Quality Technology & Quantitative Management

ISSN: (Print) 1684-3703 (Online) Journal homepage: https://www.tandfonline.com/loi/ttqm20

# A computationally efficient self-starting scheme to monitor general linear profiles with abrupt changes

Zhiming Xia & Fugee Tsung

**To cite this article:** Zhiming Xia & Fugee Tsung (2019) A computationally efficient self-starting scheme to monitor general linear profiles with abrupt changes, Quality Technology & Quantitative Management, 16:3, 278-296, DOI: 10.1080/16843703.2017.1396956

**To link to this article:** https://doi.org/10.1080/16843703.2017.1396956

Published online: 25 Nov 2017.

QUALITY TECHNOLOGY & QUANTITATIVE MANAGEMENT
2019, VOL. 16, NO. 3, 278–296
https://doi.org/10.1080/16843703.2017.1396956

Taylor & Francis
Taylor & Francis Group



# A computationally efficient self-starting scheme to monitor general linear profiles with abrupt changes

Zhiming Xia[a] and Fugee Tsung[b]

[a]School of Mathematics, Northwest University, Xi'an, China; [b]Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

**ABSTRACT**

A self-starting monitoring scheme is proposed in this paper for the simultaneous detection of variance and coefficients in linear profiles with unknown error distributions. Based on the global data, we construct a sequential Wald-type charting statistic, obtain the corresponding asymptotical distributions and further provide a recursive algorithm to quickly calculate statistics sequentially. Control limits of our charting statistics are also constructed based on their asymptotical distributions. Finally, we apply our method to analyze both artificial and real data, and numerical results show that our method performs well.

**ARTICLE HISTORY**
Accepted 23 October 2017

**KEYWORDS**
Linear profile; self-starting; SPC; Wald-type charting statistic

## 1. Introduction

The evolution of a system is generally affected by a large number of unavoidable factors or common causes, which can be characterized as a series of stable random variables or random variables with an unchanged relationship. This characterization is known as a stable or in-control system state. However, the interference of avoidable factors or special causes can result in the system losing stability and entering an out-of-control state, which needs to be detected as quickly as possible. SPC is a set of procedures to determine whether a system is in control or not and is broadly applied in many fields including financial service (see Tsung, Zhou, & Jiang, 2007), healthcare (see Woodall (2007)), medical research and others. In this paper, we focus on monitoring a linear relationship or a so-called linear profile.

Many researchers over past decades have proposed some methods for detecting linear profiles. Croarkin and Varner's (1982) scheme is based on the principle of inverse calibration and involves plotting the deviations of the measured values from the standards on a Shewhart control chart. Kang and Albin (2000) proposed a Shewhart-type control chart, multivariate Hotelling $T^2$ charts, based on successive least squares estimators of coefficients in simple linear profiles, and an EWMA chart based on average residuals between response variable and regression lines. To improve the low efficiency of $T^2$ charts, Noorossana, Amiri, Vaghefi, and Roghanian (2004) and Noorossana and Amiri (2007) proposed a MCUSUM chart based on vectors of the regression estimators, which are extended to monitor polynomial profiles by Kazemzadeh, Noorossana, and Amiri (2009). Similarly, Zou, Tsung, and Wang (2007) built a MEWMA chart scheme to monitor general linear profiles. Niaki, Abbasi, and Arkat (2007) obtained a control chart based on the generalized linear test to monitor the coefficients in conjunction with an R-chart to monitor the error variance. Zhang, Li, and Wang (2009) proposed a EWMA control chart to detect coefficients and variance simultaneously based on local likelihood ratios. The above detection approaches can monitor coefficients simultaneously but cannot identify

**CONTACT** Fugee Tsung season@ust.hk

the source of variation, which can be obtained by monitoring each coefficient individually using several univariate control charts. Kim, Mahmoud, and Woodall (2003) recommended monitoring parameters in simple linear profiles using separate EWMA charts. Similarly, Saghaei, Mehrjoo, and Amiri (2009) constructed a CUSUM chart to detect coefficients and variance separately. Recently Khedmati and Niaki (2016a, 2016b) proposed two new approaches, Khedmati and Niaki (2016a, 2016b), to monitor simple profiles in multistage processes. Variable sampling interval schemes are also considered by Kazemzadeh, Amiri, and Kouhestani (2016) and Magalhaes and Von Doellinger (2016) in monitoring parameter shifts in linear profiles.

All of the above methods assume that in-control values of parameters are known or can be accurately estimated in Phase I. Theoretically, even if Phase I data can be obtained to estimate the desired parameters, the effect of estimation on control chart performance should be carefully considered in further research (see Ghosh, Reynolds, & Hui, 1981; Quesenberry, 1993; Chen, 1997; Woodall & Montgomery, 1999). On the other hand, by constructing a self-starting detection method, we can also avoid this theoretical problem and monitor the parameters of interest quickly and directly. In the case of location models, Hawkins, Qiu, and Kang (2003), Hawkins (1987) and Hawkins and Zamba (2005a, 2005b) firstly proposed a so-called self-starting method to monitor mean or variance based on the likelihood ratio, which Zou, Zhang, and Wang (2006) extended to monitor simple linear profiles. Likewise, Zou, Zhou, Wang, and Tsung (2007) constructed two separate EWMA charts based on standardized recursive residuals (see Brown, Durbin, & Evans, 1975) to monitor coefficients and variance correspondingly in a simple linear profile. Recently, some researchers propose some new self-starting methods to simultaneously monitor shifts in mean and variance in simple linear models (see Amiri, Khosravi, & Ghashghaei, 2016a, 2016b; Li, Zhang, & Wang, 2010).

In this paper, we propose a new self-starting method to simultaneously monitor regression coefficients and variance in a general linear profile. Our method is different from existing methods in the literature. Firstly, to our best knowledge, most of non-self-starting methods construct Shewhart charts, CUSUM charts or EWMA charts based on different building blocks for quick detection, which often use local data near the present but does not make full use of global data. Our method aims to use all available data. Secondly, to resolve the calculation problem, we also proposed a set of recursive calculation formulae to reduce calculation complexity. Thirdly, the statistical performance of the above existing methods relies heavily on the normality assumption. Mahmoud and Woodall (2004) and Mahmoud (2004), etc. reported that departures from the normality assumption can greatly affect the performance of statistical monitoring methods. Our method doesn't make this assumption. Fourthly, we also give the limiting distribution of charting statistic, which is useful for calculating the thresholds for different error distributions.

The remainder of the article is organized as follows. In Section 2, we introduce linear profiles and list some necessary assumptions in detail. In Section 3, we build self-starting control charts and obtain their corresponding asymptotic distributions of our charting statistics by which control limits are given. Meanwhile, a detailed fast algorithm for computing charting statistics is also constructed. In Section 4, numerical experiments and corresponding analysis are given. All proofs for our theorems are included in the Appendix 1.

## 2. Statistical model for a general linear profile

To describe linear profiles with shifts in coefficients or variance, consider the following change-point formulation

$$y_{ij} = \begin{cases} \mathbf{x}_{ij}^T \boldsymbol{\beta}_0 + \sigma_0 \varepsilon_{ij}, & i = 1, \ldots, n_j,\ j = 1, \ldots, \tau, \\ \mathbf{x}_{ij}^T \boldsymbol{\beta}_1 + \sigma_1 \varepsilon_{ij}, & i = 1, \ldots, n_j,\ j = \tau + 1, \tau + 2, \ldots, \end{cases} \tag{1}$$

where $\mathbf{x}_{ij} = (x_{ij,1}, x_{ij,2}, \ldots, x_{ij,p})^T$ denotes the $i$th design vector at time $j$, and $\varepsilon_{ij}$ is an unobservable disturbance with $\mathrm{Var}(\varepsilon_{ij}) = 1$. The vectors $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1$ and variances $\sigma_0, \sigma_1$ are all unknown parameters.

We assume that $\boldsymbol{\delta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0 \neq \mathbf{0}$ or $\sigma_0 \neq \sigma_1$, and $\tau$ denotes the true unknown time or change-point and can take infinity when no shift occurs. The model can be described equivalently by

$$\mathbf{y}_j = \begin{cases} \mathbf{X}_j\boldsymbol{\beta}_0 + \sigma_0\boldsymbol{\varepsilon}_j, & j = 1, \ldots, \tau, \\ \mathbf{X}_j\boldsymbol{\beta}_1 + \sigma_1\boldsymbol{\varepsilon}_j, & j = \tau+1, \tau+2, \ldots, \end{cases} \tag{2}$$

where $\mathbf{y}_j = (y_{1j}, y_{2j}, \ldots, y_{n_jj})^T$, $\mathbf{X}_j = (\mathbf{x}_{1j}, \mathbf{x}_{2j}, \ldots, \mathbf{x}_{n_jj})^T$ and $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j}, \varepsilon_{2j}, \ldots, \varepsilon_{n_jj})^T$.

For further theoretical research in detail later, we add the following three assumptions.

