Hitotsubashi University

Global COE Hi-Stat Discussion Paper Series 187

Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

Statistical Inference in Possibly Integrated/Cointegrated Vector Autoregressions: Application to Testing for Structural Changes

Eiji Kurozumi
Khashbaatar Dashtseren

April 2011

Hi-Stat
Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitatchi Tokyo, 186-8601 Japan
http://gcoe.ier.hit-u.ac.jp

Hi-Stat Discussion Paper

Statistical Inference in Possibly Integrated/Cointegrated Vector Autoregressions: Application to Testing for Structural Changes[1]

Eiji Kurozumi[2] Khashbaatar Dashtseren

Department of Economics

Hitotsubashi University

April, 2011

Abstract

We develop a new approach of statistical inference in possibly integrated/cointegrated vector autoregressions. Our method is built on the two previous approaches: the lag augmented approach by Toda and Yamamoto (1995) and the artificial autoregressions by Yamamoto (1996). We show that our estimator is asymptotically normally distributed irrespective of whether the variables are stationary or nonstationary, and that the Wald test statistic for the parameter restrictions has an asymptotic chi-square distribution. Using this method, we also propose to test for multiple structural changes. We show that our test statistics have the same limiting distributions as in the standard case, irrespective of whether the variables are stationary, purely integrated, or cointegrated.

JEL classification: C12; C13; C32

Keywords: multiple breaks, stationary, unit root, cointegration

[1]This research was supported by the Global COE program of the Research Unit for Statistical and Empirical Analysis in Social Sciences, Hitotsubashi University.

[2]Correspondence: Eiji Kurozumi, Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan. E-mail: kurozumi@stat.hit-u.ac.jp

1. Introduction

This paper develops a new method of estimation and statistical inference in stationary, purely integrated, or cointegrated vector autoregressions (VAR), and also studies the tests for structural changes. The VAR models have long been used in empirical analyses to capture the relations between economic variables. Since macroeconomic variables sometimes show nonstationary behaviors, usually, before estimating the VAR models, we first test for a unit root for each variable and if we find the evidence of nonstationarity, then we test for cointegration. One of the most widely used unit root tests is the (augmented) Dickey-Fuller test by Dickey and Fuller (1979) and Said and Dickey (1984); the cointegrating rank is estimated using the system approach by Ahn and Reinsel (1990) and Johansen (1988, 1991, 1995).

Often, once the order of integration/cointegration is determined, we make a statistical inference about the coefficients. In this case, the knowledge of the order of integration/cointegration is crucial to make a valid inference; if the estimated order is different from the true one, then our statistical inference would be invalid and we may come to a wrong conclusion. That is, we have to be careful about the pre-test bias when estimating the order of integration/cointegration. However, in some cases, we are not much interested in the order of integration/cointegration but wish only to test for the hypothesis about the coefficients. In such a case, statistical methods robust to the order of integration/cointegration are useful in practical analyses in order to avoid the pre-test bias; hence, several methods have been proposed in the literature. For example, Phillips (1995) develops the fully modified VAR (FM-VAR) approach, wherein the Wald test statistic has a limiting distribution that is bounded above by a chi-square distribution, so that the test with chi-square critical values becomes asymptotically conservative. Toda and Yamamoto (1995) propose to estimate a model with intentionally augmented lags, and show that the estimated parameter of interest has a limiting normal distribution irrespective of whether the variables are (trend) stationary, integrated, or cointegrated. This lag-augmented (LA) method is further modified by Kurozumi and Yamamoto (2000) in order to reduce the estimation bias. While these methods are mainly developed for the inference about the reduced VAR models, Hsiao and Wang (2006, 2007) propose the estimation methods for structural VAR models using techniques similar to the FM-VAR and the LA-VAR approaches.

The above methods are useful if we are interested in only the coefficients associated with the lagged dependent variables, in such cases as testing for Granger-noncausality and investigating the term structure of interest rates. However, such methods are not necessarily suitable for testing hypotheses including restrictions on the coefficients associated with an intercept and a linear trend, because the corresponding estimators by the above methods depend on the stationary/nonstationary nature of the variables. Yamamoto (1996) proposes the artificial vector autoregressions (ART-VAR) approach, but he also shows that the ART-VAR method is quite inefficient in small samples.

In this paper, we develop an alternative approach to making a statistical inference about all the regression coefficients, including those associated with a constant and a linear trend. Our method is basically a combination of the LA-VAR and the ART-VAR approaches. We show that the estimator used in our approach has a limiting normal distribution irrespective of whether the variables are stationary, purely integrated, or cointegrated. As a result, the Wald test statistic for the hypothesis about the coefficients weakly converges to a chi-square distribution under the null hypothesis.

In addition to the order of integration/cointegration, we also have to consider the possibility of structural changes when we investigate the data in long samples because neglecting the presence of structural breaks may invalidate the statistical inference. For stationary models, the tests for structural changes have long been developed in the literature: the sup-type test by Andrews (1993) and the exponential-type and the averaging-type tests by Andrews, Lee and Ploberger (1996) are commonly used in practice to test for a one time structural change (and possibly multiple structural changes); the tests for multiple structural changes by taking the possible number of structural changes into account have been proposed by Bai and Perron (1998) and Qu and Perron (2007). On the other hand, for cointegrated models, we need to investigate both the order of cointegration and the existence of structural changes. Saikkonen and Lütkepohl (2000) and Lütkepohl, Saikkonen and Trenkler (2003) propose tests for the cointegrating rank with deterministic shifts with a known timing, while the unknown case is considered by Inoue (1999) and Lütkepohl, Saikkonen and Trenkler (2004). Regarding structural change tests, Quintos (1995) proposes tests for cointegrating vectors with a known break point, while Quintos (1997), Seo (1998), Hansen and Johansen (1999), and Qu (2007) develop tests for changes with unknown points. Tests for structural changes for cointegrating regressions (and not VAR models) are investigated by Bai, Lumsdaine and Stock (1998) with one time change and Kejriwal and Perron (2008, 2010b) with multiple changes. For the known break point and the known cointegrating rank, Hansen (2003) proposes the likelihood ratio test for parameter restrictions. Unfortunately, the above tests for cointegrating rank with structural changes assume that we know the existence of structural changes, whereas the tests for structural changes require knowing the cointegrating rank. Therefore, the tests for cointegration and structural changes present a circular testing problem, and thus, the existing approaches appear limited when structural changes are incorporated into possibly integrated/cointegrated VAR.

As the second contribution of this paper, we develop tests for structural changes for possibly nonstationary VAR models using the same method as explained above: we combine the LA-VAR and the ART-VAR methods. We propose the sup-type, the exponential-type, and the averaging-type tests by Andrews (1993) and Andrews, Lee and Ploberger (1996) and the double maximum tests by Bai and Perron (1998) and show that they have the same limiting null distributions as in the standard case, and hence, we can use the existing critical value tables. Again, the advantage of our method is that we do not require the knowledge of the order of integration/cointegration, and can use the same test statistics in any case. Note that recently, the trend and/or level breaks tests, which are robust to the stationary/unit root property in the stochastic term, have been investigated in the literature; see Harvey, Leybourne and Taylor (2009, 2010), Perron and Yabu (2009) and Kejriwal, and Perron (2010a) among others. These papers only consider deterministic shifts whereas we propose tests for the structural changes in the whole structure of VAR models; thus, the main purpose of this paper is different from that of the above mentioned papers.

The rest of the paper is organized as follows. Section 2 describes the model and the assumptions. In Section 3, we explain our method of estimating models, which is robust to the order of integration/cointegration. We first transform a model in order to avoid asymptotic multicollinearity in the regressors, and show that the coefficients associated with the transformed regressors are asymptotically normal. We then show that the standard Wald test statistic based on the original estimator can be expressed as a nonsingular transformation of the transformed estimator, and that it has a limiting chi-square distribution. Section 4 deals with the tests for multiple structural changes and shows that the limiting distributions of our tests are the same as in the standard case. The finite sample properties of the structural change tests are investigated in Section 5. The concluding remarks are given in Section 6. The technical derivations are relegated to the Appendix.

