Central European Journal of Economic Modelling and Econometrics

The Cointegrated VAR Model with Deterministic Structural Breaks

Emilia Gosińska* and Aleksander Welfe†

Submitted: 28.03.2022, Accepted: 8.05.2022

Abstract

The presence of a binary variable in the cointegrated VAR (CVAR) model is most often interpreted as the structural break affecting the data generating process. It is proved in the paper that to enjoy this interpretation the binary variable must appear simultaneously inside and outside the cointegration space. In order to test for the break we advocate to employ the Wald statistic, however, its critical values and the power had to be simulated separately for the possible change of the constant, the trend, and both. The experiments were designed for different sizes of the cointegrating space, number of variables, the span of the break, normally and t-distributed errors. It is shown that the power of the test depends mostly on the magnitude of the break and the sample size while other factors are of secondary importance. In order to test for the break at unknown period the supWald statistic was proposed.

Keywords: structural breaks, cointegrated VAR, WALD test, hypothesis testing

JEL Classification: C1, C12, C32

*Chair of Econometric Models and Forecasts, Faculty of Economics and Sociology, University of Łódź, Poland; e-mail: emilia.gosinska@uni.lodz.pl; ORCID: 0000-0002-5325-6144

†Chair of Econometric Models and Forecasts, Faculty of Economics and Sociology, University of Łódź, Poland; e-mail: aleksander.welfe@uni.lodz.pl; ORCID: 0000-0002-9925-3432

1 Introduction

Structural breaks are a relatively common problem, especially in empirical studies on transforming economies utilising the time series, which leads to the non-normal distribution of residuals and greatly hinders statistical inference (testing). It is usually solved by the introduction of binary variables (see Juselius, 2006). However, for a binary variable to represent a real structural break in the data generating process (DGP), it must be present inside and outside of the cointegration space in the cointegrated VAR (CVAR) model at the same time (see Section 2).
In testing for the presence of structural break in a known period, the Wald statistic is worth considering (see Sobreira and Nunes 2012, Nielsen 2004). A structural break in the deterministic component of the DGP within the system of nonstationarity variables causes that the distribution of the Wald statistic is nonstandard (see Section 3). Then, the critical values need to be simulated through the Monte Carlo experiments (see Section 4). If the structural break is not known, which is common in empirical applications, the appropriate testing procedure must also be applied (see Section 3).
Summing up, the paper focuses on determining the appropriate specification of a CVAR model and then testing it for the presence of various structural breaks (a level break, a trend break, and both) in the deterministic components. The power and size of the Wald test and the influence of specific conditions (e.g. the break point values, error distribution) are simulated and then analyzed (see Section 4).

2 CVAR in the presence of the structural breaks in the deterministic part of the DGP

The implications of the presence of the deterministic terms in a VAR process with the unit root can be explored by defining the DGP as follows:

$$\mathbf{v}_t = \mathbf{y}_t + \mathbf{H}\mathbf{d}_t, \tag{1}$$

where $\mathbf{y}_t$ denotes $M \times 1$ vector of M stochastic variables integrated of order one, $\mathbf{d}_t = \begin{bmatrix} 1 & \vdots & t \end{bmatrix}^T$, $\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \vdots & \mathbf{h}_2 \end{bmatrix}$ - $M \times 2$ matrix of parameters.
Equation (1) defines the DGP as the explicit sum of the zero mean stochastic part and the deterministic component that can contain any deterministic or binary variables (e.g. a constant, a trend, specific variables accounting for structural breaks). Therefore, the mean of the $\mathbf{y}_t$ variables is directly specified by the deterministic term and is independent of the stochastic component parameters. This is the main advantage of setting up the process in the above form (see Lütkepohl 2005, p. 256–258).
The second component of (1) decomposes into (see Johansen and Nielsen 2018):

$$\mathbf{H}\mathbf{d}_t = \mathbf{H}_0\mathbf{d}_t^* + \mathbf{H}_1\Delta\mathbf{d}_t^*, \tag{2}$$

where $\mathbf{d}_t^* = t$, $\Delta\mathbf{d}_t^* = 1$, $\mathbf{H}_0 = \mathbf{h}_2$, $\mathbf{H}_1 = \mathbf{h}_1$.
Let us assume that a single structural break affects both deterministic components of the DGP, that is constant and trend:

$$\mathbf{v}_t = \mathbf{y}_t + \mathbf{h}_1 + \mathbf{h}_2 t + \mathbf{h}_3 u_{1t} + \mathbf{h}_4 u_{2t}, \tag{3}$$

where

$$u_{1t} = \begin{cases} 1 & \text{for } t \geq t_0 \\ 0 & \text{for } t < t_0 \end{cases}, \quad u_{2t} = \begin{cases} t-(t_0-1) & \text{for } t \geq t_0 \\ 0 & \text{for } t < t_0 \end{cases}$$

and consequently in (2):

$$\mathbf{d}_t^* = \begin{bmatrix} t \\ u_{2,t} \end{bmatrix}, \quad \Delta\mathbf{d}_t^* = \begin{bmatrix} 1 \\ u_{1,t} \end{bmatrix}, \quad \mathbf{H}_0 = \begin{bmatrix} \mathbf{h}_2 & \mathbf{h}_4 \end{bmatrix}, \quad \mathbf{H}_1 = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_3 \end{bmatrix}.$$

Multiplying both sides of (3) by lag polynomial $\mathbf{\Pi}(L)$ gives

$$\mathbf{\Pi}(\mathbf{L})\mathbf{v}_t = \mathbf{\Pi}(\mathbf{L})\mathbf{y}_t + \mathbf{\Pi}(\mathbf{L})\mathbf{h}_1 + \mathbf{\Pi}(\mathbf{L})\mathbf{h}_2 t + \mathbf{\Pi}(\mathbf{L})\mathbf{h}_3 u_{1t} + \mathbf{\Pi}(\mathbf{L})\mathbf{h}_4 u_{2t}, \tag{4}$$

where $\mathbf{\Pi}(L) = \mathbf{I} - L\mathbf{\Pi}_1 - L^2\mathbf{\Pi}_2 - \ldots - L^S\mathbf{\Pi}_S$. Matrices $\mathbf{\Pi}_1$,$\mathbf{\Pi}_2$,..., $\mathbf{\Pi}_S$ include the multipliers of the VAR representation $\mathbf{\Pi}(L)\mathbf{y}_t = \xi_t$, S defines the lag length of the VAR process, and ξ_t is $M \times 1$ vector of white-noise error terms.
It can be easily shown that the components of the right hand side of the above equation are equal, respectively, to (for the details see Gosińska 2015)