**(A1)** Suppose $x_{ij,l}, i = 1, \ldots, n_j, l = 1, \ldots, p$ is fixedly designed satisfying $|x_{ij,l}| < c$ where $c$ is a positive constant. And when $n_j \to +\infty$, $\mathbf{X}_j^T\mathbf{X}_j/n_j \to \mathbf{Q}$ where $\mathbf{Q}$ is positively definite.
The assumption **A**1 is standard for change-point analysis (see Csorg & Horvath, 1997, etc.) in linear models and it will be used for the central limit theorem (CLT) and the law of large number.

**(A2)** $\varepsilon_{ij}, i = 1, \ldots, n_j, j = 1, 2, \ldots$ are i.i.d. satisfying $E\left(\varepsilon_{ij}\right) = 0, E\left(\varepsilon_{ij}^3\right) = 0, \mathrm{Var}\left(\varepsilon_{ij}\right) = 1, E\left(|\varepsilon|_{ij}^{4+\delta}\right) < \infty$ for some $\delta > 0$.
This assumption indicates that $v^2 = E(\sigma_0^2\varepsilon_{ij}^2 - \sigma_0^2)^2 < \infty$ and that the distribution of $\varepsilon_{ij}$ is approximately symmetric.

**(A3)** $n_i/n_j \to 1, i, j = 1, 2, \ldots, n_i, n_j \to \infty$.
The potential purpose of the above assumption **A**3 is that sample size $n_j$ at time $j$ is sufficiently large, which guarantees that the key building block $\xi_i$ in Theorem 1 can be approximated well and thus our charting statistic can be sufficiently approximated by its asymptotic distribution. As shown in numerical study later, a small size, $n_j$, is sometimes sufficient for a good asymptotic performance if errors $\varepsilon_{ij}$ are normally distributed and vice versa.

## 3. A self-starting chart for shifts in coefficients or variance

### 3.1. Charting statistic and SPC phase II procedure

For a fixed $t \geq 1$ and $1 \leq k \leq t-1$, we can construct the following estimators

$$\begin{aligned}
\hat{\beta}(1,k) &= \left(\mathbf{X}_{1,k}^T\mathbf{X}_{1,k}\right)^{-1}\mathbf{X}_{1,k}^T\mathbf{Y}_{1,k}, & \hat{\beta}(k+1,t) &= \left(\mathbf{X}_{k+1,t}^T\mathbf{X}_{k+1,t}\right)^{-1}\mathbf{X}_{k+1,t}^T\mathbf{Y}_{k+1,t}, \\
\hat{\mathbf{e}}(1,k) &= \mathbf{Y}_{1,k} - \mathbf{X}_{1,k}\hat{\beta}(1,k), & \hat{\mathbf{e}}(k+1,t) &= \mathbf{Y}_{k+1,t} - \mathbf{X}_{k+1,t}\hat{\beta}(k+1,t), \\
\hat{\sigma}^2(1,k) &= \left\|\hat{\mathbf{e}}(1,k)\right\|^2 / \sum_{j=1}^{k} n_j, & \hat{\sigma}^2(k+1,t) &= \left\|\hat{\mathbf{e}}(k+1,t)\right\|^2 / \sum_{j=k+1}^{t} n_j, \\
\hat{v}^2(1,k) &= \frac{\left\|\hat{\mathbf{e}}^2(1,k) - \hat{\sigma}^2(1,k)\mathbf{1}_k\right\|^2}{\sum_{j=1}^{k} n_j}, & \hat{v}^2(k+1,t) &= \frac{\left\|\hat{\mathbf{e}}^2(k+1,t) - \hat{\sigma}^2(k+1,t)\mathbf{1}_{t-k}\right\|^2}{\sum_{j=k+1}^{t} n_j}.
\end{aligned} \tag{3}$$

where $\mathbf{X}_{i,j}^T = \left(\mathbf{X}_i^T, \mathbf{X}_{i+1}^T, \ldots, \mathbf{X}_j^T\right), \mathbf{Y}_{i,j}^T = \left(\mathbf{y}_i^T, \mathbf{y}_{i+1}^T, \ldots, \mathbf{y}_j^T\right)$ for an arbitrary $1 \leq i < j \leq t$, $\mathbf{1}_k$ denotes a vector with $k$ components all equal to 1, and $\hat{\mathbf{e}}^2(0,k)$ denotes a new vector with each component equal to the square of the corresponding component in $\hat{\mathbf{e}}(0,k)$. A combined Wald-Type test statistic is constructed naturally as follow

$$\mathrm{CW}_{k,t} = \frac{\left\|\hat{\beta}(k+1,t) - \hat{\beta}(1,k)\right\|_{W_{1k}}^2}{\hat{\sigma}^2(1,k,t)} + \frac{\left\|\hat{\sigma}^2(k+1,t) - \hat{\sigma}^2(1,k)\right\|_{W_{2k}}^2}{\hat{v}^2(1,k,t)}$$

where

$$\left\|\hat{\beta}(k+1,t)-\hat{\beta}(1,k)\right\|^2_{W_{1k}} = (\hat{\beta}(k+1,t)-\hat{\beta}(1,k))^T W_{1k}(\hat{\beta}(k+1,t)-\hat{\beta}(1,k))$$
$$\left\|\hat{\sigma}^2(k+1,t)-\hat{\sigma}^2(1,k)\right\|^2_{W_{2k}} = (\hat{\sigma}^2(k+1,t)-\hat{\sigma}^2(1,k))^2 W_{2k}$$

and $W_{1k} = [(\mathbf{X}'_{1,k}\mathbf{X}_{1,k})^{-1} + (\mathbf{X}'_{k+1,t}\mathbf{X}_{k+1,t})^{-1}]^{-1}$, $W_{2k} = [(\sum_{j=1}^{k} n_j)^{-1} + (\sum_{j=k+1}^{t} n_j)^{-1}]^{-1}$ and $\hat{\sigma}^2(1,k,t) = \hat{\sigma}^2(1,k)\sum_{j=1}^{k} n_j/\sum_{j=1}^{t} n_j + \hat{\sigma}^2(k+1,t)\sum_{j=k+1}^{t} n_j/\sum_{j=1}^{t} n_j, \hat{\nu}^2(1,k,t) = \hat{\nu}^2(1,k)$ $\sum_{j=1}^{k} n_j/\sum_{j=1}^{t} n_j + \hat{\nu}^2(k,t)\sum_{j=k+1}^{t} n_j/\sum_{j=1}^{t} n_j$. In the null case of no change-points, the statistic has an asymptotic central chi-square distribution with $p+1$ degrees of freedom, which is accurately shown in Lemma 8 as

$$\mathrm{CW}_{k,t} \Longrightarrow \frac{B_{t,k}^T B_{t,k}}{\frac{k}{t}\left(1-\frac{k}{t}\right)} \stackrel{d}{=} \chi^2(p+1), k = 1,2,\ldots,t-1,$$

where $B_{t,k}$ denotes a $(p+1)$-dimensional Brown Bridge described formally in detail later. However, with a shift in coefficients or variance, the distribution becomes noncentral. The non-centrality parameter becomes nonzero with the first out-of-control observation, and then increases to a limit that depends on the magnitude of the parameter shift and on the length of the in-control history. To illustrate this, suppose that the variance remains at the in-control level but the coefficients shift by $\delta$ standard deviations immediately after observation $\tau$. Then when observation $t > \tau$, the coefficient component of our statistic for $k = \tau$ has a non-centrality parameter that can be written as

$$\frac{\delta^T\delta}{\sigma_0^2\frac{\tau}{t}(1-\frac{\tau}{t})}.$$

Maximizing $\mathrm{CW}_{k,t}$ among all possible values, $1 \le k \le t-1$, to measure the global bias at time $t$, the final charting statistic is constructed by

$$\mathrm{CW}_t = \max_{1\le k\le t-1} \mathrm{CW}_{k,t}. \tag{4}$$

To obtain corresponding thresholds $\{h_{t,\alpha}, t = 2,3,\ldots\}$ given an alarm probability $\alpha \in (0,1)$, we build the asymptotical distribution of $\mathrm{CW}_t, t = 2,3,\ldots$ in the following theorem.

**Theorem 1:** *Under assumptions A1,A2 and A3 and the null case of no change-points, we have*

$$CW_t \Longrightarrow \max_{1\le k\le t-1} \frac{B_{t,k}^T B_{t,k}}{\frac{k}{t}\left(1-\frac{k}{t}\right)}$$

*where* $B_{t,k} = W_t\left(\frac{k}{t}\right) - \frac{k}{t}W_t(1), k = 1,\ldots,t-1$ *and* $W_t(k/t)$ *are corresponding some discrete values of a standard* $(p+1)-$*dimensional Brownian motion at finite time points* $1/t,\ldots,(t-1)/t$ *such that* $W_t(k/t) = 1/\sqrt{t}\sum_{i=1}^{k}\xi_i$, $\xi_i \sim N(\mathbf{0}, \mathbf{I}_{p+1})$ *does not depend on* $t$.