2. Model and Assumptions

Let us consider the following n-dimensional VAR model of order p (VAR(p)):

$$y_t = c_0 + c_1 \frac{t}{T} + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + u_t \quad (t = 1, \cdots, T), \tag{1}$$

where y_ts are n-dimensional observations, c_0 and c_1 are $n \times 1$ coefficient vectors associated with the deterministic terms, $\Phi_1, \cdots, \Phi_p$ are $n \times n$ coefficient matrices, and u_t is a sequence of innovations. We allow y_t to be a stationary, pure integrated, or cointegrated process. More precisely, we make the following assumptions in this paper.

Assumption A1 *(a) $\{u_t\}$ is a martingale difference sequence with respect to $\mathcal{F}_t = \sigma\{u_t, u_{t-1}, \cdots\}$ with $E[u_t u_t' | \mathcal{F}_{t-1}] = \Sigma_u > 0$ for all t. (b) $\sup_t E\|u_t\|^{4+\kappa} < \infty$ for some $\kappa > 0$.*

Assumption A2 *The lag polynomial $\Phi(L) = I_n - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p$ satisfies either of the following: (a) $|\Phi(z)| = 0$ implies $|z| > 1$ or (b) some solutions of $|\Phi(z)| = 0$ equal 1 and the other solutions lie outside the unit circle. With regard to case (b), we also assume the following. (b-i) $\Pi = \sum_{j=1}^{p} \Phi_j - I_n$ can be decomposed such that $\Pi = \alpha\beta'$, where α and β are $n \times r$ matrices of rank r $(0 \leq r < n)$. (b-ii) Let $\alpha_\perp$ and $\beta_\perp$ be $n \times (n - r)$ full column rank matrices such that $\alpha'\alpha_\perp = \beta'\beta_\perp' = 0$ for $0 < r < n$ and $\alpha_\perp = \beta_\perp = I_n$ for $r = 0$. Then, $\alpha_\perp'(I_n + \sum_{j=2}^{k}(j-1)\Phi_j)\beta_\perp$ has full rank $n - r$. (b-iii) $\alpha_\perp' c_1 = 0$ if $c_1 \neq 0$, and $\alpha_\perp' c_0 = 0$ if $c_1 = 0$.*

Assumption A1 is standard in the time series literature to ensure that the weak law of large numbers (WLLN) and the functional central limit theorem (FCLT) hold. Assumption

A2 excludes the explosive case but allows y_t to be either (trend) stationary or cointegrated. In the case of cointegration, the integrated order is at most one and the I(2) case is excluded by (b-ii). In addition, by (b-iii), when $c_1 \neq 0$, y_t may have a linear trend but is not allowed to have a quadratic trend, while y_t does not have a linear trend when $c_1 = 0$. For details, see Chapter 5 in Johansen (1995).

3. Robust Estimation

Suppose that we want to estimate c_0, c_1, and $\Phi_1, \cdots, \Phi_p$, and to make a statistical inference about them but we are not interested in the integrated order of y_t. In other words, we want to establish the method of estimation and statistical inference about all the coefficients, which is robust to the integrated order. Note that if we are interested in only the lag's coefficients, then we may use the LA method by Toda and Yamamoto (1995) or the FM method by Phillips (1995). However, since we also want to investigate the coefficients associated with a constant and a linear trend, we cannot use these methods. On the other hand, Yamamoto (1996) proposes the ART-VAR approach, which provides a statistical method that is robust to the integrated order of time series, but he also points out that this method is inefficient. In fact, the original ART-VAR approach includes many additional artificial regressors, so much that the total number of regressors is doubled as compared to the original model; as such, the degrees of freedom in the original ART-VAR is much reduced. Moreover, the convergence order of the estimator of $\Phi_1, \cdots, \Phi_p$ in the ART-VAR approach is slower than $\sqrt{T}$, while our estimator is $\sqrt{T}$-consistent as will be shown later. As a result, the estimator based on the original ART-VAR is quite inefficient, and hence, the tests based on it lose power.

In order to avoid this inefficiency of the ART-VAR method, we consider a combination of the LA-VAR and the ART-VAR approaches and estimate the following regression:

$$y_t = c_0 + c_1 \frac{t}{T} + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \Phi_{p+1} y_{t-p-1} + d_0 \eta_{0t} + d_1 \eta_{1t} + u_t, \tag{2}$$

where $\Phi_{p+1} = 0$, $d_0 = d_1 = 0$, $\eta_{0t} = 1 - (1/T^{\lambda_0})\varepsilon_{0t}$, and $\eta_{1t} = (t/T)\{1 - (1/T^{\lambda_1})\varepsilon_{1t}\}$ with $0 < \lambda_0, \lambda_1 < 1/2$ (which must be pre-determined by a researcher) and with ε_{0t} and ε_{1t} being artificially generated i.i.d. random variables with unit variances. The selection of λ_0 and λ_1 will be discussed later. Although ε_{0t} and ε_{1t} can be drawn from any random generators with finite fourth moments, we assume that they are independent (pseudo) standard normal random variables generated by a computer. In regression (2), the extra lag y_{t-p-1} is introduced so that y_{t-p} and y_{t-p-1} are cointegrated under Assumption A2(b), as suggested by Toda and Yamamoto (1995), while the artificial regressors η_{0t} and η_{1t} are included as regressors along the line of Yamamoto (1996). That is, the artificial regressor η_{0t} is introduced so that the order of the constant term can be reduced by linear combinations with η_{0t} from $O_p(1)$ to $O_p(1/T^{\lambda_0})$. Similarly, the order of t/T is $O_p(1)$ but it is reduced to $O_p(1/T^{\lambda_1})$ by linear combinations of t/T and η_{1t}. Intuitively, we can consider a constant and η_{0t} (t/T and η_{1t}) to be "cointegrated" in the sense that some linear combinations between the two variables reduce the original order. This reduction in the original order plays an important role when making a statistical inference under Assumption A2(b).

Let us express regression (2) as

$$\begin{aligned} y_t &= \Theta x_t + u_t \\ &= (x_t' \otimes I_n)\theta + u_t, \end{aligned} \tag{3}$$

where $\Theta = [c_0, c_1, \Phi_1, \cdots, \Phi_p, \Phi_{p+1}, d_0, d_1]$, $x_t = [1, t/T, y_{t-1}', \cdots, y_{t-p}', y_{t-p-1}', \eta_{0t}, \eta_{1t}]'$ and $\theta = vec(\Theta)$. Since we are interested in the first $n \times (np+2)$ sub-matrix of Θ ($[c_0, c_1, \Phi_1, \cdots, \Phi_p]$), we consider the following linear hypothesis:

$$R_L \Theta R_R = Q \quad \text{or equivalently,} \quad R\theta = q, \tag{4}$$

where R_L is a $k_1 \times n$ full row rank matrix; $R_R = [R_{R,1}', 0]'$ is an $\{n(p+1)+4\} \times k_2$ full column rank matrix with $R_{R,1} = diag\{I_2, R_\phi\}$ so that the restrictions on c_0, c_1, and Φ_js for $j = 1, \cdots, p$ are separate; $R = R_R' \otimes R_L$ with full row rank $k = k_1 \times k_2$ and $q = vec(Q)$. We consider the restrictions on c_0, c_1, and Φ_js to be separate because their convergence orders are different, as will be shown later. Then, the Wald test statistic for restrictions (4) is given by

$$W_T = (R\hat{\theta} - q)' \left[R \left\{ \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \otimes \hat{\Sigma}_u \right\} R' \right]^{-1} (R\hat{\theta} - q), \tag{5}$$

where $\hat{\theta}$ is the least squares estimator of θ and $\hat{\Sigma}_u = \sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$ with $\hat{u}_t$ being the regression residuals.