$$\mathbf{\Pi}(L)\mathbf{h}_1 = -\mathbf{\Pi}\mathbf{h}_1, \tag{5a}$$

$$\mathbf{\Pi}(L)\mathbf{h}_3 u_{1t} = -\mathbf{\Pi}\mathbf{h}_3 u_{1,t-1} + \mathbf{h}_3 \Delta u_{1,t} + \sum_{s=1}^{S-1} (-\mathbf{\Gamma}_s)\mathbf{h}_3 \Delta u_{1,t-s}, \tag{5b}$$

$$\mathbf{\Pi}(L)\mathbf{h}_2 t = -\mathbf{\Pi}\mathbf{h}_2(t-1) + \mathbf{\Psi}\mathbf{h}_2, \tag{5c}$$

$$\mathbf{\Pi}(L)\mathbf{h}_4 u_{2t} = -\mathbf{\Pi}\mathbf{h}_4 u_{2,t-1} + \mathbf{\Psi}\mathbf{h}_4 u_{1,t} + \sum_{s=0}^{S-2} \left(\sum_{j=s+1}^{S-1} \mathbf{\Gamma}_j \right) \mathbf{h}_4 \Delta u_{1,t-s}, \tag{5d}$$

where $\mathbf{\Pi} = \sum_{s=1}^{S} \mathbf{\Pi}_s - \mathbf{I}$, $\mathbf{\Gamma}_i = -\sum_{s=i+1}^{S} \mathbf{\Pi}_s$, $\mathbf{\Psi} = \mathbf{I} + \sum_{s=1}^{S-1} s\mathbf{\Pi}_{s+1}$, $\mathbf{\Pi}_s$ are the matrices of the lag polynomial $\mathbf{\Pi}(L)$.
Provided matrix $\mathbf{\Pi} = \mathbf{A}\mathbf{B}^T$ is of reduced rank, the CVAR representation with a constant, a trend, and breaks in both can be written as (see different parametrisation in Trenkler et al. 2006)

$$\begin{aligned} \mathbf{\Delta v}_t = \mathbf{A}\left[\mathbf{B}^T\mathbf{v}_{t-1} + \mathbf{g}_1 + \mathbf{g}_2(t-1) + \mathbf{g}_3 u_{1,t-1} + \mathbf{g}_4 u_{2,t-1}\right] + \sum_{s=1}^{S-1} \mathbf{\Gamma}_s \mathbf{\Delta v}_{t-s} + \mathbf{f}_2 + \\ + \sum_{s=0}^{S-1} \mathbf{f}_{3,s}\Delta u_{1,t-s} + \mathbf{\Psi}\mathbf{h}_4 u_{1,t} + \sum_{s=0}^{S-2} \mathbf{f}_{4,s}\Delta u_{1,t-s} + \xi_t \end{aligned} \tag{6}$$

for $t = S+1, S+2, \ldots$, where $M \times R$ matrices $\mathbf{B}$ and $\mathbf{A}$ have standard interpretation of cointegrating vectors and weights, R is the size of the cointegration space, and

$$\begin{aligned}
\mathbf{g}_i &= -\mathbf{B}^T\mathbf{h}_i, \quad i = 1,2,3,4,\\
\mathbf{f}_2 &= \mathbf{\Psi}\mathbf{h}_2,\\
\mathbf{f}_{3,s} &= \begin{cases} \mathbf{h}_3 & \text{for } s = 0\\ -\mathbf{\Gamma}_s\mathbf{h}_3 & \text{for } s = 1,2,\ldots,S-1 \end{cases},\\
\mathbf{f}_{4,s} &= \sum_{j=s+1}^{S-1}\mathbf{\Gamma}_j\mathbf{h}_4,
\end{aligned}$$

or in a more compact form as

$$\mathbf{\Delta v}_t = \mathbf{A}\mathbf{B}^{*T}\mathbf{v}^*_{t-1} + \sum_{s=1}^{S-1}\mathbf{\Gamma}_s\mathbf{\Delta v}_{t-s} + \mathbf{F}^1\mathbf{\Delta d}^*_t + \sum_{s=0}^{S-1}\mathbf{F}^2_s\mathbf{\Delta}^2\mathbf{d}^*_{t-s} + \xi_t, \tag{7}$$

where

$$\begin{aligned}
\mathbf{v}^*_{t-1} &= \begin{bmatrix}\mathbf{v}_{t-1}\\ \mathbf{d}^*_{t-1}\end{bmatrix},\\
\mathbf{B}^{*T} &= \begin{bmatrix}\mathbf{B}^T & \mathbf{g}_2 & \mathbf{g}_4\end{bmatrix},\\
\mathbf{F}^1 &= \begin{bmatrix}\mathbf{f}^1_2 & \mathbf{f}^1_4\end{bmatrix},\\
\mathbf{F}^2_s &= \begin{bmatrix}\mathbf{0} & \mathbf{f}^2_{4,s}\end{bmatrix},\\
\mathbf{f}^1_2 &= \mathbf{\Psi}\mathbf{h}_2 + \mathbf{A}\mathbf{g}_1,\\
\mathbf{f}^1_4 &= \mathbf{\Psi}\mathbf{h}_4 + \mathbf{A}\mathbf{g}_3,\\
\Delta^2\mathbf{d}^*_{t-s} &= \begin{bmatrix}0\\ \Delta u_{1,t-s}\end{bmatrix},\\
\mathbf{f}^2_{4,s} &= \begin{cases} \mathbf{f}_{3,s} - \mathbf{A}\mathbf{g}_3 + \mathbf{f}_{4,s} & \text{for } s = 0\\ \mathbf{f}_{3,s} + \mathbf{f}_{4,s} & \text{for } s = 1,2,\ldots,S-2\\ \mathbf{f}_{3,s} & \text{for } s = S-1 \end{cases}.
\end{aligned}$$