**Remark 1:** By Theorem 1, we know that charting statistics $\{\mathrm{CW}_t, t = 2,3,\ldots\}$ are not independent, but are linked together asymptotically by a common normal series $\{\xi_j, j = 1,2,\ldots\}$ on which we can simply obtain control limits for different alarm probability $\alpha$ as shown in Table 1 even for different unknown error distributions. Specifically, asymptotical distributions of all $\mathrm{CW}_t, t = 2,3,\ldots$ are based on a common set of i.i.d. normal random variables $\xi_1,\xi_2,\ldots,\xi_j,\ldots$ in which each $\xi_j$ is a limit, respectively, by sufficiently large $n_j$ data at time $j$ as shown in the latter proof in Appendix 1. The charting statistic $\mathrm{CW}_t$ constructed in Theorem 1 converges to the maximum of a squared

standard Brown Bridge at only $t-1$ discrete time points rather than the whole time interval $[0,1]$. All information at these $t-1$ time points are caught by our charting statistic $\mathrm{CW}_t$ which is naturally approximated by such a special functional of Brown Bridge $B_{t,k}$ on $t-1$ discrete points, $k = 1,\ldots,t-1$.

**Remark 2:** Once the charting statistic $\mathrm{CW}_t$ signals for a possible out-of-control state before $t$, the corresponding maximum value point $\hat{\tau} = \arg\max_{1\le k\le t-1} \mathrm{CW}_k$ can be constructed as a reliable estimate for the potential abrupt change-point $\tau$. As such, our method can be used to test and estimate potential out-of-control points simultaneously.

Similar to (4), we can also construct the corresponding charting statistic $\mathrm{CW}_t^{(1)} = \max_{1\le k\le t-1} \mathrm{CW}_{k,t}^{(1)}$ to singly diagnose whether coefficients shift or not where

$$\mathrm{CW}_{k,t}^{(1)} = \frac{\left\|\hat{\beta}(k+1,t) - \hat{\beta}(1,k)\right\|_{W_{1k}}^2}{\hat{\sigma}^2(1,k,t)}, k = 1,2,\ldots,t-1$$

and the other charting statistic $\mathrm{CW}_t^{(2)} = \max_{1\le k\le t-1} \mathrm{CW}_{k,t}^{(2)}$ to diagnose possible shift in variance where

$$\mathrm{CW}_{k,t}^{(2)} = \frac{\left\|\hat{\sigma}^2(k+1,t) - \hat{\sigma}^2(1,k)\right\|_{W_{2k}}^2}{\hat{\nu}^2(1,k,t)}, k = 1,2,\ldots,t-1.$$

The asymptotical distributions of them can be easily obtained in Theorem 2.

**Theorem 2:** *Under assumptions* A1, A2 *and* A3 *and the null case of no change-points, we have*

$$CW_t^{(1)} \Longrightarrow \max_{1\le k\le t-1} \frac{B_{t,k}^{(1)T} B_{t,k}^{(1)}}{\frac{k}{t}\left(1-\frac{k}{t}\right)},$$

$$CW_t^{(2)} \Longrightarrow \max_{1\le k\le t-1} \frac{\left(B_{t,k}^{(2)}\right)^2}{\frac{k}{t}\left(1-\frac{k}{t}\right)},$$

*where* $B_{t,k}^{(1)}, B_{t,k}^{(2)}$ *are p-dimensional and one-dimensional Brown Bridges, respectively, which have similar structures like that in Theorem* 1.

The complete monitoring scheme can be described by Algorithm I below.

**Algorithm I** (The on-line monitoring algorithm)

**(I-1)** Observing the $t$-th sample $(\mathbf{X}_t, \mathbf{y}_t)$, obtain the charting statistic $\mathrm{CW}_t$ according to (4).

**(I-2)** Sample the next observation $(\mathbf{X}_{t+1}, \mathbf{y}_{t+1})$ when $\mathrm{CW}_t \le h_{t,\alpha}$ where $\{h_{t,\alpha}, t \ge 2\}$ is a sequence of control limits computed by giving a alarm probability $\alpha$ or the corresponding in-control average running length (IC ARL).

**(I-3)** Stop sampling and signal out-of-control alert when $\mathrm{CW}_t > h_t$.

### 3.2. Fast computation for $CW_t$

Given the $(t+1)$-th observation $(\mathbf{X}_{t+1}, \mathbf{y}_{t+1})$, it is conceivable that $\mathrm{CW}_{k,t+1}, 1 \le k \le t$ seems to lead to heavy computation. However, the truth is not so simple because practical computation can only rely on one single basic term $\mathbf{A}_{t+1} = \mathbf{X}_{1,t+1}^T \mathbf{X}_{1,t+1}, \hat{\beta}(1,t+1)$, which can be computed recursively based on historical values $\mathbf{A}_t, \hat{\beta}(1,t)$ and new observations $(\mathbf{X}_{t+1}, \mathbf{y}_{t+1})$. The complexity of the entire computation is proportional to $t^2$, i.e. $\mathcal{O}(t^2)$, with further details analyzed in Algorithm II.

**Algorithm II** (The recursive algorithm for $\mathrm{CW}_t$)

**(II-1)** Once reading $t+1$, we firstly update $(t+1)-$th group of statistics as the following

$$\begin{aligned}\mathbf{A}_{t+1} &= \mathbf{A}_t + \mathbf{X}_t'\mathbf{X}_t,\\ \hat{\beta}(1,t+1) &= \hat{\beta}(1,t) + \mathbf{A}_t^{-1}\mathbf{X}_{t+1}'\left(I + \mathbf{X}_{t+1}\mathbf{A}_t^{-1}\mathbf{X}_{t+1}'\right)^{-1}\left(\mathbf{y}_{t+1} - \mathbf{X}_{t+1}\hat{\boldsymbol{\beta}}(1,t)\right).\end{aligned}$$

Based on $\mathbf{A}_{t+1}, \hat{\beta}(1,t+1)$ and data $\mathbf{Y}_{1,t+1}, \mathbf{X}_{1,t+1}$, we can update $\hat{\mathbf{e}}(1,t+1), \hat{\sigma}^2(1,t+1)$ by the formulae (3).

$$\begin{aligned}\hat{\mathbf{e}}(1,t+1) &= \mathbf{Y}_{1,t+1} - \mathbf{X}_{1,t+1}\hat{\beta}(1,t+1),\\ \hat{\sigma}^2(1,t+1) &= \left\|\hat{\mathbf{e}}(1,t+1)\right\|^2 / \sum_{j=1}^{t+1} n_j,\\ \hat{v}^2(1,t+1) &= \left\|\hat{\mathbf{e}}^2(1,t+1) - \hat{\sigma}^2(1,t+1)\mathbf{1}_{t+1}\right\|^2 / \sum_{j=1}^{t+1} n_j,\end{aligned}$$

which leads to computation complexity $\mathcal{O}(t)$.

**(II-2)** For $1 \le k \le t$, based on II-1 we can also update $t$ groups of statistics using a quick recursive algorithm through the following formula

$$\begin{aligned}\hat{\beta}(k+1,t+1) &= \left(I - \mathbf{A}_{t+1}^{-1}\mathbf{A}_k\right)^{-1}\left(\hat{\beta}(1,t+1) - \mathbf{A}_{t+1}^{-1}\mathbf{A}_k\hat{\beta}(1,k)\right),\\ \hat{\mathbf{e}}(k+1,t+1) &= \mathbf{Y}_{k,t+1} - \mathbf{X}_{k,t+1}\hat{\beta}(k+1,t+1),\\ \hat{\sigma}^2(k+1,t+1) &= \left\|\hat{\mathbf{e}}(k+1,t+1)\right\|^2 / \sum_{j=k+1}^{t+1} n_j,\\ \hat{v}^2(k+1,t+1) &= \left\|\hat{\mathbf{e}}^2(k+1,t+1) - \hat{\sigma}^2(k+1,t+1)\mathbf{1}_{t+1-k}\right\|^2 / \sum_{j=k+1}^{t+1} n_j.\end{aligned}$$

which leads to computation complexity $\mathcal{O}(t^2)$ analyzed in detail in Remark 3.

**(II-3)** After the two steps above, two consistent estimators can be easily obtained for $\sigma_0^2, v^2$ under the null hypothesis $H_0$ as follows

$$\begin{aligned}\hat{\sigma}^2(1,k,t+1) &= \hat{\sigma}^2(1,k)\sum_{j=1}^{k} n_j / \sum_{j=1}^{t+1} n_j + \hat{\sigma}^2(k+1,t+1)\sum_{j=k+1}^{t+1} n_j / \sum_{j=1}^{t+1} n_j,\\ \hat{v}^2(1,k,t+1) &= \hat{v}^2(1,k)\sum_{j=1}^{k} n_j / \sum_{j=1}^{t+1} n_j + \hat{v}^2(k+1,t+1)\sum_{j=k+1}^{t+1} n_j / \sum_{j=1}^{t+1} n_j,\end{aligned}$$

which are also two excellent estimates even with an abrupt change $\tau$ because $k$ can be chosen as an estimate, as described in Remark 2.