In order to derive the asymptotic distribution of the Wald test statistic, we first transform the regressors in (3), and then investigate the asymptotic property of the transformed coefficients because the regressors in (3) are asymptotically multicollinear in either the stationary case or the cointegrated case. Let H be an $n(p+1)+4$ square full rank matrix given by

$$H = \begin{bmatrix} I_2 & 0 & -I_2 \\ F & D & 0 \\ 0 & 0 & I_2 \end{bmatrix} \quad \text{and} \quad H^{-1} = \begin{bmatrix} I_2 & 0 & I_2 \\ -D^{-1}F & D^{-1} & -D^{-1}F \\ 0 & 0 & I_2 \end{bmatrix},$$

where D is an $n(p+1)$ square full rank matrix and F is an $n(p+1)\times 2$ matrix. The exact definitions of D and F depend on Assumption A2, and are explained below. Using H we transform (3) as

$$\begin{aligned} y_t &= \Theta H^{-1} H x_t + u_t \\ &= \Theta^* x_t^* + u_t \\ &= (x_t^{*\prime} \otimes I_n)\theta^* + u_t, \end{aligned}$$

where $\Theta^* = \Theta H^{-1}$, $x_t^* = Hx_t$, and $\theta^* = vec(\Theta^*)$. The centered least squares estimator of θ normalized by a scaling matrix $G^* \otimes I_n$ is given by

$$(G^* \otimes I_n)(\hat{\theta}^* - \theta^*) = \left[\left(\sum_{t=1}^{T} G^{*-1} x_t^* x_t^{*\prime} G^{*-1} \right)^{-1} \otimes I_n \right] \left(\sum_{t=1}^{T} G^{*-1} x_t^* \otimes u_t \right).$$

Let $E[y_t] = \mu_0 + \mu_1(t/T)$. Then, for the stationary case under Assumption A2(a), $z_t = y_t - \mu_0 - \mu_1(t/T)$ becomes a zero-mean stationary VAR(p) process. In this case, we define

$$F = F_0 = \begin{bmatrix} -\mu_0 & -\mu_1 \\ \vdots & \vdots \\ -\mu_0 & -\mu_1 \end{bmatrix} \quad \text{and} \quad D = D_0 = I_{n(p+1)},$$

so that $x_t^* = [(1/T^{\lambda_0})\varepsilon_{0t},\ (t/T^{1+\lambda_1})\varepsilon_{1t}, z_{t-1}', \cdots, z_{t-p}', z_{t-p-1}', \eta_{0t}, \eta_{1t}]'$. As a result, the sample second moment of x_t^* using a scaling matrix given by $G^* = G_0^* = diag\{T^{1/2-\lambda_0}, T^{1/2-\lambda_1}, \sqrt{T} I_{n(p+1)+2}\}$, becomes asymptotically block diagonal as given by

$$\sum_{t=1}^{T} G_0^{*-1} x_t^* x_t^{*\prime} G_0^{*-1} \xrightarrow{p} \left[\begin{array}{ccc|cc} 1 & 0 & 0 & & 0 \\ 0 & 1/3 & 0 & & \\ 0 & 0 & \Gamma_0 & & \\ \hline & & & 1 & 1/2 \\ 0 & & & 1/2 & 1/3 \end{array} \right] = \left[\begin{array}{c|c} \Omega_{0,11} & 0 \\ \hline 0 & \Omega_{0,22} \end{array} \right] = \Omega_0,$$

where Γ_0 is the expectation of the second moment of $[z'_{t-1}, \cdots, z'_{t-p-1}]'$ and $\Omega_{0,11}$ is the first $n(p+1)+2$ square block of Ω_0. On the other hand, we have, from the FCLT,

$$\sum_{t=1}^{T} G_0^{*-1}(x_t^* \otimes u_t) \xrightarrow{d} (\Omega_0 \otimes \Sigma_u)^{1/2} B_0(r),$$

where $B_0(r)$ is an $n^2(p+1)+4n$-dimensional standard Brownian motion. Since Ω_0 is block diagonal, we have

$$(G_0^* \otimes I_n)(\hat{\theta}^* - \theta^*) \xrightarrow{d} (\Omega_0^{-1} \otimes \Sigma_u)^{1/2} B_0(r) \sim \begin{bmatrix} N\left(0, \Omega_{0,11}^{-1} \otimes \Sigma_u\right) \\ N\left(0, \Omega_{0,22}^{-1} \otimes \Sigma_u\right) \end{bmatrix}. \tag{6}$$

In the cointegrated case under Assumption A2(b), the matrices F and D are defined as

$$F = F_1 = \begin{bmatrix} -\mu_1/T & 0 \\ \vdots & \vdots \\ -\mu_1/T & 0 \\ -\beta'\mu_0 & -\beta'\mu_1 \\ -\beta'_\perp\mu_0 & -\beta'_\perp\mu_1 \end{bmatrix} \quad \text{and} \quad D = D_1 = \begin{bmatrix} I_n & -I_n & & & 0 \\ & I_n & -I_n & & \\ & & \ddots & \ddots & \\ & & & I_n & -I_n \\ 0 & & & & \begin{bmatrix} \beta' \\ \beta'_\perp \end{bmatrix}, \end{bmatrix},$$

so that $x_t^* = [(1/T^{\lambda_0})\varepsilon_{0t},\ (t/T^{1+\lambda_1})\varepsilon_{1t}, \Delta z'_{t-1}, \cdots, \Delta z'_{t-p}, (\beta' z_{t-p-1})', (\beta'_\perp z_{t-p-1})', \eta_{0t}, \eta_{1t}]'$. Note that in this case, Δz_{t-j}s and $\beta' z_{t-p-1}$ are zero mean stationary processes while $\beta'_\perp z_{t-p-1}$ is a zero mean purely integrated process. By letting $G^* = G_1^* = diag\{T^{1/2-\lambda_0}, T^{1/2-\lambda_1}, \sqrt{T} I_{np+r}, T I_{n-r}, \sqrt{T} I_2\}$, we have from the WLLN and the FCLT that

$$\sum_{t=1}^{T} G_1^{*-1} x_t^* x_t^{*\prime} G_1^{*-1} \xrightarrow{p} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & \\ 0 & 0 & \Gamma_1 & \\ \hline 0 & & & \int_0^1 \tilde{B}(r)\tilde{B}(r)'dr \end{array}\right]$$
$$= \left[\begin{array}{c|c} \Omega_{1,11} & 0 \\ \hline 0 & \int_0^1 \tilde{B}(r)\tilde{B}(r)'dr \end{array}\right] = \Omega_1,$$

where Γ_1 is the expectation of the second moment of $[\Delta z'_{t-1}, \cdots, \Delta z'_{t-p}, (\beta' z_{t-p-1})']'$, $\Omega_{1,11}$ is the first $np+r+2$ square block of Ω_1, and $\tilde{B}(r) = [B_\perp(r)', 1, r]'$ with $B_\perp(r)$ being an $(n-r)$-dimensional Brownian motion induced by $\beta'_\perp z_{t-p-1}$. On the other hand, we have

$$\sum_{t=1}^{T} G_1^{*-1}(x_t^* \otimes u_t) \xrightarrow{d} \begin{bmatrix} (\Omega_{1,11} \otimes \Sigma_u)^{1/2} B_1(r) \\ \int_0^1 \tilde{B}(r) \otimes dB_u(r) \end{bmatrix},$$

where $B_1(r)$ is an $n(np+r+2)$-dimensional standard Brownian motion while $B_u(r)$ is an n-dimensional Brownian motion induced by the partial sums of u_t. Again, since Ω_1 is block diagonal, we can see that

$$(G_1^* \otimes I_n)(\hat{\theta}^* - \theta^*) \xrightarrow{d} \begin{bmatrix} (\Omega_{1,11}^{-1} \otimes \Sigma_u)^{1/2} B_1(r) \\ \left(\int_0^1 \tilde{B}(r)\tilde{B}(r)' dr \otimes I_n\right)^{-1} \left(\int_0^1 \tilde{B}(r) \otimes dB_u(r)\right) \end{bmatrix}. \tag{7}$$

From (6) and (7), we can see that the limiting distribution of $\hat{\theta}^*$ changes depending on Assumptions A2(a) and (b): if y_t is (trend) stationary, then the estimators of all the coefficients are asymptotically normal, whereas for the I(1) case, only the first $n(np+r+2)$ elements of $\hat{\theta}^*$ have the asymptotic normality and the rest have the non-normal distribution. However, the following theorem shows that the Wald test statistic W_T is asymptotically chi-square distributed irrespective of whether y_t is stationary, integrated or cointegrated. Intuitively, this is because the original parameters of interest are expressed as a linear transformation of the first $n\{n(p+1)+2\}$ or the first $n(np+r+2)$ elements of $\hat{\theta}^*$ plus the negligible term depending on Assumptions A2(a) and (b), respectively, those of which are asymptotically normal.