The maximum likelihood estimation of (7) can be conducted by reduced rank regression of $\mathbf{\Delta v}_t$ on $\begin{bmatrix}\mathbf{v}_{t-1}\\ \mathbf{d}^*_{t-1}\end{bmatrix}$ corrected for

$$\begin{bmatrix}\mathbf{\Delta v}_{t-1} & \ldots & \mathbf{\Delta v}_{t-S+1} & \Delta\mathbf{d}^*_t & \Delta^2\mathbf{d}^*_t & \Delta^2\mathbf{d}^*_{t-1} & \ldots & \Delta^2\mathbf{d}^*_{t-S+1}\end{bmatrix},$$

assuming that the zero regressors have been removed and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{g}_2$, $\mathbf{g}_4$, $\mathbf{F}^1$, $\mathbf{F}^2_s$ consist of freely-varying parameters. Although Johansen (1995) method refers to $\begin{bmatrix}\mathbf{v}_{t-1}\\ \mathbf{d}^*_t\end{bmatrix}$, the asymptotic properties of the parameter estimators are the same in both cases, which can be shown by considering two different representations of the CVAR, assuming DGP equal to (3) and $S = 1$:

$$\boldsymbol{\Delta}\mathbf{v}_t = \mathbf{A}(\mathbf{B}^T\mathbf{v}_{t-1} - \mathbf{B}^T\mathbf{H}_0\mathbf{d}_t^*) + ((\mathbf{A}\mathbf{B}^T + \mathbf{I})\mathbf{H}_0 - \mathbf{A}\mathbf{B}^T\mathbf{H}_1)\boldsymbol{\Delta}\mathbf{d}_t^* + \\ + (\mathbf{A}\mathbf{B}^T + \mathbf{I})\mathbf{H}_1\boldsymbol{\Delta}^2\mathbf{d}_t^* + \xi_t \quad (8)$$

and

$$\boldsymbol{\Delta}\mathbf{v}_t = \mathbf{A}(\mathbf{B}^T\mathbf{v}_{t-1} - \mathbf{B}^T\mathbf{H}_0\mathbf{d}_{t-1}^*) + (\mathbf{H}_0 - \mathbf{A}\mathbf{B}^T\mathbf{H}_1)\Delta\mathbf{d}_t^* + \\ + (\mathbf{A}\mathbf{B}^T + \mathbf{I})\mathbf{H}_1\Delta^2\mathbf{d}_t^* + \xi_t. \quad (9)$$

The second representation is an equivalent to (7) for $S = 1$. The estimation procedure for (8) and (9) is based on two extended models with different sets of parameters:

$$\boldsymbol{\Delta}\mathbf{v}_t = \mathbf{A}(\mathbf{B}^T\mathbf{v}_{t-1} + \mathbf{G}\mathbf{d}_t^*) + \mathbf{F}_1^+\Delta\mathbf{d}_t^* + \mathbf{F}_2\Delta^2\mathbf{d}_t^* + \xi_t \quad (10)$$
$$\boldsymbol{\Delta}\mathbf{v}_t = \mathbf{A}(\mathbf{B}^T\mathbf{v}_{t-1} + \mathbf{G}\mathbf{d}_{t-1}^*) + \mathbf{F}_1\Delta\mathbf{d}_t^* + \mathbf{F}_2\Delta^2\mathbf{d}_t^* + \xi_t, \quad (11)$$

where $\mathbf{G} = -\mathbf{B}^T\mathbf{H}_0$, $\mathbf{F}_1^+ = (\mathbf{A}\mathbf{B}^T + \mathbf{I})\mathbf{H}_0 - \mathbf{A}\mathbf{B}^T\mathbf{H}_1$, $\mathbf{F}_2 = (\mathbf{A}\mathbf{B}^T + \mathbf{I})\mathbf{H}_1$, $\mathbf{F}_1 = \mathbf{H}_0 - \mathbf{A}\mathbf{B}^T\mathbf{H}_1$, because of

$$\mathbf{F}_1^+ + \mathbf{A}\mathbf{G} = \mathbf{F}_1, \quad (12)$$

(11) is the reparameterization of (10). Hence, (10) and (11) are the same statistical models. An analogous relation between $\mathbf{F}_1^+$ and $\mathbf{F}_1$ occurs in the case of a CVAR with a structural break in the level only (then $\mathbf{H}\mathbf{d}_t = \mathbf{h}_1 + \mathbf{h}_3 u_{1t}$).
Summing up, Equation (8) is the reparameterization of (9) and consequently of (7) for a more general case. Accordingly, the maximum likelihood estimation procedure and the asymptotic properties of the estimators correspond to those considered in Johansen (1995) and Johansen and Nielsen (2018).
It is notable that the regressors in $\mathbf{d}_t^*$ restricted to the cointegration space are one-period lagged in (9) but not in (8). Representations (9) and (7) lead to the following conclusions. Firstly, the presence of a binary variable in the cointegration space in period t amounts to assuming the occurrence of a structural break in the data generating process in period $t - 1$. Secondly, if a binary variable (or a trend break) appears in the cointegration space to account for a structural change in the DGP, the appropriate deterministic variables must also appear outside of the cointegration space. Thirdly, it is inappropriate to interpret the presence of a separate binary variable outside of the cointegration space as resulting from a structural change unless it modifies the constants of the cointegration vectors.

3 Testing for structural break in the deterministic part of DGP

Testing for the presence of a structural break is considered in two dimensions. In the first one, it depends on the type of a structural break (a break in a constant, a trend, and both). In the second one, whether the timing of the break is known or unknown is important.

3.1 The break occurs in the known period $\mathbf{t_0}$

Let us initially consider the general case of a break. Under the null hypothesis, all parameters representing a structural break in the cointegration space (in vectors $\mathbf{g}_i$, where $i = 3, 4$, and $i = 3$ for the level break, and $i = 4$ for the trend break) and short-term parameters associated with binary variables (u_{jt}, $j = 1, 2$) are equal to zero. Vector $\mathbf{h}_i$ decomposition in the stationary and nonstationary directions yields the following hypotheses:

$$\begin{aligned} &H_0\colon \mathbf{B}^T\mathbf{h}_i = 0 \wedge \mathbf{B}_\perp^T\mathbf{h}_i = 0, \\ &H_1\colon \mathbf{B}^T\mathbf{h}_i \neq 0 \vee \mathbf{B}_\perp^T\mathbf{h}_i \neq 0, \end{aligned} \tag{13}$$

where $\mathbf{B}_\perp$is an $M \times (M - R)$ full column rank matrix satisfying $\mathbf{B}^T\mathbf{B}_\perp = \mathbf{0}$.
Note that $\widehat{\mathbf{B}}_\perp(\widehat{\mathbf{B}}_\perp^T\widehat{\mathbf{B}}_\perp)^{-1}\widehat{\mathbf{B}}_\perp^T + \widehat{\mathbf{B}}(\widehat{\mathbf{B}}^T\widehat{\mathbf{B}})^{-1}\widehat{\mathbf{B}}^T = \mathbf{I}$, thus estimators $\widehat{\mathbf{B}}^T\widehat{\mathbf{h}}_i$ and $\widehat{\mathbf{B}}_\perp^T\widehat{\mathbf{h}}_i$ are asymptotically independent (see Sobreiraa and Nunes, 2012), therefore:

$$WALD = WALD_B + WALD_{B_\perp} \tag{14}$$

where WALD_B and $\text{WALD}_{B_\perp}$ are the Wald statistics for testing $\mathbf{B}^T\mathbf{h}_i = 0$ and $\mathbf{B}_\perp^T\mathbf{h}_i = 0$, respectively and are equal to

$$\begin{aligned} \text{WALD}_B &= (\mathbf{\Theta}_B\widehat{\mathbf{h}})^T[\mathbf{\Theta}_B\widehat{D}^2(\widehat{\mathbf{h}})\mathbf{\Theta}_B^T]^{-1}(\mathbf{\Theta}_B\widehat{\mathbf{h}}), \\ \text{WALD}_{B_\perp} &= (\mathbf{\Theta}_{B_\perp}\widehat{\mathbf{h}})^T[\mathbf{\Theta}_{B_\perp}\widehat{D}^2(\widehat{\mathbf{h}})\mathbf{\Theta}_{B_\perp}^T]^{-1}(\mathbf{\Theta}_{B_\perp}\widehat{\mathbf{h}}), \end{aligned}$$

where $\mathbf{\Theta}_B$ and $\mathbf{\Theta}_{B_\perp}$ are appropriate for the particular type of the structural break restriction matrices (see Appendix).
In order to determine the value of the above Wald statistic the parameters of interest in the direction of $\mathbf{B}$ and $\mathbf{B}_\perp$ (in the sense of Saikkonen and Lütkepohl 2000) must be estimated. The method consists of three steps.
Firstly, the parameters of the stochastic components $\mathbf{A}$, $\mathbf{B}$, $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, …, $\mathbf{\Gamma}_{S-1}$, $\mathbf{\Omega}$ are estimated by a conventional reduced rank regression (see Johansen 1995). Secondly, in the general case, the parameters of the deterministic part $\mathbf{h}_{\widehat{\mathbf{B}}_\perp,i}$ and $\mathbf{h}_{\widehat{\mathbf{B}},i}$, $i = 1, 2, 3, 4$ (see Equation (3)), are estimated by a generalized least squares method from the equation:

$$\widehat{\mathbf{Q}}^T\widehat{\mathbf{\Pi}}(L)\mathbf{v}_t = \sum_{i=1}^{4}\left(\widehat{\mathbf{K}}_{\widehat{\mathbf{B}},it}\mathbf{h}_{\widehat{\mathbf{B}},i} + \widehat{\mathbf{K}}_{\widehat{\mathbf{B}}_\perp,it}\mathbf{h}_{\widehat{\mathbf{B}}_\perp,i}\right), \tag{15}$$

where

$$
\begin{aligned}
\widehat{\mathbf{K}}_{\widehat{\mathbf{B}}_\perp,it} &= \widehat{\mathbf{Q}}^T\widehat{\mathbf{K}}_{it}\widehat{\mathbf{B}}_\perp(\widehat{\mathbf{B}}_\perp^T\widehat{\mathbf{B}}_\perp)^{-1},\\
\mathbf{h}_{\widehat{\mathbf{B}}_\perp,i} &= \widehat{\mathbf{B}}_\perp^T\mathbf{h}_i,\\
\widehat{\mathbf{K}}_{\widehat{\mathbf{B}},it} &= \widehat{\mathbf{Q}}^T\widehat{\mathbf{K}}_{it}\widehat{\mathbf{B}}(\widehat{\mathbf{B}}^T\widehat{\mathbf{B}})^{-1},\\
\mathbf{h}_{\widehat{\mathbf{B}},i} &= \widehat{\mathbf{B}}^T\mathbf{h}_i,\\
\widehat{\mathbf{K}}_{1t} = \widehat{\mathbf{\Pi}}(L) &= \begin{cases} \mathbf{I}, & t=1\\ \mathbf{I}-\sum_{j=1}^{t-1}\widehat{\mathbf{\Pi}}_j, & t=2,\dots,S\\ -\widehat{\mathbf{A}}\widehat{\mathbf{B}}^T, & t=S+1,\dots,T \end{cases},\\
\widehat{\mathbf{K}}_{2t} = \widehat{\mathbf{\Pi}}(L)t &= \begin{cases} \mathbf{I}, & t=1\\ t\mathbf{I}-\sum_{j=1}^{t-1}(t-j)\widehat{\mathbf{\Pi}}_j, & t=2,\dots,S\\ \widehat{\mathbf{\Psi}}-(t-1)\widehat{\mathbf{A}}\widehat{\mathbf{B}}^T, & t=S+1,\dots,T \end{cases},\\
\widehat{\mathbf{K}}_{3t} = \widehat{\mathbf{\Pi}}(L)u_{1t} &= \begin{cases} \mathbf{0}, & t<t_0\\ \mathbf{I}, & t=t_0\\ \mathbf{I}-\sum_{j=1}^{t-1}\widehat{\mathbf{\Pi}}_j, & t=t_0+1,\dots,t_0+S-1\\ -\widehat{\mathbf{A}}\widehat{\mathbf{B}}^T, & t=t_0+S,\dots,T \end{cases},\\
\widehat{\mathbf{K}}_{4t} = \widehat{\mathbf{\Pi}}(L)u_{2t} &= \\
&= \begin{cases} \mathbf{0}, & t<t_0\\ \mathbf{I}, & t=t_0\\ (t-t_0+1)\mathbf{I}-\sum_{j=1}^{t-1}(t-t_0+1-j)\mathbf{\Pi}_j, & t=t_0+1,\dots,t_0+S-1\\ \widehat{\mathbf{\Psi}}-(t-t_0)\widehat{\mathbf{A}}\widehat{\mathbf{B}}^T, & t=t_0+S,\dots,T \end{cases}.
\end{aligned}
$$

Defining the $\mathbf{Q}$ matrix as $\mathbf{Q}\mathbf{Q}^T = \mathbf{\Omega}^{-1}$ results in the spherical covariance matrix of the error term. This condition is satisfied if

$$
\widehat{\mathbf{Q}} = \left[\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{A}}(\widehat{\mathbf{A}}^T\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{A}})^{-1/2} \;\vdots\; \widehat{\mathbf{A}}_\perp(\widehat{\mathbf{A}}_\perp^T\widehat{\mathbf{\Omega}}\widehat{\mathbf{A}}_\perp)^{-1/2}\right]
$$

