Efficient Score Test for Change Detection in Vector Autoregressive Models

Marek Dvořák[1]

Abstract. The statistics and econometrics articles both contain a vast amount of testing procedures related to structural changes with unknown break dates, most of them designed for the univariate models. Work on detecting change-points in multivariate models is receiving an increasing attention nowadays. Many of the test statistics are based on likelihood ratio, see for instance paper by [5]. We will discuss the efficient score-based test for change detection in multivariate AR models. Its asymptotic properties and results from the simulation study will be presented. The idea of such methodology came from [3].

Keywords: Change point analysis, Invariance principle, VAR

JEL classification: C12, C32
AMS classification: primary 62G20; secondary 60F17, 62M10

1 Introduction

1.1 The model and its assumptions

Let us consider the following $\mathbb{VAR}(p)$ model of the form

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad t \in \mathbb{Z}, \tag{1}$$

where $\mathbf{c} \in \mathbb{R}^n$ is a vector of constants, $\mathbf{\Phi}_j \in \mathbb{R}^{n\times n}$, $j = 1,\ldots,p$ are non random matrices of autoregressive coefficients and $\{\boldsymbol{\varepsilon}_t\}_t$ is an n-dimensional noise process specified further. Let us assume that we have T consecutive observations $\mathbf{y}_1,\ldots,\mathbf{y}_T$ drawn from the model (1). Our objective is to test the null hypothesis H_0 against H_1, where

$$\begin{array}{lrl} H_0: & & \mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t,\ t = 1,\ldots,T, \\ H_1: & \exists k \in \{1,\ldots,T-1\}: & \mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t,\ t = 1,\ldots,k, \\ & & \quad = \widetilde{\mathbf{c}} + \widetilde{\mathbf{\Phi}}_1\mathbf{y}_{t-1} + \ldots + \widetilde{\mathbf{\Phi}}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t,\ t = k+1,\ldots,T \end{array} \tag{2}$$

and either $\mathbf{c} \neq \widetilde{\mathbf{c}}$ or $\mathbf{\Phi}_j \neq \widetilde{\mathbf{\Phi}}_j$ for at least one $j = 1,\ldots,p$. We will not assume any changes in the variance structure of the autoregressive model.

Let us denote $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ an n dimensional identity matrix. We define an indicator $\mathrm{I}_{[s=t]}$ with the values 1 for $s = t$ and 0 otherwise. Throughout the paper, vec operator is considered as a matrix function that stacks column vectors of a matrix below one another.

Let us formulate the following assumptions:

Assumptions:

(A.1) roots of the polynomial $\det\{\mathbf{I}_n - \mathbf{\Phi}_1 z - \ldots - \mathbf{\Phi}_p z^p\}$ lie outside the complex unit circle,

(A.2) white noise process $\boldsymbol{\varepsilon}_t$ is i.i.d. Gaussian, $\mathsf{E}[\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_t^\top] = \mathrm{I}_{[s=t]}\mathbf{\Omega}$ and $\mathbf{\Omega}$ is a positive-definite matrix,

[1]Charles University, Faculty of Mathematics and Physics/Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague, Czech Republic, e-mail: dvorakm@karlin.mff.cuni.cz

(A.3) vector $\mathbf{v}_1 = \mathsf{vec}(\mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{1-p})$ of initial observations satisfies

$$\mathbf{v}_1 = \boldsymbol{\mu}_{\boldsymbol{np}} + \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{u}_{-k},$$

where $\boldsymbol{\mu}_{\boldsymbol{np}} := \mathsf{vec}(\boldsymbol{\mu}, \boldsymbol{\mu}, \dots, \boldsymbol{\mu}) \in \mathbb{R}^{np}$, $\mathbf{u}_t := \mathsf{vec}(\boldsymbol{\varepsilon}_t, \mathbf{0}, \dots, \mathbf{0}) \in \mathbb{R}^{np}$ and

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\Phi}_1 \ \cdots \ \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \mathbf{I}_{p-1} \otimes \mathbf{I}_n & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{np \times np}.$$

Assumption (A.1) guarantees stationarity of the $\mathbb{VAR}(p)$ model. Assumption (A.2) specifies the properties of the error term $\boldsymbol{\varepsilon}_t$. The last Assumption (A.3) ensures that the initial observations of the model come from the same stationary representation as the observations $\mathbf{y}_1, \dots, \mathbf{y}_k$.