Theorem 1 *Under Assumptions A1 and A2, if the restrictions $R\theta = q$ hold, then $W_T \xrightarrow{d} \chi^2_k$.*

Using theorem 1, we can make a statistical inference about the original coefficients irrespective of the order of integration/cointegration. This is a major advantage over the exiting methods because they usually need the knowledge of the I(0)/I(1) properties of the time series for statistical inference. On the other hand, our estimator is not efficient because the convergence orders are partially reduced as compared to the standard methods, as can be seen from (6) and (7). However, our new method is useful in practical analyses when we want to avoid the pre-test bias and also when we want to test for the parameter restrictions before determining the order of integration/cointegration, as explained in the next section.

In order to apply our method in practice, we need to determine the values of λ_0 and λ_1. For simplicity, we consider the case where there is no linear trend in the regressors. In this case, the estimator of c_0^*, which corresponds to the first column of Θ^*, is asymptotically normal with the convergence order given by $T^{1/2-\lambda_0}$ in both the I(0) and I(1) cases; see the (1,1) element of G_0^* and G_1^*. In this case, it can be shown that the off-block-diagonal elements of $\sum_{t=1}^{T} G^{*-1} x_t^* x_t^{*\prime} G^{*-1}$ are at most $O_p(1/T^{\lambda_0})$, and then, we can see that

$$T^{1/2-\lambda_0}(\hat{c}_0^* - c_0^*) = \text{the dominating term} + O_p\left(\frac{1}{T^{\lambda_0}}\right).$$

Thus, we may want to set λ_0 as large as (as close to 1/2 as) possible in order to approximate the distribution of the left-hand side by the dominating term on the right-hand side, while we would prefer that λ_0 be small considering the convergence order of $\hat{c}_0^*$. From the viewpoint of statistical testing, the larger values of λ_0 will result in the size of the Wald test being better controlled, while the smaller values of λ_0 will result in the test being more powerful. That is, there is a trade-off between controlling for size and the power of the test, in terms of the values of λ_0. Since the order of the remaining term is $O_p(1/T^{\lambda_0})$ while the local alternative associated with the dominating term is of order $T^{\lambda_0-1/2}$, we propose to set $\lambda_0 = 1/4$ as a compromise to equalize the two orders. In fact, the preliminary simulations show that it is difficult to control the size of the test when λ_0 is close to zero, and that we have the problem of low power for the larger values of λ_0. By the same reason, we also consider $\lambda_1 = 1/4$ in the following sections.

4. Tests for Structural Changes

One of the important applications of the statistical method developed in the previous section is testing for structural changes. As discussed in the Introduction, the knowledge of the order of integration/cointegration is required to test for structural changes in general but we usually do not know whether the time series is stationary, purely integrated, or cointegrated. Hence, it is important and useful in practical analyses to construct structural change tests that are robust to the I(0)/I(1) properties of the time series.

Let us consider the following VAR(p) model with m structural changes for the j-th regime ($j = 1, \cdots, m+1$):

$$y_t = c_{j0} + \Phi_{j1} y_{t-1} + \cdots + \Phi_{jp} y_{t-p} + u_t \quad (t = T_{j-1}+1, \cdots, T_j), \tag{8}$$

where $T_1, \cdots, T_m$ are the break points. Note that we exclude a linear trend from the regression, and as such, we also assume that $c_{j0} = 0$ if y_t is purely integrated in the j-th regime and that c_{j0} is orthogonal to $\alpha_\perp$ if y_t is cointegrated, as explained in Assumption A2(b).

As in the previous section, we estimate (8) by augmenting an extra lag and an artificial regressor. That is, we estimate

$$y_t = c_{j0} + \Phi_{j1} y_{t-1} + \cdots + \Phi_{jp} y_{t-p} + \Phi_{jp+1} y_{t-p-1} + d_{j0}\eta_{0t} + u_t, \tag{9}$$

for $t = T_{j-1}+1, \cdots, T_j$ $(j = 1, \cdots, m+1)$, where $\Phi_{jp+1} = 0$, $d_{j0} = 0$, and η_{0t} is defined as in the previous section. In each regime, we define a matrix H_j in the same way as H, and transform (9) such that

$$\begin{aligned} y_t &= \Theta_j H_j^{-1} H_j x_t + u_t \\ &= \Theta_j^* x_{jt}^* + u_t \\ &= (x_{jt}^{*\prime} \otimes I_n)\theta_j^* + u_t, \end{aligned}$$

for the j-th regime, where $\Theta_j = [c_{j0}, \Phi_{j1}, \cdots, \Phi_{jp}, \Phi_{jp+1}, d_{j0}]$, $\Theta_j^* = \Theta_j H_j^{-1}$, $x_{jt}^* = H_j x_t$, and $\theta_j^* = vec(\Theta_j)$. In this case, the full set of parameters is expressed as $\Theta = [\Theta_1, \Theta_2, \cdots, \Theta_{m+1}]$, $\theta = [\theta_1', \theta_2', \cdots, \theta_{m+1}']'$, $\Theta^* = [\Theta_1^*, \Theta_2^*, \cdots, \Theta_{m+1}^*]$, and $\theta^* = [\theta_1^{*\prime}, \theta_2^{*\prime}, \cdots, \theta_{m+1}^{*\prime}]'$, with $\hat{\Theta}$, $\hat{\theta}$, $\hat{\Theta}^*$, and $\hat{\theta}^*$ being the corresponding least squares estimators.

Since we are interested in whether or not structural changes occurred in the original VAR model (8), the null hypothesis is given by

$$\tilde{\Theta}_1 = \tilde{\Theta}_2 = \cdots = \tilde{\Theta}_{m+1} \quad \text{or equivalently,} \quad \tilde{\theta}_1 = \tilde{\theta}_2 = \cdots = \tilde{\theta}_{m+1},$$

where $\tilde{\Theta}_j = [c_{j0}, \Phi_{j1}, \cdots, \Phi_{jp}]$ is the first $n \times (np+1)$ block of Θ_j, and $\tilde{\theta}_j = vec(\tilde{\Theta}_j)$. Then, for a given set of change points $\mathcal{T} = \{T_1, T_2, \cdots, T_m\}$, the Wald test statistic becomes

$$W_T(\mathcal{T}) = (R\hat{\theta})' \left[R \left(\hat{\Sigma}^{-1} \otimes \hat{\Sigma}_u \right) R' \right]^{-1} (R\hat{\theta}),$$

where $\hat{\Sigma}^{-1} = diag\{\hat{\Sigma}_1^{-1}, \hat{\Sigma}_2^{-1}, \cdots, \hat{\Sigma}_{m+1}^{-1}\}$ with $\hat{\Sigma}_j = \sum_{t=T_{j-1}+1}^{T_j} x_t x_t'$ for $j = 1, \cdots, m+1$, $\hat{\Sigma}_u$ is defined as in the previous section, and $R = R_R' \otimes I_n$ with

$$R_R = \begin{bmatrix} \tilde{R}_R & & 0 \\ -\tilde{R}_R & \ddots & \\ & \ddots & \tilde{R}_R \\ 0 & & -\tilde{R}_R \end{bmatrix} \quad \text{where} \quad \tilde{R}_R = \begin{bmatrix} I_{np+1} \\ 0 \end{bmatrix}.$$

Because we usually do not know the break points, we consider the sup-type test proposed by Andrews (1993) and Bai and Perron (1998).

$$\text{Sup-}F_T(m) = \sup_{\mathcal{T}\in\mathcal{T}_\epsilon} \frac{1}{mn(np+1)} W_T(\mathcal{T}),$$

where $\mathcal{T}_\epsilon = \{(\tau_1, \cdots, \tau_{m+1});\ |\tau_{j+1} - \tau_j| \geq \epsilon \quad (j = 0, \cdots, m)\}$ for a given $\epsilon > 0$ with $\tau_j = T_j/T$, $T_0 = 0$, and $T_{m+1} = T$ as per convention. Because $\hat{\Sigma}$ is asymptotically degenerate, we need to consider the transformation of the regression model to derive the limiting distribution of the above test statistic. However, we do not have to explain the derivation of the limiting distribution in detail because as shown in the proof of Theorem 1, we know that $W_T(\mathcal{T})$ is asymptotically equivalent to a test for linear restrictions on a part of $\tilde{\theta}^*$ associated with the transformed stationary variables, whose limiting distributions are expressed as a linear transformation of a standard Brownian motion as given in (6) and (7). Thus, following Theorem 6 in Bai and Perron (1998) and Kurozumi (2011), we have the following theorem.