(see Trenkler et al. 2006). The matrices of the lag polynomial $\widehat{\mathbf{\Pi}}(L)$ can be written as $\widehat{\mathbf{\Pi}}_1 = \mathbf{I}_M + \widehat{\mathbf{A}}\widehat{\mathbf{B}}^T + \widehat{\mathbf{\Gamma}}_1$, $\widehat{\mathbf{\Pi}}_2 = \widehat{\mathbf{\Gamma}}_j - \widehat{\mathbf{\Gamma}}_{j-1}$, $\widehat{\mathbf{\Pi}}_S = -\widehat{\mathbf{\Gamma}}_{S-1}$, $j = 2,\dots,S-1$.
In the third step the Wald statistic is used for testing. Three different cases can be considered: WALD in case of simultaneous break in level and trend ($i = 3, 4$ in (14) and (15)), WALD_C – break in level ($i = 3$) and WALD_T – break in trend ($i = 4$).
The asymptotic properties of estimators in the deterministic part of DGP, derived by Saikkonen and Lütkepohl (2000) and Trenkler et al. (2006), lead to the following conclusions. Firstly, the vector of parameters $\mathbf{h}_3$ is not estimated consistently in the direction of $\mathbf{B}_\perp^T$ but is bounded in probability. Secondly, the asymptotic distribution

of $\mathbf{h}_4$ in the direction of $\mathbf{B}_\perp^T$ depends on the break date. Thirdly, estimators $\mathbf{h}_3$ and $\mathbf{h}_4$ are consistent in the direction of $\mathbf{B}^T$. Due to the consistency problems of parameters' estimators relating to the deterministic part, the asymptotic distributions of the considered statistics are not straightforward. Nevertheless, the critical values and the properties of WALD, WALD_C and WALD_T can be simulated.

3.2 Unknown break period

If the period in which a break occurs is not known, the following generalization of the Wald statistic can be used (see Andrews 1993):

$$\sup \mathrm{WALD} = \sup_\tau \mathrm{WALD}(\tau), \tag{16}$$

where $\tau \in (0,1)$ defines the break date, $t_0 = \tau T$ and T the sample size.
In practice, an unknown structural break corresponds (assuming only one break point) to the value of τ, which is related to the maximum value of the Wald statistic.
Assuming that $\mathbf{h}_3(\tau)$ and $\mathbf{h}_4(\tau)$ are vector of parameters associated with:

$$u_{1t} = \begin{cases} 1 & \text{for } t \geq [\tau T] \\ 0 & \text{for } t < [\tau T] \end{cases} \quad \text{and} \quad u_{2t} = \begin{cases} t - ([\tau T] - 1) & \text{for } t \geq [\tau T] \\ 0 & \text{for } t < [\tau T] \end{cases},$$

the null hypothesis for sup WALD tests are respectively:

$H_0 : \mathbf{h}_3(\tau) = 0 \wedge \mathbf{h}_4(\tau) = 0$ – a simultaneous break in the level and the trend,
$H_0 : \mathbf{h}_3(\tau) = 0$ – a level break,
$H_0 : \mathbf{h}_4(\tau) = 0$ – a trend break.

The distributions of the above test statistics are not standard, thus the critical values need to be simulated.

4 Monte Carlo Simulations

The critical values of the Wald statistic for each case (a level break, a trend break, and both) should be simulated under the null hypothesis $\mathbf{h}_i = 0$, $i = 3, 4$ for the 5% significance level. Additionally it was assumed that parameters associated with remaining deterministic variables which are present in DGP and which are not subject to the null hypothesis are equal to zero (the Wald statistic is invariant to them).
The stochastic component of the DGP, assuming R cointegrating vectors is given by (see Toda 1994)

$$\mathbf{y}_t = \begin{bmatrix} \rho \mathbf{I}_R & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M-R} \end{bmatrix} \mathbf{y}_{t-1} + \mathbf{e}_t, \tag{17}$$

where $\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{I})$ and $|\rho| < 1$ is the autoregressive parameter. The deterministic components are set to zero in the DGP, but they are taken into account while estimating and testing.

The critical values for WALD, WALD$_T$ and WALD$_C$ tests depend on the number of observations (T), the number of variables in the system (M), the value of the autoregressive parameter (ρ), the number of cointegrating vectors (R) and the timing of the break (τ defines the break fraction, see (16)). Therefore, they need to be simulated individually for each case. For an empirical model, the critical values can be calculated using a data-driven approach (the critical values for specific cases are presented in Appendix).

The power of tests was determined assuming the alternative hypothesis, which implies the presence of the structural break in the data generating process.

Table 1a: The power of the WALD test, $\tau = 0.5$, $\rho = 0.5$, $n = 10000$

WALD			$\vartheta = 0.05$, $\omega = 0.5$			$\vartheta = 0.075$, $\omega = 0.75$			$\vartheta = 0.1$, $\omega = 1$		
	K	R	$T = 100$	$T = 150$	$T = 200$	$T = 100$	$T = 150$	$T = 200$	$T = 100$	$T = 150$	$T = 200$
$M = 4$	3	1	0.0766	0.2672	0.7115	0.1627	0.6825	0.9824	0.3436	0.9403	0.9993
	2	2	0.1037	0.4161	0.9106	0.2480	0.8848	0.9988	0.5164	0.9900	0.9999
	1	3	0.1893	0.7674	0.9966	0.5095	0.9927	1	0.8467	0.9999	1
$M = 5$	4	1	0.0662	0.1591	0.5385	0.1230	0.4826	0.9475	0.2269	0.8393	0.9968
	3	2	0.0790	0.2387	0.7459	0.1599	0.6919	0.9913	0.3391	0.9480	0.9968
	2	3	0.1123	0.4328	0.9463	0.2615	0.9114	0.9994	0.5627	0.9956	1
	1	4	0.1937	0.8081	0.9981	0.5378	0.9959	1	0.8758	0.9998	1

Table 1b: The power of the WALD$_C$ test, $\tau = 0.5$, $\rho = 0.5$, $n = 10000$