1.2 The test statistic

We start with the construction of the efficient vector test statistic that will be derived from the likelihood function under the assumptions above. Let us denote $\boldsymbol{\phi} := \mathsf{vec}\,(\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_p) \in \mathbb{R}^{n^2 p}$, $\boldsymbol{\theta} := (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top)^\top \in \mathbb{R}^{n(np+1)}$, $\mathbf{Y}_k := \mathsf{vec}(\mathbf{y}_1, \dots, \mathbf{y}_k) \in \mathbb{R}^{nk}$, $\boldsymbol{\mu}_{\boldsymbol{k}} := \mathsf{vec}(\boldsymbol{\mu}, \dots, \boldsymbol{\mu}) \in \mathbb{R}^{nk}$ and

$$\mathbf{X}_k := \begin{pmatrix} \mathbf{y}_0 - \boldsymbol{\mu} & \cdots & \mathbf{y}_{k-1} - \boldsymbol{\mu} \\ \vdots & \ddots & \vdots \\ \mathbf{y}_{1-p} - \boldsymbol{\mu} & \cdots & \mathbf{y}_{k-p} - \boldsymbol{\mu} \end{pmatrix} \in \mathbb{R}^{np \times k}. \tag{3}$$

The conditional log-likelihood function based on k observations $\mathbf{y}_1, \dots, \mathbf{y}_k$ with given $\mathbf{y}_{1-p}, \mathbf{y}_{2-p}, \dots, \mathbf{y}_0$ is of the form

$$\ell_k(\boldsymbol{\mu}, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_p, \boldsymbol{\Omega}) = -\frac{nk}{2} \log(2\pi) - \frac{k}{2} \log|\boldsymbol{\Omega}| - \frac{1}{2} \sum_{t=1}^{k} \boldsymbol{\varepsilon}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \boldsymbol{\mu} - \sum_{j=1}^{p} \boldsymbol{\Phi}_j (\mathbf{y}_{t-j} - \boldsymbol{\mu})$. Partial derivatives of ℓ_k with respect to the unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\phi}$ are of the form

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ell_k(\boldsymbol{\theta}) = \Big(\mathbf{I}_n - \sum_{j=1}^{p} \boldsymbol{\Phi}_j\Big)^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^{k} \boldsymbol{\varepsilon}_t,$$
$$\frac{\partial}{\partial \boldsymbol{\phi}} \ell_k(\boldsymbol{\theta}) = \left(\mathbf{X}_k \otimes \boldsymbol{\Omega}^{-1}\right) (\mathbf{Y}_k - \boldsymbol{\mu}_{\boldsymbol{k}}) - \Big(\mathbf{X}_k \mathbf{X}_k^\top \otimes \boldsymbol{\Omega}^{-1}\Big) \boldsymbol{\phi}$$

and "$\otimes$" denotes the Kronecker product.

The Fisher information matrix $\mathcal{I}_{\mathcal{F}}$ about the parameter vector $\boldsymbol{\theta}$ can be obtained under general regularity conditions as

$$\mathcal{I}_{\mathcal{F}}(\boldsymbol{\theta}) = -\,\mathsf{E}\left[\frac{\partial^2 \ell_k}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^\top}\right],$$

where

$$\frac{\partial^2 \ell_k}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} -k\Big(\mathbf{I}_n - \sum_{j=1}^{p} \boldsymbol{\Phi}_j\Big)^\top \boldsymbol{\Omega}^{-1} \Big(\mathbf{I}_n - \sum_{j=1}^{p} \boldsymbol{\Phi}_j\Big) & \frac{\partial^2 \ell_k}{\partial \boldsymbol{\mu} \partial \boldsymbol{\phi}^\top} \\ \frac{\partial^2 \ell_k}{\partial \boldsymbol{\phi} \partial \boldsymbol{\mu}^\top} & -\mathbf{X}_k \mathbf{X}_k^\top \otimes \boldsymbol{\Omega}^{-1} \end{pmatrix}$$

and the off-diagonal elements can be found in [6], p. 91. The asymptotic information matrix $\mathcal{I}(\boldsymbol{\theta}) = \lim_{k \to \infty} \frac{1}{k} \mathcal{I}_{\mathcal{F}}(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$ exists and is given by