Theorem 2 *Under the null hypothesis of no structural change,*

$$\textit{Sup-}F_T(m) \xrightarrow{d} \sup_{\mathcal{T}\in\mathcal{T}_\epsilon} \frac{1}{m(n^2p+n)} W(m) \quad \textit{where} \quad W(m) = \sum_{j=1}^{m} \frac{\|\tau_j W(\tau_{j+1}) - \tau_{j+1} W(\tau_j)\|^2}{\tau_j \tau_{j+1}(\tau_{j+1} - \tau_j)}.$$

The critical values of the above distributions are given in Bai and Perron (1998, 2003). Theorem 2 is concerned with the sup-type test but we can easily see that the exponential-type and the averaging-type tests by Andrews, Lee and Ploberger (1996) are available. For example, in a special case of one time change, these test statistics are given by

$$\text{Exp-}W_T(1) = \log\left[\frac{1}{(\bar{\tau} - \underline{\tau})T} \sum_{t=\underline{\tau}T+1}^{\bar{\tau}T} \exp\left(\frac{1}{2} W_T(\mathcal{T})\right)\right], \quad \text{Avg-}W_T(1) = \frac{1}{(\bar{\tau} - \underline{\tau})T} \sum_{t=\underline{\tau}T+1}^{\bar{\tau}T} W_T(\mathcal{T}),$$

where $\underline{\tau}$ and $\bar{\tau}$ are the lower and the upper bounds of the possible break fraction τ. The critical values given in Andrews and Ploberger (1994) are applicable in this case.

The above tests suppose that the number of structural changes under the alternative can be specified from the outset. However, in some cases, we do not want to prespecify m but just suppose M, the possible maximum number of breaks under the alternative. In such a case, we can construct the double maximum tests proposed by Bai and Perron (1998). The test statistics are given by

$$\text{UDmax-}F_T(M) = \max_{1\le m\le M} \text{Sup-}F_T(m), \quad \text{WDmax-}F_T(M) = \max_{1\le m\le M} \frac{c(k,\alpha,1)}{c(k,\alpha,m)}\text{Sup-}F_T(m),$$

where $c(k,\alpha,m)$ is a critical value of Sup-$F_T(m)$ with significance level α. Again, as in Theorem 2, we can see that the limiting distributions of the above two test statistics are the same as those given by Bai and Perron (1998), and hence, we can use the critical values given in Bai and Perron (1998, 2003).

5. Simulation Results

To demonstrate the performance of the proposed tests in finite samples, we conduct two sets of numerical simulations. The first set is for the investigation of the single break tests, namely Sup-$F(1)$, Exp-$W_T(1)$, and Avg-$W_T(1)$; the second set is for Sup-$F(m)$, UDmax-$F_T(M)$, and WDmax-$F_T(M)$ with two breaks.

5.1 Single break case

We first report the simulation results regarding the size and power properties of the structural break tests dealing with a single break. We consider, as a data generating process, the following AR(1) process with a single break in a constant as a leading case:

DGP1:

$$y_t = \begin{cases} c_{10} + \phi y_{t-1} + \varepsilon_t. & \text{for } t = 1, ..., T_1 \\ c_{20} + \phi y_{t-1} + \varepsilon_t & \text{for } t = T_1 + 1, ..., T, \end{cases} \tag{10}$$

where ε_t is *i.i.d.* $N(0,1)$ with initial value $y_0 = 0$.

The null rejection probabilities are simulated for the grid values of ϕ covering the range $[0,1]$ with increments of 0.05. The parameters c_{10} and c_{20} are set to zero because the test statistics are invariant to their values. The sample sizes used are T = 50, 100, 150, and 200, with $5,000$ replications for each. The trimming parameter is set at $\epsilon = 0.15$, so that the admissible range of change point is $[0.15, 0.85]$. Figure 1 plots the finite sample sizes of Sup-$F(1)$, Exp-$W_T(1)$, and Avg-$W_T(1)$; in all the cases, the rejection frequencies are calculated at the nominal 5% level. We can see from the figure that the empirical sizes of all the tests are close to the nominal size for $\phi < 0.9$. On the other hand, when ϕ is close to one, the sup-type test slightly overly rejects the null hypothesis whereas the averaging-type test tends to be conservative. The empirical size of the sup-type test is not much affected by the true values of ϕ, and is close to the nominal size in all the cases.

To investigate the finite sample power properties, we consider DGP (10) with a break at the half of the sample. The sample sizes of $T = 100$ and $T = 200$ are used for the power simulations, with four distinct configurations for each. These configurations correspond to the stationary ($\phi = 0.5$), the moderate deviation from a unit root ($\phi = 0.9$), the near unit root ($\phi = 0.95$), and the unit root ($\phi = 1$) cases. The data are generated accordingly, for a grid of values for the magnitude of the break, $\delta = c_{20} - c_{10}$, covering the range $[0, 2]$ in steps of 0.05.

From Figures 1 and 2, we can see that the powers increase as the magnitude of the break or the sample size increases. We can also see that these powers are sensitive to the values of ϕ; the power of each test statistic is uniformly higher for the larger values of ϕ. Among the three tests, Sup-$F(1)$ and Exp-$W_T(1)$ show similar finite sample properties and generally have higher power than Avg-$W_T(1)$, which is consistent with the theoretical investigation by Kim and Perron (2009). In particular, Avg-$W_T(1)$ has poor finite sample properties with regard to power when $\phi = 1$. Considering the size and the power properties, in our approach, the sup-type test seems most reliable for the single break case.

5.2 Two breaks case

To assess the finite sample properties of Sup-$F_T(m)$ for $m \geq 2$ and the double maximum tests, we use the following data generating process, in addition to DGP1:

DGP2:

$$\begin{aligned}
y_t &= c_{10} + \phi y_{t-1} + u_t, \qquad t = 1, \cdots, T_1 \\
y_t &= c_{20} + \phi y_{t-1} + u_t, \qquad t = T_1 + 1, \cdots, T_2 \\
y_t &= c_{30} + \phi y_{t-1} + u_t, \qquad t = T_2 + 1, \cdots, T
\end{aligned} \tag{11}$$

where $y_0 = 0$ and $u_t \sim i.i.d.N(0, 1)$. Again, the trimming parameter ϵ is set at 0.15 while the maximum number of allowable breaks is set at $M = 3$.

Table 1 reports the simulation results for the sizes of Sup-$F_T(m)$, UDmax-$F_T(M)$, and WDmax-$F_T(M)$. We can see from the table that the empirical sizes of Sup-$F_T(m)$ with $m = 1$ and 2 and UDmax-F_T are relatively close to the nominal sizes whereas Sup-$F_T(3)$ is oversized. As in the first simulation, all the tests tend to overly reject the null hypothesis when ϕ is close to one, although the overall size distortion is mitigated as the sample size increases.

We then investigate the empirical powers of the tests in the presence of a single break (DGP1) and two breaks (DGP2). We first consider the case of one break in an intercept occurring at the half of the sample, that is, $\lambda_0 = 0.5$. Table 2 reports the results for three values of break magnitudes $\delta = c_{20} - c_{10} = 0.5, 0.7, 1.0$ for each specification. As can be seen from Table 2, the power (expectedly) increases as the break magnitude increases. We can also see that Sup-$F_T(3)$ has the highest power in all the specifications, although this may be because of the oversize distortions under the null hypothesis. The other tests have similar finite sample properties with regard to power.

Finally, we consider the case where there exist two breaks in an intercept occurring at the locations $(\tau_1, \tau_2) = (1/3, 2/3)$ and $(0.4, 0.6)$. We generate DGP2 with break magnitudes satisfying $\delta = c_{20} - c_{10} = c_{30} - c_{20}$ for $T = 100$. From Table 3, we can see that the overall properties are preserved in this case. We obtained a similar result for $T = 200$, but do not report it to save space.

6. Conclusion

In this paper, we proposed an estimation method for the VAR processes that are (trend) stationary, purely integrated, or cointegrated. Our approach is based on the lag augmented method by Toda and Yamamoto (1995) and the ART-VAR approach by Yamamoto (1996). Although our method is not necessarily efficient, it has an advantage over the existing methods in that the Wald test statistic for linear restrictions weakly converges to a chi-square distribution irrespective of whether the time series is stationary, integrated, or cointegrated; thus, we can make a statistical inference about the coefficients without the knowledge of the order of integration/cointegration. The useful application of our method is testing for structural changes. We show that the sup-type, the exponential-type, and the averaging-type tests, and the $UDmax$ and $WDmax$ tests based on our estimators are available using the existing tables of critical values. Our approach can be seen as a benchmark for a robust statistical inference about the VAR models, and would prove useful to develop more efficient methods in the future.