**(II-4)** Compute the sequence $\mathrm{CW}_{k,t+1}, 1 \le k \le t+1$ and the final statistic $\mathrm{CW}_t$ In the above algorithm, we take a recursive computation method and avoid computing these statistics directly, especially calculating the inverse matrix $(\mathbf{X}_{1,t+1}'\mathbf{X}_{1,t+1})^{-1}$, which often results in high computation complexity.

**Remark 3:** To further understand the above algorithm, we will analyze its computational complexity for computing $\mathrm{CW}_{t+1}$ when the $(t+1)$-th data arrives. In Step II-1, the leading computational cost is caused by $\hat{\mathbf{e}}(1,t+1), \hat{\sigma}^2(1,t+1), \hat{v}^2(1,t+1)$ which all need to compute $t+1$ times basic

operations. Thus the computational complexity in Step II-1 is $\mathcal{O}(t)$. In Step II-2, we need $t$ iterations and the main single iteration cost lies in the update of $\hat{\mathbf{e}}(k+1,t+1), \hat{\sigma}^2(k+1,t+1), \hat{v}^2(k+1,t+1)$ which all need to compute $t+1-k$ times basic operations. The total computational complexity in Step II-2 will be $\sum_{k=1}^{t}(t+1-k) = t(t+1)/2 = \mathcal{O}(t^2)$.

### 3.3. Calculating control Limits

For a given in-control (IC) average run length (ARL), such as $1/\alpha$, the control limit $h_{t,\alpha}$ can be obtained by solving the following equation

$$\begin{aligned} &P\left(\mathrm{CW}_1 > h_{1,\alpha}\right) = \alpha, \\ &P\left(\mathrm{CW}_t > h_{t,\alpha} | \mathrm{CW}_k \le h_{k,\alpha}, 1 \le k \le t-1\right) = \alpha, t \ge 2. \end{aligned} \tag{5}$$

which is a probability of signal at process reading $t$, conditional on the absence of a signal prior to that moment (see Zou, Qiu, & Hawkins, 2008; Hawkins et al., 2003). Here, we do not assume the specific probability distribution of $\varepsilon_{ij}$ in model (1) and it is impossible to obtain an accurate form of $h_{t,\alpha}$ in (5). Fortunately, we infer the asymptotical distribution of $\mathrm{CW}_t$ in Theorems 1 and 2. As it is difficult to solve a $h_{t,\alpha}$ even based on a known asymptotical marginal distribution because there are no explicit expressions of joint probability distribution for $\max_{1\le k\le t-1} B_{t,k}^T B_{t,k} \Big/ \left(\frac{k}{t}\left(1-\frac{k}{t}\right)\right), t = 2,\ldots,$ we plan to obtain control limits $h_{t,\alpha}, t = 2,\ldots$ by simulation according to Theorem 1 based on a sequence of standard normal random variables $\{\xi_j\}$. We repeat the simulation 100,000 times to generate a sequence of $\xi_j \sim N(0,\mathbf{I}_3), j = 1,\ldots,500$ and for each repetition, we separately calculate $\max_{1\le k\le t-1} B_{t,k}^T B_{t,k} \Big/ \left(\frac{k}{t}\left(1-\frac{k}{t}\right)\right), t = 2,\ldots,500$ according to Theorem 1. The covering values $\alpha$ are set at $0.01, 0.005, 0.002$. Similarly, based on Theorem 2, we can also obtain control limits $h_{t,\alpha}^{(1)}$ and $h_{t,\alpha}^{(2)}$ for $\mathrm{CW}_t^{(1)}$ with $p = 2$ and $\mathrm{CW}_t^{(2)}$, respectively, which are included in Table 2. Some results in Tables 1 and 2 will be used for numerical study in the next section.

Based on our numerical experience and the empirical results in Hawkins et al. (2003) or Zou et al. (2008) which discuss similar simulation-based control limits, $h_{t,\alpha}, h_{t,\alpha}^{(1)}, h_{t,\alpha}^{(2)}$ are gradually stationary as $t$ increases as we observe in Tables 1 and 2. The control limits $h_{t,\alpha}$ in Table 1 will be used to simultaneously monitor possible shifts in coefficients ($p = 2$) or variance in the latter simulation. Results in Table 2 can be used in both simultaneous monitoring and diagnostic procedure as shown in real-data analysis.

### 3.4. Diagnostic procedure

After obtaining a signal through our proposed charting statistic, i.e. $\mathrm{CW}_t \ge h_t$, then a general common task is to diagnose which should be the reason of out-of-control signal.

- **Case 1** If $\mathrm{CW}_t^{(1)} \ge h_t^{(1)}$ and $\mathrm{CW}_t^{(2)} < h_t^{(2)}$, then we think signal comes from change-point in mean or regression coefficients of the linear profiles.
- **Case 2** If $\mathrm{CW}_t^{(1)} < h_t^{(1)}$ and $\mathrm{CW}_t^{(2)} \ge h_t^{(2)}$, then a shift in variance of the linear profiles occurs.
- **Case 3** If $\mathrm{CW}_t^{(1)} \ge h_t^{(1)}$ and $\mathrm{CW}_t^{(2)} \ge h_t^{(2)}$, then we think a common change-point occurs simultaneously in mean and variance of the linear profiles.

## 4. Simulations and applications

### 4.1. Monte Carlo simulation

In this subsection, we present numerical results about the performance of our method based on different structural change models. The linear profile with possible shifts in coefficients and variance at change-point $\tau$ is designed as the following form

**Table 1.** Control limit $h_{t,\alpha}$ for time $t$ probability of type 1 error $\alpha$.

| | Threshold $h_{t,\alpha}$ | | | |
|---|---|---|---|---|
| $t$ | 0.01 | 0.005 | 0.002 | 0.001 |
| 2 | 11.375 | 12.889 | 14.696 | 16.218 |
| 3 | 12.125 | 13.551 | 15.592 | 16.920 |
| 4 | 12.577 | 14.139 | 16.345 | 17.745 |
| 5 | 12.902 | 14.521 | 16.608 | 18.115 |
| 6 | 13.159 | 14.706 | 16.726 | 18.211 |
| 7 | 13.211 | 14.899 | 16.937 | 18.662 |
| 8 | 13.364 | 14.964 | 17.098 | 18.797 |
| 9 | 13.525 | 15.227 | 17.383 | 18.782 |
| 10 | 13.500 | 15.219 | 17.354 | 18.893 |
| 15 | 13.785 | 15.534 | 17.753 | 19.391 |
| 20 | 13.786 | 15.578 | 18.051 | 19.601 |
| 25 | 13.960 | 15.717 | 17.975 | 19.663 |
| 30 | 13.931 | 15.756 | 17.892 | 19.450 |
| 35 | 14.037 | 15.799 | 17.932 | 19.543 |
| 40 | 13.998 | 15.878 | 18.221 | 19.852 |
| 45 | 13.978 | 15.762 | 18.088 | 19.689 |
| 50 | 13.992 | 15.803 | 18.095 | 19.587 |
| 60 | 14.285 | 15.957 | 18.216 | 19.889 |
| 70 | 14.065 | 15.860 | 18.321 | 20.152 |
| 80 | 13.962 | 15.875 | 18.176 | 19.828 |
| 90 | 14.002 | 15.760 | 18.130 | 20.062 |
| 100 | 13.955 | 15.874 | 18.067 | 19.762 |
| 200 | 14.039 | 16.027 | 18.318 | 20.060 |
| 500 | 14.287 | 16.030 | 18.243 | 19.922 |

Note: The number in table is based on $1.0 \times 10^5$ repetitions.

