most of previous papers examine two-dimensional models. In three-dimensional case the structure of causal links may be more extended.

was already mentioned former studies concentrated on similar topic provided evidence of poor performance of modified Wald procedure in case of nonstationary variables. Thus, it seems to be reasonable to formulate:

> Hypothesis 1 – 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 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:

> Hypothesis 2 – Residual-based bootstrap method usually improves size performance of TY test.

In practice the proper specification of VAR model is often difficult to obtain. One of the most common problems is the misspecification of lag parameter. Previous studies[5] show that in this case the size performance of TY test (asymptotic variant) may significantly worsen. It may be interesting how bootstrap-based technique performs in this case. Therefore, we should test:

> Hypothesis 3 – Misspecification of lag parameter in VAR model leads to considerable aggravation of size performance of TY only in asymptotic variant.

Despite the fact that bootstrap methods are often a useful tool to overcome problem of size distortion in 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 (also covering cases where standard assumptions do not hold[6]). Therefore, this paper contains the verification of following:

> Hypothesis 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 TY test[7]. 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:

> Hypothesis 5 – When standard assumptions hold, the increase of sample size improves size performance of TY test (asymptotic variant).

In order to apply bootstrap technique researcher must establish number of bootstrap replications. In previous papers this number varied significantly (from dozens to hundreds). It may be interesting to investigate if change of number of bootstrap replication may lead to

---

[5] See [10] and [20].

[6] This paper examines possibilities where some standard assumptions about structure of considered VAR models and TY methodology are unfulfilled.

[7] See [4], [10] and [20].

significant improvement of size performance of TY test in some specific cases (namely, cases of relatively significant size distortion). This problem may be captured in verification of following:

> Hypothesis 6 – There is a relationship between number of bootstrap replication and size performance of TY test in some specific cases.

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

## 3. Methodology and 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:

$$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + \varepsilon_t \qquad (1)$$

where $y_t = (y_t^1, ..., y_t^n)'$, $c = (c_1, ..., c_n)'$ and $\varepsilon_t = (\varepsilon_{1,t}, ..., \varepsilon_{n,t})'$ [8] are $n$-dimensional vectors and $\{A_i\}_{i=1}^{p}$ is a set of $n \times n$ matrices of parameters for appropriate lags. The order $p$ of the process is assumed to be known. Furthermore, we shall assume that error vector is an independent white noise process with nonsingular covariance matrix $\Sigma_\varepsilon$ (which elements are constant over time[9]). We also assume that the condition $E|\varepsilon_{k,t}|^{2+s} < \infty$ holds true for all $k=1,\dots,n$ and some $s>0$. The Toda-Yamamoto (see [22]) idea of testing for causal effects is based on estimating the augmented VAR($p+d$) model (circumflex indicates OLS estimator of specific parameter):

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

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

$$\mathrm{H}_0: \ a_{jk}^{s} = 0 \quad (3)$$

[8] In this paper transpose of matrix $M$ is denoted by $M'$.

[9] In this paper cases where these standard assumptions do not hold are also investigated.

for $s$=1,…,$p$

where $A_s = \left[a_{pq}^s\right]_{p,q=1,\dots,n}$ for $s$=1,…,$p$. According to Toda and Yamamoto [22] 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 following compact notation ($T$ denotes the considered sample size):

**Table 1**: Compact notation used to formulate TY test statistics:

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

The initial point of considered procedure is the calculation of $S_U := \frac{\hat{\delta}\hat{\delta}'}{T}$ — the variance-covariance matrix of residuals from unrestricted augmented model (i.e. model (2)). Then we can define $\beta := vec(c,A_1,\dots,A_p,0_{n\times nd})$ and $\hat{\beta} := vec(\hat{c},\hat{A}_1,\dots,\hat{A}_p,\dots,\hat{A}_{p+d})$ where $vec(\cdot)$ denotes column stacking operator and $0_{n\times nd}$ stands for $n\times nd$ 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:

$$\mathrm{TY} := \left(C\hat{\beta}\right)'\left(C\left((ZZ')^{-1}\otimes S_U\right)C'\right)^{-1}\left(C\hat{\beta}\right) \quad (4)$$