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{pmatrix} \Big(\mathbf{I}_n - \sum_{j=1}^{p} \boldsymbol{\Phi}_j\Big)^\top \boldsymbol{\Omega}^{-1} \Big(\mathbf{I}_n - \sum_{j=1}^{p} \boldsymbol{\Phi}_j\Big) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_y(0) \otimes \boldsymbol{\Omega}^{-1} \end{pmatrix} =: \begin{pmatrix} \mathcal{I}_{1,1}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{2,2}(\boldsymbol{\theta}) \end{pmatrix},$$

see [6], p. 91–92; and $\boldsymbol{\Gamma}_y(0)$ is the variance matrix of $\mathbf{y}_1, \dots, \mathbf{y}_p$.

Let $\widehat{\boldsymbol{\theta}}_T := \big(\widehat{\boldsymbol{\mu}}_T^\top, \widehat{\boldsymbol{\phi}}_T^\top\big)^\top$ be the maximum likelihood (=ML) estimators of the unknown parameters based on the full sample $\mathbf{y}_1, \ldots, \mathbf{y}_T$. The efficient score test statistic is an $r := n(np+1)$-vector of the following form

$$\widehat{\mathbf{B}}(\tau) := \frac{1}{\sqrt{T}} \cdot \mathcal{I}^{-\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_T) \cdot \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\mu}} \ell_{\lfloor T\tau \rfloor}(\widehat{\boldsymbol{\theta}}_T) \\ \frac{\partial}{\partial \boldsymbol{\phi}} \ell_{\lfloor T\tau \rfloor}(\widehat{\boldsymbol{\theta}}_T) \end{pmatrix},$$

where $\lfloor x \rfloor$ is integer part of x and $k = \lfloor T\tau \rfloor$, $0 \le \tau \le 1$.

Throughout the article, we will omit the subscript $_T$ in the ML estimators for notation simplicity. Let us denote $\mathbf{K} = \sum_{j=1}^p \boldsymbol{\Phi}_j$ and its corresponding ML estimate as $\widehat{\mathbf{K}} = \sum_{j=1}^p \widehat{\boldsymbol{\Phi}}_j$.

2 Main Result

The aim of the article will be the proof of the following theorem:

Theorem 1. *Let us suppose that the sequence $\{\mathbf{y}_t\}$ satisfies $\mathbb{VAR}(p)$ model of the form (1) and Assumptions (A.1) – (A.3) be fulfilled. Then, under H_0, there exists a r-dimensional sequence of Brownian bridges $\mathbf{B}(\tau)$ with independent components $B_j(\tau)$, $0 \le \tau \le 1$, $j = 1, \ldots, r$, such that*

$$\max_{j=1,\ldots,r} \sup_{0\le\tau\le1} \big|\widehat{B}_j(\tau) - B_j(\tau)\big| = o_{\mathsf{P}}(1). \tag{4}$$

Proof. The proof can be done separately with respect to each parameter and the basic steps are similar to those in [3]. Due to the consistency of the ML estimators, it follows that $\big\|\mathcal{I}(\widehat{\boldsymbol{\theta}}) - \mathcal{I}(\boldsymbol{\theta})\big\| = o_{\mathsf{P}}(1)$.