Appendix

Proof of Theorem 1: Let $G = diag\{T^{1/2-\lambda_0}, T^{1/2-\lambda_1}, \sqrt{T} I_{k_2-2}\}$ be a $k_2 \times k_2$ scaling matrix. Then, because $\theta^* = (H'^{-1} \otimes I_n)\theta$, $\hat{\theta}^* = (H'^{-1} \otimes I_n)\hat{\theta}$, and $x_t^* = Hx_t$, the Wald test statistic (5) can be expressed as

$$W_T = \left[(G \otimes I_{k_1})R^*(\hat{\theta}^* - \theta^*)\right]' \left[(G \otimes I_{k_1})R^* \left\{ \left(\sum_{t=1}^{T} x_t^* x_t^{*\prime}\right)^{-1} \otimes \hat{\Sigma}_u \right\} R^{*\prime}(G \otimes I_{k_1})\right]^{-1} \left[(G \otimes I_{k_1})R^*(\hat{\theta}^* - \theta^*)\right], \tag{12}$$

where $R^* = R(H' \otimes I_n) = R_R' H' \otimes R_L$. In the stationary case under Assumption A2(a), noting that H is defined using $F = F_0$ and $D = D_0 = I_{n(p+1)}$ and that the last $n+2$ rows of R_R are all zeroes, we can see that

$$\begin{aligned} (G \otimes I_{k_1})R^* &= GR_R' H' \otimes R_L \\ &= R_R' G_0^* H' \otimes R_L \\ &= R_R'(\tilde{H}_0' + O(T^{-\underline{\lambda}}))G_0^* \otimes R_L, \end{aligned}$$

where $\underline{\lambda} = \min(\lambda_0, \lambda_1)$ and $\tilde{H}_0$ is defined in the same way as H with $F = F_0$ replaced by zeroes. Then, from (6) and the structure of $\tilde{H}_0$, we have

$$\begin{aligned} (G \otimes I_{k_1})R^*(\hat{\theta}^* - \theta^*) &= (R_R' \tilde{H}_0' \otimes R_L)(G_0^* \otimes I_n)(\hat{\theta}^* - \theta^*) + o_p(1) \\ &\xrightarrow{d} N\left(0,\ R_{R,1}' \Omega_{0,11}^{11} R_{R,1} \otimes R_L \Sigma_u R_L'\right), \end{aligned} \tag{13}$$

where $\Omega_{0,11}^{11}$ is the first $np+2$ square block of $\Omega_{0,11}^{-1}$ and

$$\begin{aligned}
& (G \otimes I_{k_1}) R^* \left\{ \left(\sum_{t=1}^{T} x_t^* x_t^{*\prime} \right)^{-1} \otimes \hat{\Sigma}_u \right\} R^{*\prime} (G \otimes I_{k_1}) \\
= & \left\{ R_R' \tilde{H}_0' \left(\sum_{t=1}^{T} G_0^{*-1} x_t^* x_t^{*\prime} G_0^{*-1} \right)^{-1} \tilde{H}_0 R_R \right\} \otimes R_L \hat{\Sigma}_u R_L + o_p(1) \\
\xrightarrow{p} & R_{R,1}' \Omega_{0,11}^{11} R_{R,1} \otimes R_L \Sigma_u R_L'. \qquad (14)
\end{aligned}$$

Thus, from (13) and (14), we can see that W_T weakly converges to a chi-square distribution with $k = k_1 \times k_2$ degrees of freedom.

Similarly, in the cointegrated case under Assumption A2(b), we can see that

$$(G \otimes I_{k_1}) R^* = R_R' (\tilde{H}_1' + O(T^{-\lambda})) G_1^*,$$

where $\tilde{H}_1$ is defined in the same way as H with $F = F_1$ replaced by zeroes. Then, from (7), we have

$$\begin{aligned}
(G \otimes I_{k_1}) R^* (\hat{\theta}^* - \theta^*) &= (R_R' \tilde{H}_1' \otimes R_L)(G_1^* \otimes I_n)(\hat{\theta}^* - \theta^*) + o_p(1) \\
&\xrightarrow{d} N = \left(0,\ R_{R,1}' \tilde{H}_{11}' \Omega_{1,11}^{-1} \tilde{H}_{11} R_{R,1} \otimes R_L \Sigma_u R_L' \right). \qquad (15)
\end{aligned}$$

where $\tilde{H}_{11}$ is the first $(np+2)$ square block of $\tilde{H}_1$ and

$$\begin{aligned}
& (G \otimes I_{k_1}) R^* \left\{ \left(\sum_{t=1}^{T} x_t^* x_t^{*\prime} \right)^{-1} \otimes \hat{\Sigma}_u \right\} R^{*\prime} (G \otimes I_{k_1}) \\
= & \left\{ R_R' \tilde{H}_1' \left(\sum_{t=1}^{T} G_1^{*-1} x_t^* x_t^{*\prime} G_1^{*-1} \right)^{-1} \tilde{H}_1 R_R \right\} \otimes R_L \Sigma_u R_L + o_p(1) \\
\xrightarrow{p} & R_{R,1}' \tilde{H}_{11}' \Omega_{1,11}^{-1} \tilde{H}_{11} R_{R,1} \otimes R_L \Sigma_u R_L'. \qquad (16)
\end{aligned}$$

Then, from (15) and (16), we obtain the desired weak convergence.■

References

[1] Ahn, S. K. and G. C. Reinsel (1990). Estimation for Partially Nonstationary Multivariate Autoregressive Models. *Journal of the American Statistical Association* 85, 813-823.

[2] Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* 61, 821-856 (Erratum, 71, 395-397).

[3] Andrews, D. W. K., I. Lee and W. Ploberger (1996). Optimal Changepoint Tests for Normal Linear Regression. *Journal of Econometrics* 70, 9-38.

[4] Andrews, D. W. K. and W. Ploberger (1994). Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative, *Econometrica* 62, 1383-1414.

[5] Bai, J., R. L. Lumsdaine and J. H. Stock (1998). Testing For and Dating Common Breaks in Multivariate Time Series. *Review of Economic Studies* 65, 395-432.

[6] Bai, J. and P. Perron (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66, 47-78.

[7] Bai, J. and P. Perron (2003). Critical Values for Multiple Structural Change Tests. *Econometrics Journal* 6, 72-78.

[8] Dickey, D. A. and W. A. Fuller (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association* 74, 427-431.

[9] Hansen, H. and S. Johansen (1999). Some Tests for Parameter Constancy in Cointegrated VAR-Models. *Econometrics Journal* 2, 306-333.

[10] Hansen, P. R. (2003). Structural Changes in the Cointegrated Vector Autoregressive Model. *Journal of Econometrics* 114, 261-295.

[11] Harvey, D., S. J. Leybourne and A. M. R. Taylor (2009). Simple, Robust and Powerful Tests of the Breaking Trend Hypothesis. *Econometric Theory* 25, 995-1029.

[12] Harvey, D., S. J. Leybourne and A. M. R. Taylor (2010). Robust Methods for Detecting Multiple Level Breaks in Autocorrelated Time Series. *Journal of Econometrics* 157, 342-358.

[13] Hsiao, C. and S. Wang (2006). Modified Two-Stage Least-Squares Estimators for the Estimation of a Structural Vector Autoregressive Integrated Process. *Journal of Econometrics* 135, 427-463.

[14] Hsiao, C. and S. Wang (2007). Lag-Augmented Two- and Three-Stage Least Squares Estimators for Integrated Structural Dynamic Models. *Econometrics Journal* 10, 49-81.

[15] Inoue, A. (1999). Tests of Cointegrating Rank with a Trend-Break. *Journal of Econometrics* 90, 215-237.

[16] Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* 12, 231-254.

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

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

[19] Kejriwal, M. and P. Perron (2008). The Limiting Distribution of the Estimates in Cointegrated Regression Models with Multiple Structural Changes. *Journal of Econometrics* 146, 59-73.