WALD$_C$			$\omega = 1$			$\omega = 1.5$			$\omega = 2$		
	K	R	$T = 100$	$T = 150$	$T = 200$	$T = 100$	$T = 150$	$T = 200$	$T = 100$	$T = 150$	$T = 200$
$M = 4$	3	1	0.7380	0.8918	0.9440	0.8988	0.9653	0.9816	0.9538	0.9845	0.9931
	2	2	0.6253	0.7634	0.8348	0.7910	0.8888	0.9261	0.8766	0.9353	0.9567
	1	3	0.4554	0.5612	0.6225	0.6116	0.6905	0.7435	0.7061	0.7718	0.8063
$M = 5$	4	1	0.7313	0.9170	0.9683	0.9223	0.9794	0.9937	0.9664	0.9932	0.9972
	3	2	0.6678	0.8349	0.9056	0.8596	0.9408	0.9671	0.9314	0.9746	0.9888
	2	3	0.5895	0.7283	0.7992	0.7757	0.8567	0.9004	0.8624	0.9178	0.9438
	1	4	0.4336	0.5418	0.5993	0.5884	0.6781	0.7204	0.6821	0.7590	0.7951

If there is a simultaneous break in level and trend the DGP is defined as follows

$$\mathbf{v}_t = \mathbf{y}_t + \mathbf{h}_3 u_{1t} + \mathbf{h}_4 u_{2t}, \tag{18}$$

where $\mathbf{y}_t$ is defined by (17), while u_{1t} and u_{2t} are given by

Table 1c: The power of the WALD_T test, $\tau = 0.5$, $\rho = 0.5$, $n = 10000$

WALD_T			$\vartheta = 0.05$			$\vartheta = 0.1$			$\vartheta = 0.15$		
	K	R	$T = 100$	$T = 150$	$T = 200$	$T = 100$	$T = 150$	$T = 200$	$T = 100$	$T = 150$	$T = 200$
$M = 4$	3	1	0.0525	0.1349	0.4126	0.1182	0.5610	0.9478	0.2495	0.8752	0.9935
	2	2	0.0759	0.2845	0.8145	0.2466	0.9277	0.9994	0.6221	0.9949	0.9997
	1	3	0.1409	0.7093	0.9925	0.6963	0.9995	1.0000	0.9710	1.0000	1.0000
$M = 5$	4	1	0.0518	0.0993	0.2574	0.0829	0.3437	0.8561	0.1602	0.7007	0.9739
	3	2	0.0610	0.1548	0.5492	0.1404	0.7285	0.9908	0.3641	0.9570	0.9994
	2	3	0.0793	0.3213	0.8815	0.3194	0.9661	0.9997	0.7430	0.9970	1.0000
	1	4	0.1508	0.7591	0.9969	0.7519	0.9996	1.0000	0.9824	1.0000	1.0000

$$u_{1t} = \begin{cases} 1 & \text{for } t \geq [\tau T], \\ 0 & \text{for } t < [\tau T], \end{cases} \quad \text{and} \quad u_{2t} = \begin{cases} t - ([\tau T] - 1) & \text{for } t \geq [\tau T], \\ 0 & \text{for } t < [\tau T]. \end{cases}$$

It was assumed that the structural break in trend affects every variable in the system with the same magnitude, $\mathbf{h}_4 = \vartheta[1]_{M\times 1}$, where $[1]_{M\times 1}$ denotes a vector of ones, and $\vartheta = \{0.05, 0.1, 0.15\}$. The parameters in vector $\mathbf{h}_3$ are assumed to be proportional to the standard deviation of the process, $\mathbf{h}_3 = \omega\delta_{\mathbf{y}}$, where $\delta_{\mathbf{y}}$ is a $M\times 1$ vector of standard deviations for variables in $\mathbf{y}$ and $\omega = 10\cdot\vartheta$, which implies that $\omega = \{0.5, 0.75, 1\}$ (see Perron and Yabu 2007). For each individual case, the appropriate critical values were simulated.

Break in level assumes $\mathbf{h}_4 = 0$ in (18). The values of the parameters associated with the structural break (vector $\mathbf{h}_3$) were set to be proportional to the mean of the process, $\mathbf{h}_3 = \omega\mathbf{h}_1$. The power of the WALD_C test was simulated for three different values of ω, $\omega = \{1, 1.5, 2\}$.

In the case of break in trend $\mathbf{h}_3 = 0$ in (18), and as previously, $\mathbf{h}_4 = \vartheta[1]_{M\times 1}$ but $\vartheta = \{0.05, 0.075, 0.1\}$. Tables 1a, 1a, 1c present how the power of particular tests varies with the amplitude of the break (ω, ϑ), the sample size $T \in \{100, 150, 200\}$, the number of variables ($M = 4, 5$), and the number of cointegrating vectors $R \in (0, M)$. In power simulations it was assumed that the structural break occurs in the middle of the sample ($\tau = 0.5$), while the autoregressive parameter is $\rho = 0.5$. It can be concluded that in all cases (WALD, WALD_T, WALD_C) the power of test increases with the sample size and the magnitude of the break as could be expected. Similar power manifests itself for the same number of common stochastic trends (K).

In the next experiments additional properties of the proposed tests were analyzed. Firstly, it was checked whether the power of the tests is sensitive to the error terms distribution.

The power was recalculated for the t-Student(5) distributed residuals with 5 degrees of freedom. For all tests the process generating the stochastic component was defined by (17) with $\mathbf{e}_t \sim t$-Student(5). The results lead to the conclusion that the power of tests is not significantly affected by the error's distribution (see Table 2).

Table 2: The impact of errors' distribution on the power of the WALD, WALD_T, WALD_C for different break sizes for T=200, M=4, R=2

WALD		$\vartheta = 0.05$, $\omega = 0.5$		$\vartheta = 0.075$, $\omega = 0.75$		$\vartheta = 0.1$, $\omega = 1$	
K	R	Normal	t-Student(5)	Normal	t-Student(5)	Normal	t-Student(5)
3	1	0.7115	0.5316	0.9824	0.9374	0.9993	0.9975
2	2	0.9106	0.7226	0.9988	0.9889	0.9999	0.9997
1	3	0.9966	0.9509	1	0.9998	1	1.0000
WALD_C		$\omega = 1$		$\omega = 1.5$		$\omega = 2$	
K	R	Normal	t-Student(5)	Normal	t-Student(5)	Normal	t-Student(5)
3	1	0.9440	0.9375	0.9816	0.9785	0.9931	0.9917
2	2	0.8348	0.8298	0.9261	0.9212	0.9567	0.9511
1	3	0.6225	0.6184	0.7435	0.7394	0.8063	0.8087
WALD_T		$\vartheta = 0.05$		$\vartheta = 0.1$		$\vartheta = 0.15$	
K	R	Normal	t-Student(5)	Normal	t-Student(5)	Normal	t-Student(5)
3	1	0.4126	0.2403	0.9478	0.8249	0.9935	0.9749
2	2	0.8145	0.5333	0.9994	0.9895	0.9997	0.9996
1	3	0.9925	0.9211	1.0000	1.0000	1.0000	1.0000