(i) First we consider the case of detecting change in $\boldsymbol{\mu}$. The component of the efficient score vector that we use is

$$\frac{1}{\sqrt{T}} \left[\frac{\partial}{\partial \boldsymbol{\mu}} \ell_k(\boldsymbol{\theta})\right]_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \frac{1}{\sqrt{T}} \big(\mathbf{I}_n - \widehat{\mathbf{K}}\big)^\top \widehat{\boldsymbol{\Omega}}^{-1} \sum_{t=1}^k \Big(\mathbf{y}_t - \widehat{\boldsymbol{\mu}} - \sum_{j=1}^p \widehat{\boldsymbol{\Phi}}_j(\mathbf{y}_{t-j} - \widehat{\boldsymbol{\mu}})\Big) + \tag{5}$$

$$+ \frac{1}{\sqrt{T}} \big(\mathbf{I}_n - \mathbf{K}\big)^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^k \Big(\mathbf{y}_t - \overline{\mathbf{y}} - \sum_{j=1}^p \boldsymbol{\Phi}_j(\mathbf{y}_{t-j} - \overline{\mathbf{y}})\Big) - \tag{6}$$

$$- \frac{1}{\sqrt{T}} \big(\mathbf{I}_n - \mathbf{K}\big)^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^k \Big(\mathbf{y}_t - \overline{\mathbf{y}} - \sum_{j=1}^p \boldsymbol{\Phi}_j(\mathbf{y}_{t-j} - \overline{\mathbf{y}})\Big), \tag{7}$$

where $\overline{\mathbf{y}} := T^{-1}\sum_{t=1}^T \mathbf{y}_t$. We will show that the second addend (6) converges to the Brownian bridge process and also that the difference between (5) and (7) is negligible. This will complete the first part of the proof. Let M be a positive constant which can be different from term to term. For the second addend (6) we obtain that

$$\begin{aligned} &\sup_{0\le\tau\le1} \left\| \frac{1}{\sqrt{T}} \big(\mathbf{I}_n - \mathbf{K}\big)^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^{\lfloor T\tau \rfloor} \Big(\mathbf{y}_t - \overline{\mathbf{y}} - \sum_{j=1}^p \boldsymbol{\Phi}_j(\mathbf{y}_{t-j} - \overline{\mathbf{y}})\Big) - \mathbf{B}(\tau) \right\| \le \\ &\le M \cdot \left(\sup_{0\le\tau\le1} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \overline{\mathbf{y}}) - \boldsymbol{\Sigma}\mathbf{B}(\tau) \right\| + \sup_{0\le\tau\le1} \left\| \frac{1}{\sqrt{T}} \sum_{j=1}^p \boldsymbol{\Phi}_j \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_{t-j} - \overline{\mathbf{y}}) \right\| \right) =: \\ &= M\big(A_T + B_T\big), \end{aligned}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Omega}\big(\mathbf{I}_n - \mathbf{K}^\top\big)^{-1}$. As $\mathbf{B}(\tau) = \mathbf{W}(\tau) - \tau\mathbf{W}(1)$, it holds that

$$A_T \le \sup_{0\le\tau\le1} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Sigma}\mathbf{W}(\tau) \right\| + \sup_{0\le\tau\le1} \left\{ |\tau| \cdot \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Sigma}\mathbf{W}(1) \right\| \right\}.$$

Both latter terms are $o_{\mathsf{P}}(1)$ according to Proposition 3 and hence $A_T = o_{\mathsf{P}}(1)$. Term B_T is treated as follows:

$$\begin{aligned} B_T \le M \sup_{1\le k\le T} \sup_{1\le s\le p} \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{k-s} (\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{W}(k-s) \right\| + M \sup_{1\le k\le T} \frac{k}{T\sqrt{T}} \left\| \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{W}(T) \right\| + \\ + M \frac{1}{\sqrt{T}} \sup_{1\le k\le T} \sup_{1\le s\le p} \left\| \mathbf{W}(k-s) - \frac{k}{T}\mathbf{W}(T) \right\| \end{aligned}$$

which is together $o_{\mathsf{P}}(1)$ due to Proposition 3 and Theorem 2.