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 which elements take only the value of zero or one. Each of $p$ rows of matrix $C$ corresponds to restriction of one parameter in $\beta$. The value of every element in each row of $C$ is one if the associated parameter in $\beta$ is zero under the null hypothesis and it is zero otherwise. There is no association between matrix $C$ and last $n^2d$ elements in $\beta$. This approach allows us to write the null hypothesis of non-Granger causality in the following form:

$$\mathrm{H}_0 \colon C\beta' = 0 . \quad (5)$$

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

In order to examine the size properties of the TY test some I(1) models are considered. Causality tests are conducted in 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:

$$\Delta y_t = c + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \quad (6)$$

where $\Pi = -I + \sum_{i=1}^{p} A_i$ 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[10]:

- The roots of the characteristic polynomial:

$$\det(I_n - A_1 z - A_2 z^2 - \ldots - A_p z^p) \quad (7)$$

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

If the first assumption holds then the considered process is neither explosive (roots in the unit circle) or seasonally cointegrated (roots on the boundary of the unit circle different from $z=1$, for more details on this issue see Hylleberg, Engle, Granger, and Yoo [14] or Johansen and Schaumburg [15]). The second assumption ensures that there are at least $p$-$r$ 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 $p$-$r$.

[10] These assumptions are sufficient to prove so-called Johansen-Granger representation theorem (for more details see [16] and [17]).

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

$$y_t = c + Ay_{t-1} + \varepsilon_t \quad (8)$$

where $c = (0{,}01 \quad 0{,}01 \quad 0{,}01)'$ in all cases and matrix $A$ provides specific cointegration properties (see previously presented assumptions). For details about matrices used in simulation study explore the following table:

**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 | $A_3$ |

Directly form 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[11]). Beside various schemes of algebraic structure some specific distributions of error vectors are also examined. At this place it should be noted that in previous studies concentrated on similar topics the error term was usually $N(0_{n\times 1}, \sigma^2 I_n)$ distributed ($n$ stands for considered dimension) for some positive $\sigma$ (see Hacker and Hatemi [10], Dolado and Lütkepohl[12] [4] or Mantalos [20]). In this paper the size properties of TY test are examined for variety of types of time structure of error term[13]. Some fundamental information is contained in the following table:

---

[11] In three-dimensional VAR model the relationship between $y^3$ and $y^1$ as well as between $y^3$ and $y^2$ may have indirect impact on links between $y^2$ and $y^1$.

[12] In [4] authors also consider case of nonzero covariance between components of error term.

[13] In some considered specifications the standard assumptions for TY method do not hold.

**Table 3**: Models used to generate distribution of error term[14]:

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

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 [12] or Chou and Zhou [1]). 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 bootstrap approach may significantly improve investigated size properties. Fourth examined possibility ($E_4$) is related to the idea of 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). Similar type of time dependence structure in error term series was examined by Hacker and Hatemi (see [10], 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 $N(0,1)$

[14] Random draw for error term is always based on i.i.d. variables (normal, discrete uniform).

distribution) are used as the initial observations for VAR models. To make the results of 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 $E_2=(E_{2,t})_{t=1,...,T}$ series the following transformation of $E_1=(E_{1,t})_{t=1,...,T}$ series is applied:

$$E_{2,t}=ZE_{1,t} \quad (9)$$

where $t=1,\ldots,T$ and $ZZ'=\Sigma$ (Cholesky decomposition). The values of $E_1$ series are also used in process of generation of $E_4$ series and $E_3$ series (for first $\frac{T}{2}$ observations). In order to generate $E_5$ series initial observations are once again drawn from $N(0,1)$ distribution and $\begin{pmatrix} w_{1,t} & w_{2,t} & w_{3,t} \end{pmatrix}'=E_{1,t}$ for $t=1,\ldots,T$.