Table 3: The impact of τ on the power of WALD, WALD_T, WALD_C tests for T=200, M=4, R=2

				WALD					
ω / ϑ	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
0.05/0.05	0.5743	0.5674	0.7324	0.8554	0.9106	0.9358	0.9279	0.8996	0.7993
0.075/0.75	0.8884	0.9320	0.9896	0.9979	0.9988	0.9992	0.9985	0.9979	0.9894
1.0/0.1	0.9827	0.9966	0.9998	0.9999	0.9999	1.0000	1.0000	1.0000	0.9998
				WALD_C					
ω	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
1	0.9134	0.8753	0.8512	0.8368	0.8348	0.8370	0.8553	0.8810	0.9128
1.5	0.9602	0.9441	0.9257	0.9284	0.9261	0.9217	0.9342	0.9429	0.9632
2	0.9806	0.9681	0.9606	0.9536	0.9567	0.9609	0.9628	0.9666	0.9820
				WALD_T					
ϑ	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
0.05	0.1064	0.3600	0.6458	0.7739	0.8145	0.7801	0.6467	0.3678	0.1023
0.1	0.3076	0.9489	0.9967	0.9987	0.9994	0.9986	0.9976	0.9631	0.3443
0.15	0.6643	0.9991	1.0000	1.0000	0.9997	0.9999	0.9999	0.9996	0.7146

Furthermore, the influence of the break point on the power of the tests was analysed. Nine values for the break fraction were considered $\tau \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ for the system of four variables with two cointegrating vectors. The power of the WALD_C test proved to be invariant to the timing of the break. On the contrary, the WALD and WALD_T have the smallest power if the structural break appears at the beginning or at the end of the sample while for $\tau \in [0.3, 0.7]$ the power is not influenced significantly by the timing of the break (see Table 3).

Table 4: The impact of ρ on the power of WALD, WALD_C, WALD_T tests for T=200, M=4, R=2

					WALD				
ω / ϑ	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
0.05/0.05	0.9997	0.9990	0.9967	0.9807	0.9106	0.7112	0.3766	0.1558	0.0882
0.075/0.75	1.0000	1.0000	1.0000	0.9998	0.9988	0.9833	0.8415	0.4721	0.2178
1.0/0.1	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9863	0.8159	0.4587
					WALD_C				
ω	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
1	0.8516	0.8539	0.8433	0.8393	0.8348	0.8188	0.8005	0.7720	0.7485
1.5	0.9318	0.9289	0.9282	0.9247	0.9261	0.9105	0.9087	0.8892	0.8799
2	0.9643	0.9591	0.9602	0.9591	0.9567	0.9492	0.9455	0.9368	0.9301
					WALD_T				
ϑ	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$	$\rho = 0.5$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
0.05	0.9993	0.9971	0.9894	0.9464	0.8145	0.5175	0.2252	0.0878	0.0550
0.1	1.0000	1.0000	1.0000	0.9999	0.9994	0.9886	0.8693	0.4183	0.1450
0.15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9991	0.9840	0.8122	0.3526

The impact of the value of the autoregressive parameter in the DGP on the power of the proposed tests was estimated for $\rho \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. The estimation results are presented in Table 4. The power of the WALD_C test is relatively insensitive to the value of ρ, whereas the power of the WALD and WALD_T tests is the smallest for high values of ρ, i.e. $\rho \in [0.8, 0.9]$ and small magnitude of the break.

In the fourth simulation experiment the size properties of tests were considered. The empirical rejection frequencies of the true null hypothesis are very close to 5% for all tests (see Table 5).

Finally, the properties of $\sup \mathrm{WALD}$, $\sup \mathrm{WALD}_T$, $\sup \mathrm{WALD}_C$ were analyzed by calculating the probabilities of obtaining the maximum value of the Wald statistic in the same break date, which was assumed in the data generating process. In the first step, data with a structural break in the deterministic component (the level, the trend, and both) were generated for $\tau \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$. In the second step, a τ with the maximum value of the Wald statistic was determined. The number of replications was equal to 500.

In case of models with level break and simultaneous break in level and trend, the

Table 5: The size of the WALD, WALD_C, WALD_T tests

		WALD		WALD_C		WALD_T	
	R	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$M = 4$	1	0.0525	0.0531	0.0503	0.0528	0.0455	0.0498
	2	0.0537	0.0504	0.0516	0.0501	0.0548	0.0492
	3	0.0516	0.0482	0.0512	0.0511	0.0485	0.0499
$M = 5$	1	0.0495	0.0515	0.0444	0.0526	0.0468	0.0511
	2	0.0526	0.0508	0.0450	0.0477	0.0515	0.0447
	3	0.0529	0.0471	0.0468	0.0466	0.0537	0.0460
	4	0.0499	0.0514	0.0496	0.0497	0.0507	0.0510

Table 6: The probabilities of obtaining the maximum value of the Wald statistic in the same break date, which was assumed in the data generating process for sup WALD, sup WALD_T, sup WALD_C

	sup WALD						
ϑ, ω	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$
$\vartheta = 0.05$, $\omega = 0.5$	0.5540	0.6120	0.6280	0.6560	0.6560	0.6080	0.5580
$\vartheta = 0.1$, $\omega = 1$	0.9020	0.9040	0.9080	0.9120	0.8980	0.8860	0.8800
$\vartheta = 0.15$, $\omega = 1.5$	0.9620	0.9660	0.9720	0.9720	0.9640	0.9640	0.9600
	sup WALD_T						
ϑ	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$
0.05	0.1440	0.2140	0.4680	0.6940	0.4400	0.1900	0.1280
0.1	0.2180	0.3940	0.7340	0.9260	0.6780	0.3340	0.1460
0.15	0.3340	0.6620	0.9160	0.9760	0.8440	0.5140	0.2340
	sup WALD_C						
ω	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$
0.5	0.584	0.6140	0.6460	0.693	0.6720	0.6340	0.585
1	0.97	0.9840	0.9780	0.971	0.9780	0.9900	0.9700
1.5	0.999	1	1	0.999	0.9980	1	0.9980

results are not influenced by the break point (see Table 6), while in case of sup WALD_τ the concerned probability is the highest for $\tau = 0.5$. As expected, in each case the probability of identifying the true break point increases as ω and ϑ increase.

5 Conclusions

The research has shown that a structural break occurring in the DGP in period $t - 1$ requires the presence of a binary variable in the cointegration vectors in period t. The appropriate binary variable must be simultaneously added to the outside of the cointegration space.