The supremum of the norm of the difference between (5) and (7) is bounded by

$$\left\|\left(\mathbf{I}_n-\widehat{\mathbf{K}}\right)^\top\widehat{\boldsymbol{\Omega}}^{-1}-\left(\mathbf{I}_n-\mathbf{K}\right)^\top\boldsymbol{\Omega}^{-1}\right\|\cdot\left(\sup_{1\leq k\leq T}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\left(\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t\right)\right\|+\sup_{1\leq k\leq T}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\boldsymbol{\varepsilon}_t\right\|\right)+$$
$$+M\sup_{1\leq k\leq T}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\left[(\widehat{\boldsymbol{\mu}}-\overline{\mathbf{y}})+\sum_{j=1}^{p}\left(\widehat{\boldsymbol{\Phi}}_j-\boldsymbol{\Phi}_j\right)\left(\mathbf{y}_{t-j}-\overline{\mathbf{y}}\right)-\sum_{j=1}^{p}\widehat{\boldsymbol{\Phi}}_j\left(\widehat{\boldsymbol{\mu}}-\overline{\mathbf{y}}\right)\right]\right\|. \tag{8}$$

According to Proposition 4, $\left\|\left(\mathbf{I}_n-\widehat{\mathbf{K}}\right)^\top\widehat{\boldsymbol{\Omega}}^{-1}-\left(\mathbf{I}_n-\mathbf{K}\right)^\top\boldsymbol{\Omega}^{-1}\right\|=\mathcal{O}\left(\sqrt{T^{-1}\log\log T}\right)$ a.s. Expansion (10) and Proposition 4 yield that $\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\left(\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t\right)=\mathcal{O}(\sqrt{\log\log T})$ a.s. Law of the iterated logarithm for i.i.d. random vectors $\boldsymbol{\varepsilon}_t$ gives us $\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\boldsymbol{\varepsilon}_t\right\|=\mathcal{O}(\sqrt{\log\log T})$ a.s. The latest supremum in (8) is $o_{\mathsf{P}}(1)$ due to Proposition 4 and the fact that $\|\widehat{\boldsymbol{\mu}}-\overline{\mathbf{y}}\|=o(T^{\frac{1}{\nu}-1})$ a.s., for some $\nu>2$. This completes the proof for the part (i).

(ii) Let us suppose we want to test changes in $\boldsymbol{\Phi}_s$, $s=1,\ldots,p$. Let $\mathbf{v}_t:=\mathsf{vec}\left(\mathbf{y}_{t-1},\ldots,\mathbf{y}_{t-p}\right)\in\mathbb{R}^{np}$, $\boldsymbol{\mu}_{\boldsymbol{p}}:=\mathsf{vec}(\boldsymbol{\mu},\ldots,\boldsymbol{\mu})\in\mathbb{R}^{np}$, $\mathbf{M}_t:=\left(\mathbf{v}_t-\boldsymbol{\mu}_{\boldsymbol{p}}\right)^\top\otimes\mathbf{I}_n\in\mathbb{R}^{n\times n^2p}$ and $\widehat{\mathbf{M}}_t:=\left(\mathbf{v}_t-\widehat{\boldsymbol{\mu}_{\boldsymbol{p}}}\right)^\top\otimes\mathbf{I}_n$. Then

$$\left[\frac{\partial}{\partial\boldsymbol{\phi}}\ell_k(\boldsymbol{\theta})\right]_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}=\sum_{t=1}^{k}\widehat{\mathbf{M}}_t^\top\widehat{\boldsymbol{\Omega}}^{-1}\left(\mathbf{y}_t-\widehat{\boldsymbol{\mu}}-\widehat{\mathbf{M}}_t\widehat{\boldsymbol{\phi}}\right)=\sum_{t=1}^{k}\widehat{\mathbf{M}}_t^\top\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\varepsilon}}_t.$$

Replacing $\widehat{\boldsymbol{\Omega}}$ with $\boldsymbol{\Omega}$ does not change the asymptotic distribution. By the invariance principle for the martingale difference sequence $\left\{\mathbf{M}_t^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right\}$ it holds

$$\sup_{0\leq\tau\leq1}\left\|\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{\lfloor T\tau\rfloor}\left(\mathbf{M}_t^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right)-\frac{\lfloor T\tau\rfloor}{T}\sum_{t=1}^{T}\left(\mathbf{M}_t^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right)-\boldsymbol{\Sigma}_1\mathbf{B}(\tau)\right)\right\|=o_{\mathsf{P}}(1), \tag{9}$$

for some $\boldsymbol{\Sigma}_1>0$. We now have to show that the error committed by replacing the parameters in the formula (9) with their maximum likelihood parameters is negligible:

$$\left(\sum_{t=1}^{k}\left(\widehat{\mathbf{M}}_t^\top\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\varepsilon}}_t\right)-\sum_{t=1}^{k}\left(\mathbf{M}_t^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right)\right)+\left(-\frac{k}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{M}}_t^\top\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\varepsilon}}_t\right)+\frac{k}{T}\sum_{t=1}^{T}\left(\mathbf{M}_t^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right)\right)=:\mathbf{R}_{k,T}+\mathbf{S}_{k,T}.$$