**Table 2.** Control Limits $h_{t,\alpha}^{(1)}, h_{t,\alpha}^{(2)}$ for $CW_t^{(1)}, CW_t^{(2)}$, respectively.

| | Threshold $h_{t,\alpha}^{(1)}$ | | | Threshold $h_{t,\alpha}^{(2)}$ | | |
|---|---|---|---|---|---|---|
| $t$ | 0.01 | 0.005 | 0.002 | 0.01 | 0.005 | 0.002 |
| 2 | 6.651 | 7.959 | 9.532 | 9.256 | 10.611 | 12.409 |
| 3 | 7.224 | 8.536 | 10.414 | 9.882 | 11.294 | 13.174 |
| 4 | 7.601 | 8.947 | 10.885 | 10.281 | 11.783 | 13.813 |
| 5 | 7.848 | 9.196 | 11.107 | 10.578 | 12.060 | 13.946 |
| 6 | 8.025 | 9.416 | 11.379 | 10.771 | 12.362 | 14.377 |
| 7 | 8.038 | 9.499 | 11.232 | 10.853 | 12.435 | 14.520 |
| 8 | 8.219 | 9.754 | 11.602 | 11.028 | 12.501 | 14.429 |
| 9 | 8.196 | 9.690 | 11.643 | 11.167 | 12.853 | 14.710 |
| 10 | 8.217 | 9.636 | 11.478 | 11.194 | 12.693 | 14.787 |
| 20 | 8.425 | 9.921 | 12.073 | 11.388 | 12.996 | 15.168 |
| 30 | 8.603 | 10.108 | 11.929 | 11.563 | 13.221 | 15.372 |
| 40 | 8.600 | 10.189 | 12.118 | 11.546 | 13.315 | 15.484 |
| 50 | 8.494 | 10.130 | 12.026 | 11.571 | 13.250 | 15.409 |
| 100 | 8.605 | 10.066 | 12.095 | 11.606 | 13.261 | 15.498 |
| 200 | 8.688 | 10.245 | 12.489 | 11.692 | 13.189 | 15.408 |
| 500 | 8.058 | 10.321 | 12.235 | 11.568 | 13.547 | 15.515 |

Note: The number in table is based on $1.0 \times 10^5$ repetitions.

$$y_{ij} = \begin{cases} (1, x_{ij})\beta + \sigma_0 \varepsilon_{ij}, & i = 1, \ldots, n,\ j = 1, \ldots, \tau, \\ (1, x_{ij})\beta + (1, x_{ij})\delta + \sigma_1 \varepsilon_{ij}, & i = 1, \ldots, n,\ j = \tau + 1, \tau + 2, \ldots, \end{cases} \tag{6}$$

where $\beta = (2, 2)^T$ denotes two-dimensional initial regression coefficients which is set fixed, and $\delta = (\delta_1, \delta_2)^T$ denotes possible coefficients shifts which will be set varying, and $\varepsilon_{ij} \sim N(0, 1)$. We set $\sigma_0 = 1, n = 10$ and $x_{ij}$ equal-spaced in $[-3, 3], i = 1, \ldots, n$, i.e. $x_{ij} = -3, -2.33, \ldots, 2.33, 3$.

**Table 3.** IC and OC ARL with $\alpha = 0.005$ and $\sigma_0 = 1$.

| | | | $\tau$ | | | |
|---|---|---|---|---|---|---|
| $(\delta_1, \delta_2)$ | $\sigma_1 - \sigma_0$ | Charts | 5 | 20 | 50 | 100 |
| (0, 0) | 0 | CW | 200.05 | 200.00 | 200.67 | 200.78 |
| | | | (31.18) | (31.18) | (31.20) | (31.18) |
| | | SSMax-CUSUM | 200.03 | 200.03 | 200.38 | 200.90 |
| | | | (35.24) | (35.24) | (35.24) | (34.90) |
| | | EWMA | 200.06 | 200.08 | 200.74 | 200.31 |
| | | | (30.11) | (30.10) | (30.07) | (30.10) |
| (0.2, 0.2) | 0 | CW | 89.13 | 10.45 | 7.39 | 6.57 |
| | | | (14.76) | (9.47) | (5.70) | (4.29) |
| | | SSMax-CUSUM | 165.31 | 153.37 | 115.86 | 72.67 |
| | | | (23.48) | (17.39) | (16.36) | (8.41) |
| | | EWMA | 148.46 | 115.03 | 50.17 | 23.41 |
| | | | (20.55) | (16.01) | (10.03) | (7.47) |
| (0.5, 0.5) | 0 | CW | 1.82 | 1.73 | 1.80 | 1.74 |
| | | | (1.16) | (0.67) | (0.65) | (0.56) |
| | | SSMax-CUSUM | 89.85 | 37.07 | 10.03 | 4.59 |
| | | | (17.14) | (16.77) | (3.79) | (3.71) |
| | | EWMA | 70.49 | 7.01 | 3.22 | 3.08 |
| | | | (14.66) | (2.55) | (1.57) | (1.55) |
| (0, 0) | 1 | CW | 2.51 | 1.43 | 1.39 | 1.42 |
| | | | (1.29) | (0.59) | (0.68) | (0.61) |
| | | SSMax-CUSUM | 33.01 | 18.75 | 9.30 | 4.32 |
| | | | (8.82) | (6.81) | (5.23) | (4.83) |
| | | EWMA | 3.13 | 1.78 | 1.64 | 1.45 |
| | | | (1.39) | (0.89) | (0.77) | (0.69) |
| (0, 0) | 2 | CW | 1.13 | 1.04 | 1.02 | 1.04 |
| | | | (0.33) | (0.17) | (0.14) | (0.14) |
| | | SSMax-CUSUM | 8.25 | 3.60 | 1.26 | 1.13 |
| | | | (10.25) | (9.79) | (0.90) | (0.61) |
| | | EWMA | 1.16 | 1.08 | 1.08 | 1.05 |
| | | | (0.34) | (0.19) | (0.17) | (0.24) |
| (0.5, 0.5) | 2 | CW | 1.09 | 1.04 | 1.04 | 1.01 |
| | | | (0.21) | (0.10) | (0.10) | (0.08) |
| | | SSMax-CUSUM | 3.66 | 1.10 | 1.10 | 1.27 |
| | | | (0.82) | (0.21) | (0.20) | (0.15) |
| | | EWMA | 1.09 | 1.04 | 1.04 | 1.05 |
| | | | (0.22) | (0.14) | (0.14) | (0.09) |

Note: ARL and corresponding standard deviation (in parentheses) in table are based on $1.0 \times 10^5$ repetitions.

In Table 3, we will compare different methods' performance measured by out-of-control average run length (OC ARL) given the rough same in-control average run length (IC ARL). The detection performance of our statistic can be usually measured by out-of-control the average run length (OC ARL). The key varying parameters, $\delta = (\delta_1, \delta_2)^T, \sigma_1 - \sigma_0$, are set by four different cases: (a) no shifts in coefficients or variance, (b) only shift in coefficients, (c) only shift in variance, and (d) simultaneous shifts in both. Further, we set $\tau = 5, 20, 50, 100$ and $\alpha = 0.005$ which indicates corresponding IC-ARL=200. For $p = 2$ in our data model, we choose appropriate control limits $h_{t,\alpha}$ from Table 1. Different performances of our proposed method, SSMax-CUSUM by Amiri et al. (2016a) and EWMA by Zou et al. (2007) are list in Table 3 and all results are based on 1000 repetitions.

From the results in Table 3, we observe the following conclusions:

(1) When $\delta = (0, 0)$ and $\sigma_1 - \sigma_0 = 0$, i.e. no shift in coefficients or variance, all IC-ARLs are roughly 200. These results show our asymptotic control limits (5) performs well. Thus, it is fair to compare different methods under the same IC-ARL.
(2) When small shift $\delta = (0.2, 0.2)^T$ occurs only in coefficient, i.e. $\sigma_1 - \sigma_0 = 0$, ARL of CW is smaller than that of the other two methods, which seems to show our method is more sensitive to small jump size $\delta$. It is a reasonable result because CW is designed to make use of all historical

**Table 4.** Correct diagnosis-proportion.

| | | $\tau$ | | | |
|---|---|---|---|---|---|
| $(\delta_1, \delta_2)$ | $\sigma_1 - \sigma_0$ | 5 | 20 | 50 | 100 |
| (0.2, 0.2) | 0 | 0.82 | 0.98 | 0.98 | 0.99 |
| (0.5, 0.5) | 0 | 0.98 | 0.98 | 0.98 | 0.99 |
| (0, 0) | 1 | 0.73 | 0.92 | 0.93 | 0.93 |
| (0, 0) | 2 | 0.92 | 0.98 | 1.00 | 1.00 |
| (0.5, 0.5) | 2 | 0.99 | 0.99 | 1.00 | 1.00 |

Note: The number in table is based on $1.0 \times 10^5$ repetitions.

data and obtain more information. When $\delta$ increase from $(0.2, 0.2)^T$ to $(0.5, 0.5)^T$), all ARLs perform better.

(3) When small shift $\sigma_1 - \sigma_0 = 1$ occurs only in variance, performances of three methods seems to be the same. Actually in most cases numerical results of CW seem to be a little better except for $\tau = 5$. When $\sigma_1 - \sigma_0$ increases from 1 to 2, there is no significant difference among these methods.

(4) When simultaneous shifts occur in both coefficients and variance, i.e. $\delta = (0.5, 0.5)^T, \sigma_1 - \sigma_0 = 2$, ARL of CW is a little smaller than others.

Once obtaining a signal by our proposed charting statistic, i.e. $\text{CW}_t \geq h_t$, then a further interesting target is to use our diagnosis procedure to determine possible reasons of out-of-control signal: (a). coefficients shift, (b). variance shift or (c) both of them. For evaluating the performance of diagnosis procedure, we run a simulation study under the same conditions considered in Table 3 based on $1.0 \times 10^5$ repetitions and calculate the corresponding the correct proportion of reason-choosing. All results are reported in Table 4. For example, under $\tau = 5, \delta_1 = 0.2, \delta_2 = 0.2, \sigma_1 - \sigma_0 = 0$, the correct diagnosis proportion, 0.82, means to choose the reason of coefficients shifts for $0.82 \times 10^5$ times during $10^5$ repetitions. Results in Table 4 show that our diagnosis performs well.