The Monte Carlo simulations prove that the power of the proposed tests while used for testing for the presence of structural break in the deterministic part of data generating process increases with the sample size and the magnitude of the break and is not significantly affected by the distribution of the error terms (normal versus t-Student(5)).

Acknowledgements

The authors have received many valuable comments from the participants of Macromodels International Conference and workshops at the Warsaw School of Economics. All remaining errors are the authors' responsibility. Financial support from National Science Centre under OPUS 21: DEC-2021/41/B/HS4/04317 grant is gratefully acknowledged.

References

[1] Andrews D. W. K., (1993), Tests for parameter instability and structural change with unknown change point, *Econometrica* 61(4), 821–856, DOI: 10.2307/2951764.

[2] Gosińska E., (2015), Modelowanie procesów ekonomicznych generowanych przez niestacjonarne procesy stochastyczne ze zmianą strukturalną (Modelling of economic processes generated by nonstationary stochastic processes with structural break), unpublished Phd thesis, Biblioteka Uniwersytetu Łódzkiego.

[3] Johansen S., Nielsen M., (2018), The cointegrated vector autoregressive model with general deterministic terms, *Journal of Econometrics* 202, 214–229, DOI: 10.1016/j.jeconom.2017.10.003.

[4] Johansen S., (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

[5] Juselius K., (2006), *The Cointegrated VAR Model: Methodology and Applications*, Oxford University Press.

[6] Lütkepohl H., (2005), *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.

[7] Nielsen H. B., (2004), Cointegration analysis in the presence of outliers, *Econometrics Journal* 7, 249–271.

[8] Perron P., Yabu T., (2009), Testing for shifts in trend with an integrated or stationary noise component, *Journal of Business and Economic Statistics* 27(3), 369–396. DOI: 10.1198/jbes.2009.07268.

[9] Saikkonen P., Lütkepohl H., (2000), Testing for the cointegrating rank of a VAR process with structural shifts, *Journal of Business and Economic Statistics* 18, 451–464. DOI: 10.2307/1392226.

[10] Sobreira N., Nunes L.C., (2012), Testing for Broken Trends in Multivariate Time Series, Universidade Nova de Lisboa Working Paper, available at: `https://www.academia.edu/15294230/Testing_for_Broken_Trends_in_Multivariate_Time_Series`.

[11] Toda H. Y., (1994), Finite sample properties of likelihood ratio tests for cointegrating ranks when linear trends are present, *The Review of Economics and Statistics* 76(1), 66–79. DOI: 10.2307/2109827.

[12] Trenkler C., Saikkonen P., Lütkepohl H., (2006), Testing for the Cointegrating Rank of a VAR Process with Level Shift and Trend Break, EUI Working Papers, ECO No. 2006/29, available at: `https://cadmus.eui.eu//handle/1814/6306`.

Appendix

Defining $\mathbf{h} = \begin{bmatrix} \mathbf{h}^T_{\widehat{\mathbf{B}},1} & \mathbf{h}^T_{\widehat{\mathbf{B}}_\perp,1} & \mathbf{h}^T_{\widehat{\mathbf{B}},2} & \mathbf{h}^T_{\widehat{\mathbf{B}}_\perp,2} & \mathbf{h}^T_{\widehat{\mathbf{B}},3} & \mathbf{h}^T_{\widehat{\mathbf{B}}_\perp,3} & \mathbf{h}^T_{\widehat{\mathbf{B}},4} & \mathbf{h}^T_{\widehat{\mathbf{B}}_\perp,4} \end{bmatrix}^T$, restriction matrices for WALD are as follows

$$\mathbf{\Theta}_B = \begin{bmatrix} \mathbf{0}_{R\times 2M} & \mathbf{I}_{R\times R} & \mathbf{0}_{R\times(M-R)} & \mathbf{0}_{R\times R} & \mathbf{0}_{R\times(M-R)} \\ \mathbf{0}_{R\times 2M} & \mathbf{0}_{R\times R} & \mathbf{0}_{R\times(M-R)} & \mathbf{I}_{R\times R} & \mathbf{0}_{R\times(M-R)} \end{bmatrix},$$

$$\mathbf{\Theta}_{B_\perp} = \begin{bmatrix} \mathbf{0}_{(M-R)\times 2M} & \mathbf{0}_{(M-R)\times R} & \mathbf{I}_{(M-R)\times(M-R)} & \mathbf{0}_{(M-R)\times R} & \mathbf{0}_{(M-R)\times(M-R)} \\ \mathbf{0}_{(M-R)\times 2M} & \mathbf{0}_{(M-R)\times R} & \mathbf{0}_{(M-R)\times(M-R)} & \mathbf{0}_{(M-R)\times R} & \mathbf{I}_{(M-R)\times(M-R)} \end{bmatrix},$$

analogously for WALD_C

$$\mathbf{\Theta}_B = \begin{bmatrix} \mathbf{0}_{R\times 2M} & \mathbf{I}_{R\times R} & \mathbf{0}_{R\times(M-R)} & \mathbf{0}_{R\times M} \end{bmatrix},$$
$$\mathbf{\Theta}_{B_\perp} = \begin{bmatrix} \mathbf{0}_{(M-R)\times 2M} & \mathbf{0}_{(M-R)\times R} & \mathbf{I}_{(M-R)\times(M-R)} & \mathbf{0}_{(M-R)\times M} \end{bmatrix},$$

and for WALD_T

$$\mathbf{\Theta}_B = \begin{bmatrix} \mathbf{0}_{R\times 3M} & \mathbf{I}_{R\times R} & \mathbf{0}_{R\times(M-R)} \end{bmatrix},$$
$$\mathbf{\Theta}_{B_\perp} = \begin{bmatrix} \mathbf{0}_{(M-R)\times 3M} & \mathbf{0}_{(M-R)\times R} & \mathbf{I}_{(M-R)\times(M-R)} \end{bmatrix}.$$

Table A1: Simulated critical values of WALD, WALD_C and WALD_T ($\tau = 0.3$, $\rho = 0.5$, $n = 100000$)

			WALD			WALD_C			WALD_T		
	K	R	T=100	T=150	T=200	T=100	T=150	T=200	T=100	T=150	T=200
M=2	1	1	27.907	21.258	18.682	16.22	13.517	12.48	21.32	15.858	13.757
M=3	2	1	49.409	35.507	30.327	29.1	22.871	20.38	38.095	27.170	22.846
	1	2	59.638	38.913	32.825	33.36	24.571	21.507	48.997	31.176	25.655