Term $\mathbf{R}_{k,T}$ can be treated as follows:

$$\mathbf{R}_{k,T}=\sum_{t=1}^{k}\left[\left((\mathbf{v}_t-\boldsymbol{\mu}_{\boldsymbol{p}}-\widehat{\boldsymbol{\mu}_{\boldsymbol{p}}}+\boldsymbol{\mu}_{\boldsymbol{p}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\left(\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t+\boldsymbol{\varepsilon}_t\right)\right]-\sum_{t=1}^{k}\left[\left((\mathbf{v}_t-\boldsymbol{\mu}_{\boldsymbol{p}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right]=$$
$$=\sum_{t=1}^{k}\left[\left((\mathbf{v}_t-\boldsymbol{\mu}_{\boldsymbol{p}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\left(\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t\right)\right]+\sum_{t=1}^{k}\left[\left((\boldsymbol{\mu}_{\boldsymbol{p}}-\widehat{\boldsymbol{\mu}_{\boldsymbol{p}}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right]+$$
$$+\sum_{t=1}^{k}\left[\left((\boldsymbol{\mu}_{\boldsymbol{p}}-\widehat{\boldsymbol{\mu}_{\boldsymbol{p}}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\left(\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t\right)\right]=:\mathbf{R}_{k,T}^{(1)}+\mathbf{R}_{k,T}^{(2)}+\mathbf{R}_{k,T}^{(3)}.$$

By Proposition 4 and the law of the iterated logarithm for $\{\varepsilon_t\}$

$$\|\mathbf{R}_{k,T}^{(2)}\|=\left\|\left((\boldsymbol{\mu}_{\boldsymbol{p}}-\widehat{\boldsymbol{\mu}_{\boldsymbol{p}}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\sum_{t=1}^{k}\boldsymbol{\varepsilon}_t\right\|=\mathcal{O}\left(\sqrt{\frac{\log\log T}{T}}\right)\cdot\mathcal{O}(\sqrt{T\log\log T})=\mathcal{O}(\log\log T)\quad\text{a.s.}$$

Before analyzing $\mathbf{R}_{k,T}^{(1)}$ and $\mathbf{R}_{k,T}^{(3)}$ we write

$$\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t=\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}-\sum_{j=1}^{p}\widehat{\boldsymbol{\Phi}}_j\left(\mathbf{y}_{t-j}-\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}+\boldsymbol{\mu}\right)+\sum_{j=1}^{p}\boldsymbol{\Phi}_j\left(\mathbf{y}_{t-j}-\boldsymbol{\mu}\right)=$$
$$=\left(\mathbf{I}_n-\mathbf{K}\right)(\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}})+\sum_{j=1}^{p}\left(\boldsymbol{\Phi}_j-\widehat{\boldsymbol{\Phi}}_j\right)\left(\mathbf{y}_{t-j}-\boldsymbol{\mu}\right)+\sum_{j=1}^{p}\left(\widehat{\boldsymbol{\Phi}}_j-\boldsymbol{\Phi}_j\right)\left(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\right). \tag{10}$$

Imposing the latter expansion (10) into $\mathbf{R}_{k,T}^{(1)}$ and $\mathbf{R}_{k,T}^{(3)}$, using Proposition 4 and the invariance principle we obtain that $\left\|\mathbf{R}_{k,T}^{(3)}\right\|=\mathcal{O}(\log\log T)$ a.s. and $\left\|\mathbf{R}_{k,T}^{(1)}\right\|=\mathcal{O}(\log\log T)$ a.s. Term $\mathbf{S}_{k,T}$ can be analyzed in the similar way, because $\mathbf{S}_{k,T}=-\frac{k}{T}\mathbf{R}_{T,T}$. Hence, $\sup_{1\leq k\leq T}\|\mathbf{R}_{k,T}+\mathbf{S}_{k,T}\|=\mathcal{O}(\log\log T)$ a.s., which concludes the proof of the theorem. □