To examine the size properties of 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 is 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 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[15]. First considered augmented VAR model (2) is estimated through OLS methodology with the null hypothesis assumed (that is: $y^2$ does not Granger cause $y^1$). 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 probability measure equal to $\frac{1}{T}$) from the set $\{\hat{\varepsilon}_i^m\}_{i=1,...,T}$ (as a result we get the set $\{\hat{\varepsilon}_i^{**}\}_{i=1,...,T}$);
- Subtract the mean to guarantee the mean of bootstrap residuals is zero (this way we create the set $\{\hat{\varepsilon}_i^*\}_{i=1,...,T}$ , such that $\hat{\varepsilon}_{k,i}^*=\hat{\varepsilon}_{k,i}^{**}-\frac{\sum_{j=1}^{T}\hat{\varepsilon}_{k,j}^{**}}{T}$, $i=1,..,T$, $k=1,2,3$);

[15] For more details on this issue see Davison and Hinkley [3] and Hacker and Hatemi [10].

- Generate the simulated data $\{y_i^*\}_{i=1,...,T}$ through the use of original data ($\{y_i\}_{i=1,...,T}$), coefficient estimates from the regression ($\hat{c},\{\hat{A}_i\}_{i=1,...,p+d}$) and the bootstrap residuals $\{\hat{\varepsilon}_i^*\}_{i=1,...,T}$);
- 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 get empirical critical values (bootstrap critical values) next. The suitable procedure (which allows to conduct every type of simulation presented in this article) written in Gretl is available from the author upon request.

## 4. Empirical results

In this section results of conducted causality tests are presented. The following tables contain the rejection rates obtained while testing the null hypothesis in TY test with the application of both standard asymptotic distribution theory and residual-based bootstrap approach. In recent years the problem of establishment of adequate significance levels for diagnostic applications has been intensively discussed. Some researchers recommended relatively large levels (Maddala [19]) while others argued that typical values are the best choice (MacKinnon [18]). As it was already mentioned in this article typical significance levels are considered. Thus the results of presented simulations are more comparable with the similar research conducted by Hacker and Hatemi [10] and Mantalos [20]. 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:

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

where $Ts$ denotes considered nominal size (1%, 5%, 10%) and $N_r$=1000 stands for number of repetitions[16]. This is how the intervals [0,4%;1,6%], [3,6%;6,4%], [8,1%;11,9%] were established for 1%, 5% and 10% significance levels respectively. This approach leads to the criteria of bad performance, namely, actual test size is significantly distorted whenever it lies outside the suitable confidence interval. In the following tables these findings are indicated by shaded areas. 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:

[16] $N_r$=*1000* was also used in [4], [10] and [20]. Considered type of confidence intervals was used in [4] and [20].

**Table 4**: Size of TY test for Granger causality – no-cointegration case:

| Algebraic structure | Distribution of error term | Lag $p$ | $\chi^2$ 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% |

**Table 5**: Size of TY test for Granger causality – case of two cointegrating vectors:

| Algebraic structure | Distribution of error term | Lag $p$ | $\chi^2$ 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% |

**Table 6**: Size of TY test for Granger causality – case of one cointegrating vector:

| Algebraic structure | Distribution of error term | Lag $p$ | $\chi^2$ 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% |

After analyzing results contained in table 4 one can easily see that asymptotic distribution theory was found to cause serious size distortions in almost all cases. The largest distortions were indicated in case of structural change in error term distribution ($E_3$). Furthermore, it should be noted that whenever critical values were taken from suitable $\chi^2$ distribution the over-rejection was indicated, which seems to prove that Hypothesis 1 is true. The application of bootstrap method improved the size properties of TY test for all significance levels in case of $E_1$, $E_4$ and $E_5$ distribution. These results provided strong basis for claiming that Hypothesis 2 is also true. Although the significant over-rejection was still found for $E_3$ error distribution (except 10% level), size distortions were much smaller than in non-bootstrap approach. However, one must note that bootstrap test was found to under-reject the null hypothesis in case of $E_2$ distribution, which led to significant size distortions for 5% and 10% significance levels (even worse performance than for $\chi^2$ distribution). The outcomes obtained by Hacker and Hatemi [10] in corresponding research conducted for similar two-dimensional cases ($A_1$ model, $E_1$ and $E_5$ error term) are in line with 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 TY test based on asymptotic distribution theory tends to over-reject the null hypothesis also when there exist cointegration between considered variables[17]. Secondly, they provided basis for claiming that the application of bootstrap methods leads to reduction of actual test size in comparison to asymptotic method. However, this reduction is still insufficient for $A_2$ algebraic structure and $E_3$ error distribution scheme (still over-rejection) and too intensive for $A_3$ and $E_2$ case (under-rejection, worse performance in comparison to $\chi^2$ distribution on 5% and 10% significance levels).