Based on results in Table 3, we can compare performances between CW and other methods more intuitively by graphs. Results in Table 3 show that performances of EWMA and SSMax-CUSUM are roughly the same especially under the case of small jump sizes. For simplicity, we only choose EWMA to compare with our method, CW, as shown in Figure 1 which describes the relationship between ln ARL and varying shifts in coefficients or variance of model (6). In the left figure (a), we set $\tau = 20, \delta_1 = 0, \sigma_1 - \sigma_0 = 0$ and let $\delta_2$ varies in $[-1, 1]$. In the right figure (b), we set the same $\tau = 20$ and $\delta = (0, 0)^T$ and let $\ln r = \ln \sigma_1/\sigma_0$ varies in $[-1, 1]$. In the left figure (a), which shows the performance for different $\delta$ values, we can observe the following results.

(1) When the shift in $\delta_2$ is small, such as $\delta_2 \in [-0.25, 0.25]$, EWMA is not sensitive enough for the small shift $\delta_2$, which is very similar conclusion to Table 3. However, CW performs better consistently among the whole interval.

(2) When shift in $\delta_2$ is large, both EWMA and CW perform well, although ALRs of EWMA is slightly larger than those of CW. The right figure (b) demonstrates performances for different values of $r = \sigma_1/\sigma_0$. When $\ln r \leq 0$, or equivalently $\sigma_1 \leq \sigma_0$, EWMA cannot detect the shift, as is also pointed out by Zou et al. (2007). However, when $\ln r > 0$, both EWMA and CW performs well although there exists slight difference.

To our best knowledge, existing methods for simultaneously monitoring of coefficients and variance of linear profiles depend on the assumption of normal errors which restricts their applications in practice. Our method, CW, can be used in more general case based on Theorems 1 and 2. To demonstrate CW's performance under non-normal errors, we consider the following data-generating process with mixture-of-Gaussian errors (MOG-errors)

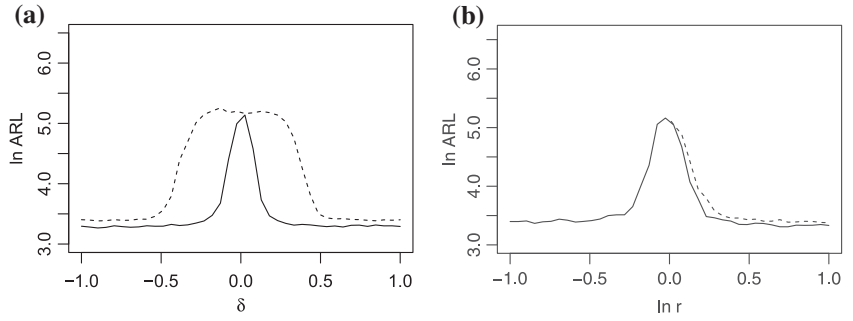


**Figure 1.** ARL for different shifts in coefficients (a) or variance (b) based on EWMA (dotted line) and CW (solid line).

**Table 5.** IC and OC ARL under MOG-errors.

| | | $\tau$ | | | |
|---|---|---|---|---|---|
| $(\delta_1, \delta_2)$ | $\sigma_1 - \sigma_0$ | 5 | 20 | 50 | 100 |
| (0, 0) | 0 | 200.91 | 200.20 | 200.66 | 200.13 |
| | | (20.08) | (20.08) | (20.08) | (20.08) |
| (0.2, 0.2) | 0 | 3.11 | 2.26 | 2.29 | 2.19 |
| | | (1.99) | (1.89) | (1.66) | (1.56) |
| (0.5, 0.5) | 0 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | (0.21) | (0.20) | (0.17) | (0.15) |
| (0, 0) | 1 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | (0.14) | (0.10) | (0.00) | (0.00) |
| (0, 0) | 2 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | (0.00) | (0.00) | (0.00) | (0.00) |
| (0.5, 0.5) | 2 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | (0.00) | (0.00) | (0.00) | (0.00) |

Note: ARL and corresponding standard deviation (in parentheses) in table are based on $1.0 \times 10^5$ repetitions.

$$y_{ij} = \begin{cases} (1, x_{ij})\beta + e_{ij}, & i = 1, \ldots, n,\ j = 1, \ldots, \tau, \\ (1, x_{ij})\beta + (1, x_{ij})\delta + e_{ij}, & i = 1, \ldots, n,\ j = \tau + 1, \tau + 2, \ldots \end{cases} \tag{7}$$

where $\beta = (2, 2)^T$ and errors $e_{ij}$ obey mixture of Gaussian, i.e.

$$e_{ij} \sim \begin{cases} 0.5N(0, \sigma_0^2) + 0.5N(0, \tilde{\sigma_0}^2), & i = 1, \ldots, n,\ j = 1, \ldots, \tau, \\ 0.5N(0, \sigma_1^2) + 0.5N(0, \tilde{\sigma_1}^2), & i = 1, \ldots, n,\ j = \tau + 1, \tau + 2, \ldots \end{cases} \tag{8}$$

with $\tilde{\sigma_0} = \sigma_0/2, \tilde{\sigma_1} = \sigma_1/2$ for simplicity. Obviously $e_{ij}$ are non-normal because $P(e_{ij} < u) = 0.5\Phi(u; 0, \sigma_0^2) + 0.5\Phi(u; 0, \tilde{\sigma_0}^2), j \le \tau$ or $0.5\Phi(u; 0, \sigma_1^2) + 0.5\Phi(u; 0, \tilde{\sigma_1}^2), j \ge \tau + 1$ where $\Phi(u; \mu, \sigma^2)$ denotes the cumulative probability distribution with mean $\mu$ and variance $\sigma^2$.

Similar to Table 3, we also set $\beta = (2, 2)^T, \sigma_0 = 1$ and $\alpha = 0.005$. TO satisfy assumptions in Theorem 1 under non-normal errors, we set a larger sample size, $n = 30$, and $x_{ij}$ equal-spaced in $[-3, 3]$. Performances of CW under different $\delta, \sigma_1 - \sigma_0$ and $\tau$ are shown in Table 5 and all results are based on $10^5$ repetitions.

From Table 5, we observe some interesting phenomena.

(1) The first interesting phenomena is that the corresponding OC ARL In Table 5 is smaller than that in Table 3 under the same $(\delta_1, \delta_2)$ and $\sigma_1 - \sigma_0$, especially OC ARL=3.11 $\ll$ 89.13 when $(\delta_1, \delta_2) = (0.2, 0.2), \sigma_1 - \sigma_0 = 0, \tau = 5$. Actually, it is not difficult to interpret this observation

**Table 6.** $y_{ij}^{(r)}$ for in-control DRIE profile.

| $j$ | $y_{ij}$ (unit: $\mu$m) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3.47 | 1.95 | 1.43 | 0.38 | −0.20 | −0.32 | 0.20 | 0.57 | 1.50 | 2.64 | 4.02 |
| 2 | 3.92 | 2.16 | 1.53 | 0.03 | 0.15 | −0.22 | 0.70 | 0.62 | 1.16 | 1.68 | 3.98 |
| 3 | 4.04 | 2.68 | 1.10 | 0.51 | 0.26 | 0.17 | −0.56 | 0.92 | 1.15 | 1.69 | 3.71 |
| 4 | 3.76 | 2.54 | 1.91 | 1.21 | −0.07 | −0.09 | 0.06 | 0.63 | 1.17 | 2.26 | 4.05 |
| 5 | 2.75 | 2.82 | 1.26 | 0.92 | 0.11 | −0.71 | −0.02 | 1.39 | 1.14 | 1.80 | 3.75 |
| 6 | 3.60 | 3.00 | 1.02 | 0.55 | −0.09 | 1.24 | −0.10 | 0.53 | 1.51 | 2.20 | 4.28 |
| 7 | 3.70 | 1.55 | 1.51 | 0.48 | −0.95 | −0.37 | −0.18 | 0.47 | 1.65 | 2.56 | 3.91 |
| 8 | 3.69 | 2.78 | 0.30 | 0.53 | 0.93 | −0.10 | 0.62 | 1.00 | 1.68 | 2.62 | 4.05 |
| 9 | 3.94 | 3.31 | 2.21 | 0.11 | 0.50 | −0.84 | 0.13 | 0.90 | 1.30 | 2.42 | 3.55 |
| 10 | 3.72 | 3.01 | 0.96 | 0.40 | −0.27 | 0.75 | 0.07 | 0.27 | 1.36 | 2.46 | 3.72 |
| 11 | 3.97 | 2.89 | 1.61 | 0.68 | 0.52 | 0.36 | −0.45 | 0.46 | 0.80 | 1.99 | 3.35 |
| 12 | 3.89 | 2.12 | 1.15 | 0.94 | 0.13 | 0.09 | −0.01 | 0.68 | 1.77 | 2.83 | 4.26 |
| 13 | 4.23 | 2.34 | 1.32 | 0.72 | 0.13 | 0.68 | 1.02 | 0.03 | 1.35 | 3.20 | 4.34 |
| 14 | 3.95 | 2.20 | 1.02 | 0.78 | 0.64 | 0.52 | 0.01 | 0.83 | 1.23 | 2.39 | 3.72 |
| 15 | 4.28 | 2.97 | 1.41 | 0.48 | 0.49 | −0.84 | −0.07 | 0.26 | 1.09 | 3.06 | 4.08 |
| 16 | 4.18 | 3.23 | 1.74 | 0.70 | 0.31 | 0.53 | −0.36 | 0.58 | 1.38 | 2.73 | 4.32 |
| 17 | 4.35 | 2.44 | 1.61 | 0.84 | 1.21 | 0.30 | 0.49 | 1.30 | 1.31 | 2.14 | 4.01 |
| 18 | 3.19 | 2.71 | 1.20 | 0.57 | 0.31 | −0.51 | 0.49 | -0.75 | 1.78 | 1.53 | 4.30 |