3 Test and small simulation study

The application of Theorem 1 for detecting abrupt changes in $\mathbb{VAR}(p)$ model is as follows: If one wants to test for changes in d parameters, $d = 1, \ldots, r$, on the overall significance level α, the null hypothesis H_0 is rejected if

$$\max_{1<k<T} \left| \widehat{B}_j\big(k/T\big) \right| \geq C(\alpha^*), \quad \text{for at least one } j, \tag{11}$$

where $C(\alpha^*)$ is a critical value and $\alpha^* = 1 - (1-\alpha)^{\frac{1}{d}}$. If (11) holds we conclude that there is a change in parameter θ_j, $j = 1, \ldots, r$. Since (4) holds, critical value $C(\alpha^*)$ can be obtained from the limiting process:

$$\mathsf{P}\Big[\sup_{0\leq\tau\leq1} \big|B_1(\tau)\big| > x\Big] = 1 + \sum_{k=-\infty}^{\infty} (-1)^{k+1} \mathrm{e}^{-2k^2x^2},$$

see [1], p. 103. In order to explore accuracy of the approximation, the simulation study was carried out. We considered a bivariate $\mathbb{VAR}(1)$ model of the form $\mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$, $t = 1, \ldots, 200$, where $\boldsymbol{\varepsilon}_t$ was the centered i.i.d. Gaussian sequence of random vectors with variance matrix $\mathbf{\Omega}$ and

$$\mathbf{\Phi} = \begin{pmatrix} 0.5 & 0.1 \\ 0.2 & 0.1 \end{pmatrix} \quad \text{and} \quad \mathbf{\Omega} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}.$$

Figure 1 illustrates a reasonable convergence of the empirical distribution function of $\sup_{0\leq\tau\leq1} \big|\widehat{B}_j(\tau)\big|$, $j = 1, \ldots, 6$, based on 1 000 simulations of the latter model to the limiting distribution. The accuracy does not get worse too much even if some more persistent autoregressive matrix is used. One obtains satisfactory results also for moderate sample sizes. We do not report any figures to save space.

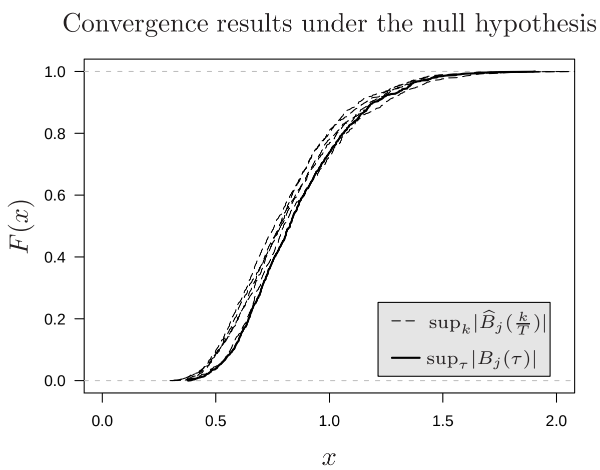

Figure 1: Empirical distribution functions of $\sup_\tau \big|\widehat{B}_j(\tau)\big|$ and the limiting distribution.

4 Conclusion

The article presents the statistical procedure to detect changes in the mean and autoregressive parameters of $\mathbb{VAR}(p)$ models. We have shown that the test statistic based on the efficient score vector converges to the Brownian bridge. Simulation results confirm that the convergence is relatively rapid even for moderate length of time series. Repeated Monte-Carlo experiments for moderate sample sizes reveal that the components of the score vector can be regarded as asymptotically uncorrelated and asymptotically Gaussian based on the p-values of multivariate Shapiro-Wilk test. The extension of the procedure to cover change detection in the variance matrix is also possible and it is the subject of the future study.

A Supplement

The first theorem describes the boundedness of increments of the multivariate Wiener process.