[20] Kejriwal, M. and P. Perron (2010a). A Sequential Procedure to Determine the Number of Breaks in Trend with an Integrated or Stationary Noise Component. *Journal of Time Series Analysis* 31, 305-328.

[21] Kejriwal, M. and P. Perron (2010b). Testing for Multiple Structural Changes in Cointegrated Regression Models. *Journal of Business and Economic Statistics* 28, 503-522.

[22] Kim, D. and P. Perron (2009). Assessing the Relative Power of Structural Break Tests Using a Framework Based on the Approximate Bahadur Slope. *Journal of Econometrics* 149, 26-51.

[23] Kurozumi, E. (2011). Tests for Multiple Structural Changes in the Nonlinear/GMM Framework. Manuscript, Hitotsubashi University.

[24] Kurozumi, E. and T. Yamamoto (2000). Modified Lag Augmented Vector Autoregressions. *Econometric Reviews* 19, 207-231.

[25] Lütkepohl, H., P. Saikkonen and C. Trenkler (2003). Comparison of Tests for the Cointegrating Rank of a VAR Process with a Structural Shift. *Journal of Econometrics* 113, 201-229.

[26] Lütkepohl, H., P. Saikkonen and C. Trenkler (2004). Testing for the Cointegrating Rank of a VAR process with Level Shift at Unknown Time. *Econometrica* 72, 647-662.

[27] Perron, P. and T. Yabu (2009). Testing for Shifts in Trend With an Integrated or Stationary Noise Component. *Journal of Business and Economic Statistics* 27, 369-396.

[28] Phillips, P. C. B. (1995). Fully Modified Least Squares and Vector Autoregression. *Econometrica* 63, 1023-1078.

[29] Qu, Z. (2007). Searching for Cointegration in a Dynamic System. *Econometrics Journal* 10, 580-604.

[30] Qu, Z. and P. Perron (2007). Estimating and Testing Structural Changes in Multivariate Regressions. *Econometrica* 75, 459-502.

[31] Quintos, C. E. (1995). Sustainability of the Deficit Process with Structural Shifts. *Journal of Business and Economic Statistics* 13, 409-417.

[32] Quintos, C. E. (1997). Stability Tests in Error Correction Models. *Journal of Econometrics* 82, 289-315.

[33] Said, E. S. and D. A. Dickey (1984). Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika* 71, 599-607.

[34] Saikkonen, P. and H. Lütkepohl (2000). Testing for the Cointegrating Rank of a VAR Process with Structural Shifts. *Journal of Business and Economic Statistics* 18, 451-464.

[35] Seo, B. (1998). Tests for Structural Change in Cointegrated Systems. *Econometric Theory* 14, 222-259.

[36] Toda, H. Y., and T. Yamamoto (1995). Statistical Inference in Vector Autoregressions with Possibly Integrated Processes. *Journal of Econometrics* 66, 225-250.

[37] Yamamoto, T. (1996). A Simple Approach to the Statistical Inference in Linear Time Series Models Which May Have Some Unit Roots. *Hitotsubashi Journal of Economics* 37, 87-100.

Table 1: Finite Sample Sizes of Sup-F_T, UDmax-F_T and WDmax-F_T

	specification	Sup $m=1$	Sup $m=2$	Sup $m=3$	UDmax	WDmax
$T=100$	$\phi=0.5$	0.06	0.08	0.11	0.07	0.08
	$\phi=0.95$	0.07	0.08	0.11	0.08	0.09
	$\phi=1.0$	0.08	0.07	0.13	0.09	0.10
$T=200$	$\phi=0.5$	0.05	0.07	0.09	0.06	0.07
	$\phi=0.95$	0.06	0.07	0.09	0.07	0.09
	$\phi=1.0$	0.08	0.07	0.11	0.08	0.09

Table 2: Finite Sample Powers of Sup-F_T, UDmax-F_T and WDmax-F_T (Single Break)

			Sup			WDmax	WDmax
		specification	$m=1$	$m=2$	$m=3$		
$T=100$	$\phi=0.5$	$\delta=0.5$	0.54	0.49	0.52	0.54	0.51
		$\delta=0.7$	0.76	0.71	0.71	0.77	0.74
		$\delta=1.0$	0.97	0.91	0.90	0.97	0.95
	$\phi=0.95$	$\delta=0.5$	0.72	0.84	0.92	0.83	0.90
		$\delta=0.7$	0.80	0.88	0.95	0.88	0.92
		$\delta=1.0$	0.86	0.93	0.97	0.93	0.96
	$\phi=1.0$	$\delta=0.5$	0.65	0.73	0.84	0.71	0.76
		$\delta=0.7$	0.68	0.71	0.79	0.72	0.76
		$\delta=1.0$	0.86	0.81	0.85	0.86	0.88
$T=200$	$\phi=0.5$	$\delta=0.5$	1.0	0.99	0.98	1.0	0.95
		$\delta=0.7$	1.0	0.98	0.98	1.0	0.92
		$\delta=1.0$	1.0	1.0	1.0	1.0	1.0
	$\phi=0.95$	$\delta=0.5$	0.75	0.80	0.93	0.88	0.90
		$\delta=0.7$	0.84	0.92	0.93	0.90	0.95
		$\delta=1.0$	0.97	0.97	1.0	0.98	1.0
	$\phi=1.0$	$\delta=0.5$	0.71	0.73	0.78	0.71	0.76
		$\delta=0.7$	0.73	0.76	0.82	0.78	0.81
		$\delta=1.0$	0.88	0.87	0.86	0.90	0.89

Table 3: Finite Sample Powers of Sup-F_T, UDmax-F_T and WDmax-F_T (Two Breaks, $T = 100$)

			Sup			WDmax	WDmax
(τ_1, τ_2)		specification	$m = 1$	$m = 2$	$m = 3$		
$(1/3, 2/3)$	$\phi = 0.5$	$\delta = 0.5$	0.87	0.79	0.82	0.86	0.85
	$\phi = 1.0$	$\delta = 0.5$	0.51	0.56	0.68	0.55	0.62
$(0.4, 0.6)$	$\phi = 0.5$	$\delta = 0.5$	0.93	0.88	0.87	0.93	0.91
	$\phi = 1.0$	$\delta = 0.5$	0.66	0.66	0.74	0.68	0.73