Source: Adapted from Zou et al. (2007a).

by calculating the variance of error $e_{ij}$ under $\sigma_0 = 1$

$$\mathrm{Var}(e_{ij}) = 0.5 \times \sigma_0^2 + 0.5 \times \tilde{\sigma}_0^2 = 5/8 < \sigma_0^2 = 1, \quad j \le \tau$$

which indicates smaller error variance and thus make the out-of-control signal be easily found by our method.

(2) We also observe that when $\delta = (0,0)^T, \sigma_1 - \sigma_0 = 0$, all of IC-ARLs approximate to 200, which is consistent to the theoretical conclusion indicated by Theorem 1 and corresponding control limits in Table 1. The reason is that MOG-errors satisfy the key assumption **A3** and thus asymptotical distribution in Theorem 1 performs well.

### 4.2. An example: deep reactive ion etching process

Zou et al. (2007) used the example of deep reactive ion etching (DRIE), a process used in semiconductor manufacturing. One of the most important characteristics of a semiconductor is the profile of the trenches, with smooth and vertical sidewalls strongly preferred. Other profile shapes due to underetching and overetching are unacceptable.

It is impossible to model the entire DRIE profile using a general linear profile. Engineers believe that the left and right corners of each profile contain the information needed to identify out-of-control conditions. Because the corners are symmetrical, one corner can be used to monitor the process. After rotating the left corner of the profile by $45^o$ counterclockwise, Zou et al. (2007) modelled the profile using a quadratic polynomial model

$$y_{ij} = \beta_0 \times x_i^2 + \sigma_0 \varepsilon_{ij}, \;\; i = 1, \ldots, 11, \; j = 1, \ldots, 18$$

where $x_i$s are equally spaced as $-2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2$, and 2.5, and $y_{ij}$s are corresponding readings of the rotated corner curve.

Based on in-control real data $y_{ij}^{(r)}, i = 1, \ldots, 11, j = 1, \ldots, 18$ are listed in Table 6, Zou et al. (2007) obtained an estimated in-control model $y_{ij}^{(r)} = 0.62x_i^2 + \sigma_0 \varepsilon_{ij}, j = 1, \ldots, 18$ with $\sigma_0 = 0.4$. As out-of-control samples are difficult and expensive to produce in a laboratory, so Zou et al. (2007) artificially increased the curvature coefficient from 0.62 to 0.67. The whole linear profile model can

**Table 7.** $y_{ij}^{(a)}$ for DRIE profile with shifts in coefficients and variance.

| $j$ | $y_{ij}$ (unit: $\mu$m) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4.59 | 2.77 | 2.26 | 2.53 | 1.15 | 1.25 | 1.24 | 2.40 | 2.94 | 3.36 | 4.74 |
| 2 | 5.06 | 3.29 | 2.21 | 1.95 | 0.67 | 1.37 | 0.86 | 1.23 | 1.96 | 2.99 | 5.07 |
| 3 | 4.79 | 3.82 | 2.38 | 1.12 | 1.03 | 1.55 | 1.08 | 1.37 | 2.66 | 3.71 | 5.53 |
| 4 | 4.76 | 3.45 | 2.18 | 0.85 | 1.60 | 1.69 | 1.41 | 1.61 | 2.00 | 3.24 | 4.95 |
| 5 | 4.71 | 3.24 | 2.38 | 2.28 | 1.60 | 1.69 | 1.19 | 1.59 | 3.07 | 3.80 | 4.46 |
| 6 | 6.09 | 3.65 | 2.67 | 2.36 | −0.35 | 0.65 | 0.01 | 1.84 | 2.11 | 3.14 | 4.79 |
| 7 | 4.73 | 3.12 | 3.33 | 2.96 | 1.15 | 1.43 | 1.07 | 2.63 | 3.42 | 3.08 | 6.05 |
| 8 | 5.43 | 4.28 | 3.19 | 1.85 | 1.45 | 0.95 | 1.47 | 1.19 | 1.83 | 3.62 | 4.84 |
| 9 | 5.16 | 3.90 | 2.87 | 0.92 | 1.24 | 0.87 | 0.51 | 1.81 | 2.98 | 4.43 | 6.30 |
| 10 | 5.10 | 4.12 | 2.72 | 2.20 | 1.36 | 1.44 | 1.68 | 2.15 | 3.14 | 4.00 | 5.46 |
| 11 | 5.56 | 4.28 | 2.55 | 1.43 | 0.27 | 0.36 | 1.30 | 2.60 | 3.27 | 4.03 | 4.85 |
| 12 | 6.01 | 2.23 | 2.70 | 1.96 | 1.18 | 1.06 | 1.40 | 2.05 | 2.24 | 3.95 | 6.34 |
| 13 | 5.69 | 3.38 | 3.10 | 1.75 | 1.77 | 0.53 | 0.22 | 1.36 | 1.83 | 3.42 | 5.14 |
| 14 | 6.69 | 3.79 | 2.10 | 1.89 | 1.06 | 1.51 | 2.11 | 2.74 | 2.20 | 4.24 | 5.70 |

**Table 8.** Using $CW_t$ to monitor and diagnose DRIE profiles simultaneously.

| | | | $CW_{k,t}, k = 1, 2, \ldots, t-1$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | $CW_t$ | $h_t^{(1)}$ | $t = 24$ | $t = 25$ | $t = 26$ | $t = 27$ | $t = 28$ | $t = 29$ | $t = 30$ | $t = 31$ | $t = 32$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 15 | 4.50 | 11.30 | 0.13 | 0.50 | 1.65 | 3.73 | 4.04 | 5.13 | 6.268 | 10.29 | 12.66 |
| 16 | 3.82 | 11.40 | 1.23 | 1.98 | 3.99 | 6.90 | 6.79 | 8.13 | 9.535 | 14.63 | 17.50 |
| 17 | 4.58 | 11.42 | 1.64 | 2.67 | 5.07 | 8.67 | 8.88 | 10.45 | 12.03 | 17.81 | 21.02 |
| 18 | 3.68 | 11.37 | 2.08 | 3.10 | 5.89 | 9.77 | 9.41 | 11.05 | 12.79 | 19.45 | 23.01 |
| 19 | 4.24 | 11.33 | 2.68 | 4.09 | 7.37 | 12.20 | 12.36 | 14.26 | 16.20 | 23.76 | 27.70 |
| 20 | 3.62 | 11.38 | 3.63 | 5.03 | 8.92 | 14.10 | 13.42 | 15.36 | 17.44 | 26.03 | 30.29 |
| 21 | 4.43 | 11.41 | 6.23 | 8.07 | 12.67 | 19.18 | 18.97 | 21.07 | 23.35 | 33.23 | 37.86 |
| 22 | 4.51 | 11.52 | 12.12 | 13.14 | 18.78 | 26.02 | 24.27 | 26.29 | 28.71 | 40.29 | 45.33 |
| 23 | 5.59 | 11.47 | 17.99 | 17.74 | 22.06 | 30.47 | 31.21 | 32.93 | 35.24 | 47.92 | 52.94 |
| 24 | 17.99 | 11.56 | | 1.78 | 6.57 | 13.54 | 13.82 | 15.92 | 18.21 | 29.99 | 34.77 |
| 25 | 17.75 | 11.45 | | | 5.76 | 12.81 | 12.94 | 14.93 | 17.20 | 30.06 | 34.90 |
| 26 | 22.07 | 11.50 | | | | 7.85 | 11.32 | 12.74 | 14.57 | 27.99 | 32.55 |
| 27 | 30.48 | 11.52 | | | | | 5.39 | 6.47 | 7.96 | 20.83 | 25.13 |
| 28 | 31.22 | 11.51 | | | | | | 2.21 | 4.57 | 19.07 | 24.00 |
| 29 | 32.93 | 11.54 | | | | | | | 2.40 | 18.58 | 23.36 |
| 30 | 35.24 | 11.56 | | | | | | | | 18.07 | 22.28 |
| 31 | 47.92 | 11.50 | | | | | | | | | 4.94 |
| 32 | 52.94 | 11.64 | | | | | | | | | |

be described by the following form

$$y_{ij} = \begin{cases} 0.62 \times x_i^2 + 0.4\varepsilon_{ij}, & i = 1, \ldots, 11,\ j = 1, \ldots, 23, \\ 0.67 \times x_i^2 + 0.6\varepsilon_{ij}, & i = 1, \ldots, 11,\ j = 19, \ldots, 32 \end{cases} \tag{9}$$

To test the efficiency of our monitoring method at detecting coefficients and variance simultaneously, we take similar steps in creating a new group of data with increases in both coefficients from 0.62 to 0.67 and variance from 0.4 to 0.6 after the fifth sample. The simulated data $y_{ij}^{(a)}, i = 1, \ldots, 11, j = 1, \ldots, 14$ are given in Table 7.