Theorem 2. *If* $\mathbf{W}$ *is a multivariate Brownian motion with independent components, it holds*

$$\limsup_{k\to\infty} \sup_{0\le s\le p} \|\mathbf{W}(k-s)-\mathbf{W}(k)\| = \mathcal{O}\big(\sqrt{\log k}\big) \quad \text{a.s.}$$

Proof. Straightforward using Theorem 1.2.1 of [2]. □

Proposition 3. *Under Assumptions (A.1) − (A.3) it holds that*

$$\sup_{0\le\tau\le 1} \left\| \frac{1}{\sqrt{T}} \Big(\sum_{t=1}^{\lfloor T\tau\rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) \Big) - \boldsymbol{\Sigma}\mathbf{W}(\tau) \right\| = o_{\mathsf{P}}(1).$$

Proof. The invariance principle for the stationary linear processes yields

$$\left\| \sum_{t=1}^{s} (\mathbf{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Sigma}\mathbf{W}(s) \right\| = o\big(s^{\frac{1}{2}-\delta}\big) \quad \text{a.s., } \delta > 0.$$

Let $s := \lfloor T\tau\rfloor$, $0 \le \tau \le 1$. Then

$$\frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{\lfloor T\tau\rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Sigma}\mathbf{W}(T\tau) \right\| = o(1) \quad \text{a.s.} \quad T\to\infty, \tag{12}$$

from which

$$\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau\rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Sigma}\mathbf{W}(\tau) \right\| = \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau\rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \frac{1}{\sqrt{T}}\boldsymbol{\Sigma}\mathbf{W}(T\tau) + \frac{1}{\sqrt{T}}\boldsymbol{\Sigma}\mathbf{W}(T\tau) - \boldsymbol{\Sigma}\mathbf{W}(\tau) \right\| \le$$

$$\le \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{\lfloor T\tau\rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Sigma}\mathbf{W}(T\tau) \right\| + \|\boldsymbol{\Sigma}\| \cdot \left\| \frac{1}{\sqrt{T}}\mathbf{W}(T\tau) - \mathbf{W}(\tau) \right\|. \tag{13}$$

The first addend in (13) is $o(1)$ a.s. due to (12). For the second addend in (13) we use the well known fact that $\frac{1}{\sqrt{T}}\mathbf{W}(T\tau) \overset{d}{\equiv} \mathbf{W}(\tau)$ from which follows that $\frac{1}{\sqrt{T}}\mathbf{W}(T\tau) - \mathbf{W}(\tau) = o_{\mathsf{P}}(1)$. The result is therefore $o_{\mathsf{P}}(1)$. Because $h(x) = \sup_{0\le t\le 1} x(t)$ is continuous then the statement of the proposition holds. □

Proposition 4. *Under Assumptions (A.1)–(A.3) it holds* $\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\| = \mathcal{O}\big(\sqrt{T^{-1}\log\log T}\big)$ a.s., $\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}\| = \mathcal{O}\big(\sqrt{T^{-1}\log\log T}\big)$ a.s. *and* $\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\| = \mathcal{O}\big(\sqrt{T^{-1}\log\log T}\big)$ a.s., *as* $T\to\infty$.

Proof. The proof follows from the asymptotic equivalence of ML and least squares estimators and the multivariate extension of Theorem 2.1 in [4]. □

Acknowledgements

The work was supported by the grants SVV 265315/2012 and GAUK 586712/2012.

References

[1] Billingsley P.: *Convergence of Probability Measures*. 2nd ed., Wiley, New York, 1999.

[2] Csörgő M. and Révész P.: *Strong Approximations in Probability and Statistics*. Akadémiai Kiadó, Budapest, 1981.

[3] Gombay E.: Change Detection in Autoregressive Time Series. *Journal of Multivariate Analysis* **99**(3), (2008), 451–464.

[4] Gombay E. and Serban D.: Monitoring Parameter Change in AR(p) Time Series Models. *Journal of Multivariate Analysis* **100**(4), (2009), 715–725.

[5] Qu Z. and Perron P.: Estimating and Testing Structural Changes in Multivariate Regressions. *Econometrica* **75**(2), (2007), 459–502.

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