In practice it is often difficult to establish the lag parameter properly before estimating VAR model. Despite the variety of econometric methods (AIC, BIC, FPE information criteria, more recent Hatemi's [11] criterion) many researchers are still struggling to decide what value of lag length chose for further analysis. In the context of our investigation this problem was examined by the repetition of all causality tests in case of 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:

**Table 7**: Size of TY test for Granger causality – no-cointegration case, misspecified parameter $p$

| Algebraic structure | Distribution of error term | Lag $p$ | $\chi^2$ 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% |

**Table 8**: Size of TY test for Granger causality – case of two cointegrating vectors, misspecified parameter $p$

| Algebraic structure | Distribution of error term | Lag $p$ | $\chi^2$ 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% |

[17] In [4] and [20] cointegration rank is no greater than one.

**Table 9**: Size of TY test for Granger causality – case of one cointegrating vector, misspecified parameter $p$

| Algebraic structure | Distribution of error term | Lag $p$ | $\chi^2$ 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% |

It seems to be obvious that 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 results contained in table 7 (no-cointegration case) one can easily see that standard approach (based on $\chi^2$ distribution) causes even stronger over-rejection (higher rejection rates and more shaded areas) than in corresponding case (table 4). On the other hand the results obtained with application of bootstrap method belong to suitable confidence intervals in all except for one case (in comparison to corresponding case). For model with two cointegrating vectors ($A_2$) the actual test size (case of $\chi^2$ distribution) is too high in all except for 3 cases. This means that misspecification of parameter $p$ considerably worsens size performance of TY test. Furthermore, actual size of bootstrap test was found to lie outside confidence interval for exactly the same combination of considered significance levels and error term schemes like in corresponding case (table 5). The standard asymptotic approach was also found to cause serious over-rejection for $A_3$ structure in almost all cases. On the other hand actual test size based on bootstrap method was distorted only for $E_2$ (under-rejection) and $E_3$ (over-rejection) case. In general, size performance of TY test worsened significantly only for 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 $E_2$ and $E_3$ case).

Additionally, to examine the size performance of TY test in both considered variants causality tests were conducted for longer sample. One should expect standard asymptotic approach to perform relatively better in this case. Suitable tests were conducted for sample size $T$=100 and no-cointegration model with parameter $p$=1 and $p$=2 [18]. For comparability with previous results (obtained for $T$=40) 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 similar investigation conducted for small sample ($T$=40):

[18] In [10] considered sample size is also equal to $T$=40 (small sample) and $T$=100 (medium sample).

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

| Algebraic Structure | Distribution of error term | Lag $p$ | $\chi^2$ distribution | | | Bootstrap distribution | | |
|---|---|---|---|---|---|---|---|---|
| | | | 1% | 5% | 10% | 1% | 5% | 10% |
| $A_1$ | $E_1$ | 1 | 1,1% (1,7%) | 6,2% (6,1%) | 12% (13,2%) | 0,9% (0,8%) | 4,2% (4,6%) | 9,5% (10,9%) |
| | $E_1$ | 2 | 1,3% (2,1%) | 5,6% (10,6%) | 13,5% (16%) | 1,1% (0,9%) | 4,9% (4,5%) | 10,3% (10,7%) |

The analysis of above table confirmed the hypothesis that size properties of TY test for Granger causality are improving with the increase of sample size. Although for 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 in suitable confidence intervals in all cases. On the other hand it should be noted that for other 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 number of replications. In previous research concentrated on similar investigation this value varied significantly. Horovitz [13] used 100 replications, Mantalos [20] — 200, Hacker and Hatemi [10] — 800 while Davidson and MacKinnon [2] used 1000 replications to create bootstrap distribution each time. Increase of number of replications may often have important impact on improvement of performance of TY test size. However in some situations bootstrap methods are likely to fail, regardless the number of replications used (see Horovitz [12]). This paper takes part in the discussion of mentioned problem as it contains results of some simulations based on different number 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 the comparability with 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 TY test conducted with application of $\chi^2$ 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 results of suitable simulations:

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

| Algebraic structure | Distribution of error term | Lag $p$ | $\chi^2$ 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 |

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

## 5. Concluding remarks

The aim of this paper was to examine the size properties of Toda-Yamamoto test for Granger causality in case of relatively small sample size. The simulation study was conducted for integrated order-1 trivariate VAR models, 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 conducted simulation study in case of properly specified lag parameters indicate that standard asymptotic approach causes significant over-rejection in almost all considered cases. The application of residual-based bootstrap method improved the size performance of TY test, however, in the case of structural break and high correlation the actual size was still far away from nominal one.

The misspecification of lag parameter caused much worse performance of TY test when asymptotic theory was applied. In general the performance of the bootstrap method has not worsened in such 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 bootstrap-based technique brought no clear suggestion about the relationship between number of bootstrap replications and actual size of test.

The outcomes contained in this article should be useful tips for other researchers using considered variants of 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 avoiding the problem of over-rejection of the considered null hypothesis. However, conducted simulation study confirms that this method cannot be used without a second thought since it is likely to fail for specific models.

**References:**


[1] Chou P.H. and Zhou G. (2006) Using Bootstrap to Test Portfolio Efficiency. *Annals of economics and finance*, **2**, 217–249.

[2] Davidson R. and MacKinnon J.G (1996) The size distortion of bootstrap tests. *Working paper, Department of Economics, Queen's University, Canada*.

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

[4] Dolado J.J. and Lütkepohl H. (1996) Making Wald tests work for cointegrated VAR systems. *Econometrics Reviews,* **15**, 369–386.

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

[6] Engle R.F. and Granger C.W.J. (1987) Cointegration and error correction: representation, estimation and testing. *Econometrica,* **55**, 251–281.

[7] Granger C.W.J. (1969) Investigating causal relations by econometric models and cross–spectral methods. *Econometrica*, **37**, 424–438.

[8] Granger C.W.J. and Newbold P. (1974), Spurious regression in econometrics. *Journal of Econometrics*, **2**, 111–120.

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

[10] 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**, 1489–1500.

[11] Hatemi-J A. (2003) A new method to choose optimal lag order in stable and unstable VAR models. *Applied Economics Letters*, **103**, 135–137.

[12] Horowitz J.L. (1995) *Advances in Economics and Econometrics: Theory and Applications, Chapter 7 : Bootstrap methods in econometrics: Theory and numerical performance*. Cambridge University Press.

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

[14] Hylleberg S., Engle R.F., Granger C.W.J. and Yoo S. (1990) Seasonal Integration and Cointegration. *Journal of Econometrics*, **44**, 215–238.

[15] Johansen S. and Schaumburg E. (1988) Likelihood Analysis of Seasonal Cointegration. *Journal of Econometrics*, **88**, 301–339.

[16] Johansen S. (1991) Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, **59**, 1551–1580.

[17] Johansen S. (1996) *Likelihood Based Inference in Cointegrated Vector Autoregressive Models 2nd edn.*. Oxford University Press.

[18] MacKinnon J.G. (1992) Model specification tests and artificial regressions. *Journal of Economic Literature*, **30**, 102–146.

[19] Maddala G.S. (1992) *Introduction to Econometrics*, *2nd ed*. New York: Maxwell Macmillan.

[20] Mantalos P. (2000) A Graphical Investigation of the Size and Power of the Granger-Causality Tests in Integrated-Cointegrated VAR Systems. *Studies in Nonlinear Dynamics & Econometrics*, Volume 4, Issue 1, 2000, Article no. 2
[21] Phillips C.P. (1986) Understanding the spurious regression in econometrics. *Journal of Econometrics*, **33**, 311–340.
[22] Toda H.Y. and Yamamoto T. (1995) Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, **66**, 225–250.