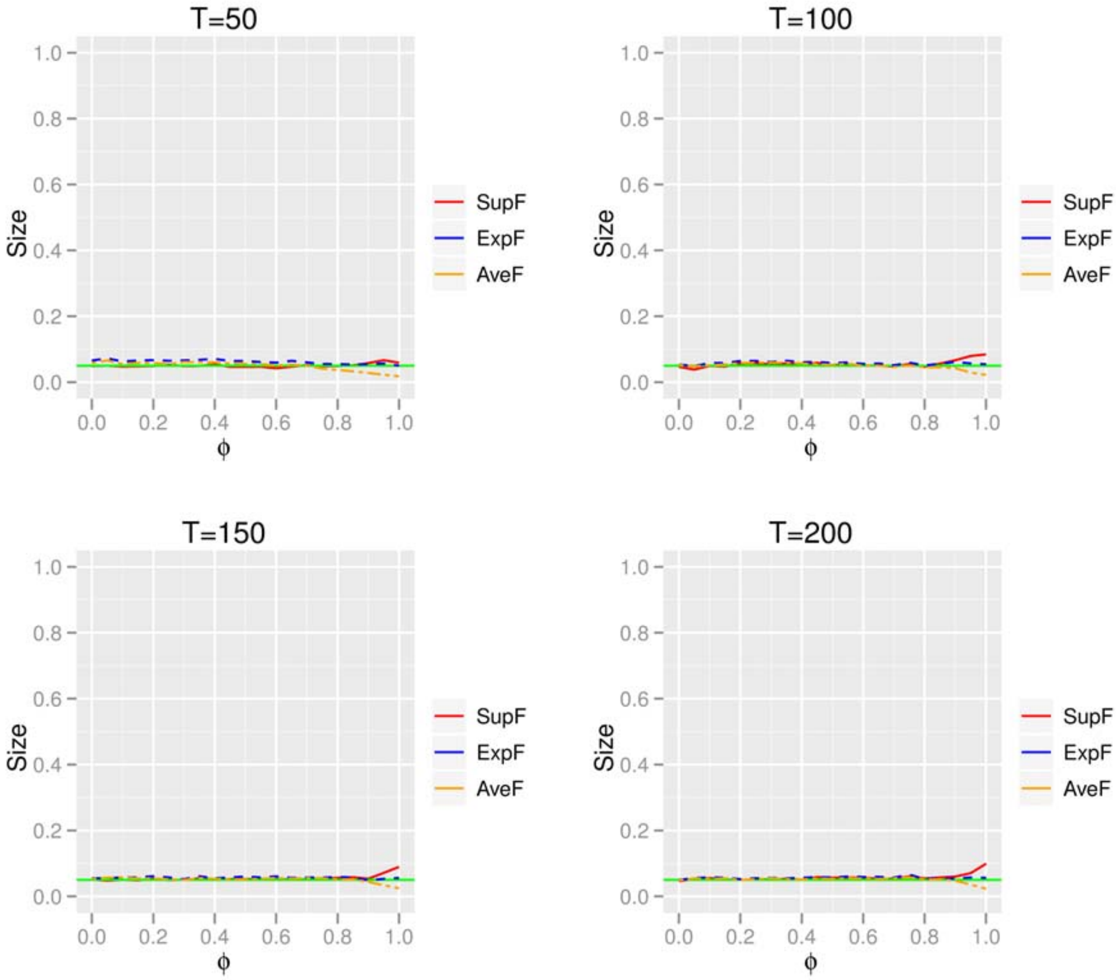

Figure 1: Finite sample size comparisons at 5% level

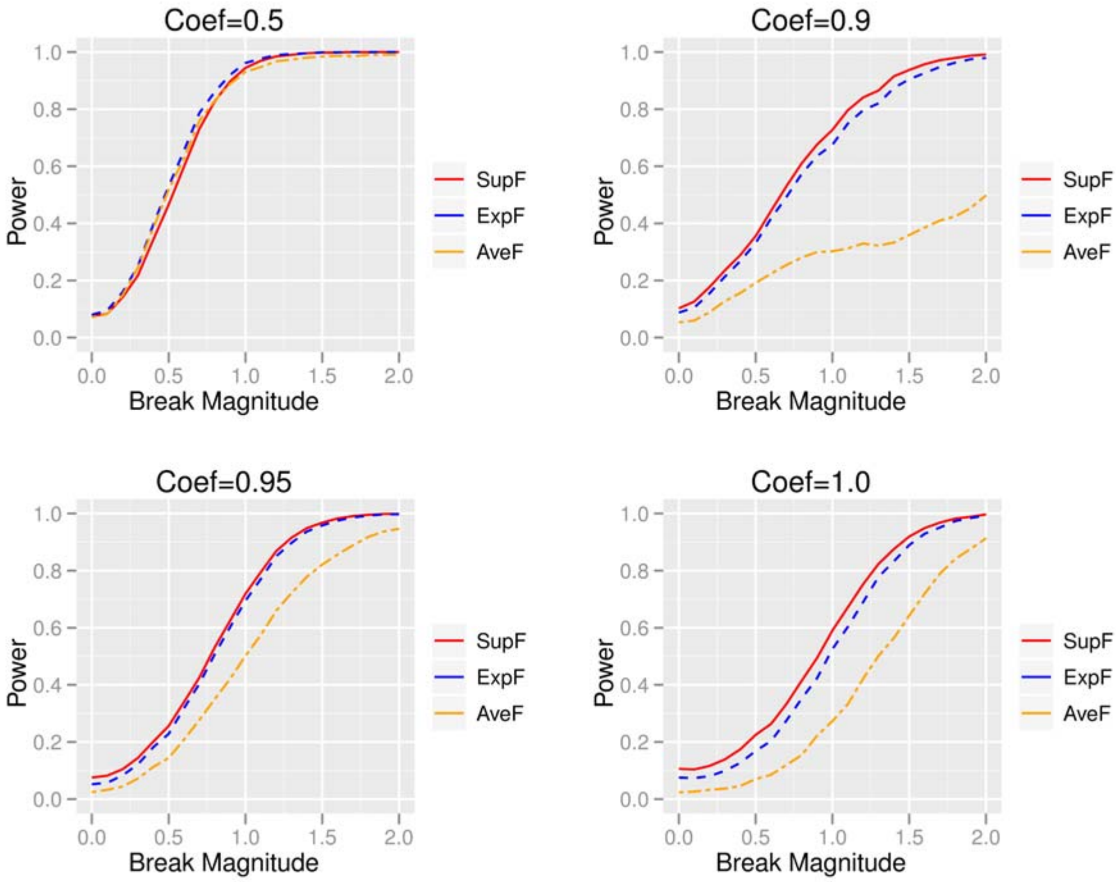

Figure 2: Finite sample power comparisons at 5% level. Sample size is $T = 100$.

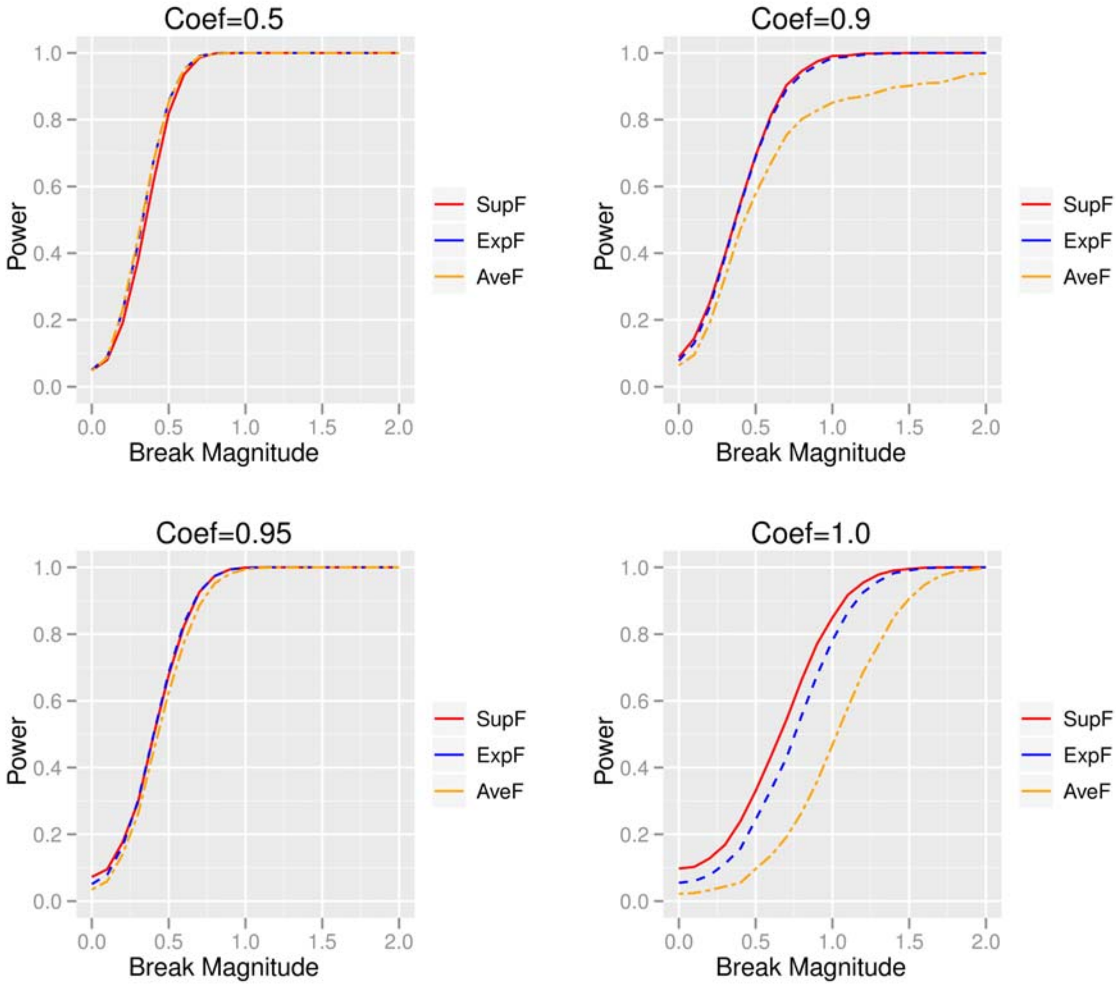

Figure 3: Finite sample power comparisons at 5% level. Sample size is $T = 200$.