In the following, we set $\alpha = 0.01$ and use our charting statistic $CW_t$ to monitor whether there exists a shift or not and further estimate the specific shift location. All results are summarized in Table 8. From these results, we can obtain several observations:

(1) The 24-th data point is the first data after simultaneous shifts. Our method can quickly give a signal at this point because 17.99 > 11.56 owing to the use of global data and thus has better efficiency.
(2) From the 24-th to 32-th time points, CW signals continuously, which means that our method can always remember the shift historical information. But some methods only using local data may miss shift information after a long time.
(3) After signalling at time $t$, we can also identify potential change-point by observing $\mathrm{CW}_{k,t}$ from $k = 1$ to $k = t - 1$. For example, after our method gives a signal at the 24-th point, we can further observe the $\mathrm{CW}_{k,t}, k = 1, 2, \ldots, 23$ and find the maximum value at the 23-th point which indicates $\hat{\tau} = 23$.

After signal by CW method, we can use $\mathrm{CW}_t^{(1)}, \mathrm{CW}_t^{(2)}$ to judge which reason, mean or variance, leads to this alarm. In this real-data analysis, we calculate $\mathrm{CW}_{24}^{(1)} = 12.19 > h_{24}^{(2)} = 11.56, \mathrm{CW}_{24}^{(2)} = 22.80 > h_{24}^{(2)} = 11.56$ which indicates both of mean and variance cause this alarm.

## 5. Conclusion

The aim of the study in this paper is to propose a new method for sequentially detecting shifts in regression coefficients and variance simultaneously. Once the out-of-control state signals, our method can be also used to locate the specific out-of-control time point by calculating the maximum value point. Finally, we can diagnose which reason, coefficients or variance, leads to the alarm signalled by our charting statistic. But our method are based on the large-sample assumption to make sure that the asymptotic distribution for charting statistic works, which is sometimes easy to satisfy under normal errors and is sometimes too strict in practice, especially under non-normal errors. How to loosen these requirements becomes one of our researchs in future.

## Acknowledgements

The authors are grateful for constructive comments from the editor and all referees.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

Prof. Tsung was supported by RGC GRF [grant number 16203917]; Prof. Xia was supported by National Natural Science Funds of China [grant number 11771353], [grant number 11201372]; Research Grants Council, University Grants Committee [grant number 619913].

## Notes on contributors

***Zhiming Xia*** is Head of applied statistics teaching and research section at Northwest University and is interested in change-point analysis and applied statistics.

***Fugee Tsung*** is Academician of IAQ, Fellow of ASA, ASQ, IISE, and HKIE and is interested in the research of quality control and data analytics.

## References

Amiri, A., Ghashghaei, R., & Khosravi, P. (2016b). *A self-starting control chart for simultaneous monitoring of mean and variance of autocorrelated simple linear profile*. 2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Tehran, 2016 Dec 4 (pp. 209–213).

Amiri, A., Khosravi, P., & Ghashghaei, R. (2016a). A self-starting control chart for simultaneous monitoring of mean and variance of simple linear profiles. *International Journal of Engineering-Transactions C: Aspects, 29*(9), 1263–1272.
Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of Royal Statistical Society B, 37*, 149–192.
Chen, G. M. (1997). The mean and standard deviation of the run length distribution of $\overline{X}$ charts when control limits are estimated. *Statistica Sinica, 7*(3), 789–798.
Croarkin, C., & Varner, R. (1982). *Measurement Assurance for Dimensional Measurements on Integrated-Circuit Photomasks* (NBS Technical Note 1164). Washington DC, USA: U.S. Department of Commerce.
Csorg, M., & Horvath, L. (1997). *Limit theorems in change-point analysis*. Chichester: Wiley.
Ghosh, B. K., Reynolds, M. R., Jr, & Hui, Y. V. (1981). Shewhart X-charts with estimated process variance. *Communications in Statistics-Theory and Methods, 10*(18), 1797–1822.
Hawkins, D. M. (1987). Self-starting cusum charts for location and scale. *The Statistician, 36*(4), 299–316.
Hawkins, D. M., Qiu, P., & Kang, C. W. (2003). The change-point model for statistical process control. *Journal of Quality Technology, 35*(4), 355–366.
Hawkins, D. M., & Zamba, K. D. (2005a). A change-point model for a shift in variance. *Journal of Quality Technology, 37*(1), 21–31.
Hawkins, D. M., & Zamba, K. D. (2005b). Statistical process control for shifts in mean or variance using a change-point formulation. *Technometrics, 47*(2), 164–173.
Kang, L. & Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *Journal of Quality Technology, 32*(4), 418–426.
Kazemzadeh, R. B., Amiri, A., & Kouhestani, B. (2016). Monitoring simple linear profiles using variable sample size schemes. *Journal of Statistical Computation and Simulation, 86*(15), 2923–2945.
Kazemzadeh, R. B., Noorossana, R., & Amiri, A. (2009). Monitoring polynomial profiles in quality control applications. *The International Journal of Advanced Manufacturing Technology, 42*(7), 703–712.
Khedmati, M., & Niaki, S. T. A. (2016a). Monitoring simple linear profiles in multistage processes by a Max-EWMA control chart. *Computers & Industrial Engineering, 98*(c), 125–143.
Khedmati, M., & Niaki, S. T. A. (2016b). A new control scheme for phase-II monitoring of simple linear profiles in multistage processes. *Quality and Reliability Engineering International, 32*(7), 2559.
Kim, K., Mahmoud, M. A., & Woodall, W. H. (2003). On the monitoring of linear profiles. *Journal of Quality Technology, 35*(3), 317–328.
Li, Z., Zhang, J., & Wang, Z. (2010). Self-starting control chart for simultaneously monitoring process mean and variance. *International Journal of Production Research, 48*(15), 4537–4553.
Magalhaes, M. S. D., & Von Doellinger, R. O. S. (2016). Monitoring linear profiles using an adaptive control chart. *The International Journal of Advanced Manufacturing Technology, 82*(5–8), 1433–1445.
Mahmoud, M. A. (2004). The monitoring of linear proles and the inertial properties of control charts. Blacksburg, VA: Virginia Polytechnic Institute and University State (Virginia Tech).
Mahmoud, M. A., & Woodall, W. H. (2004). Phase I analysis of linear profiles with calibration applications. *Technometrics, 46*(4), 380–391.
Niaki, S. T. A., Abbasi, B., & Arkat, J. (2007). A generalized linear statistical model approach to monitor profiles. *International Journal of Engineering, Transactions A: Basics, 20*(3), 233–242.
Noorossana, R., & Amiri, A. (2007). Enhancement of linear profiles monitoring in Phase II. *AmirKabir Journal of Science and Technology, 18*(66-B), 19–27. Persian.
Noorossana, R., Amiri, A., Vaghefi, S. A., & Roghanian, E. (2004). *Monitoring quality characteristics using linear profile*. Proceedings of the 3rd International Industrial EngineeringConference. Tehran, Iran.
Puntanen, S., Styan, G. P. H., & Isotalo, J. (2011). *Matrix tricks for linear statistical models*. Berlin: Springer-Verlag.
Quesenberry, C. (1993). The effect of sample size on estimated limits for $\overline{X}$ and $X$ control charts. *Journal of Quality Technology, 25*(4), 237–247.
Saghaei, A., Mehrjoo, M., & Amiri, A. (2009). A CUSUM-based method for monitoring simple linear profiles. *The International Journal of Advanced Manufacturing Technology, 45*(11), 1252–1260.
Tsung, F., Zhou, Z. H., & Jiang, W. (2007). Applying manufacturing batch techniques to fraud detection with incomplete customer information. *IIE Transactions, 39*(6), 671–680.
Woodall, W.H. (2007). The use of control charts in health-care and public-health surveillance. *Journal of Quality Technology*, *52*(2), 253–256.
Woodall, W. H., & Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology, 31*(4), 376–386.
Zhang, J., Li, Z., & Wang, Z. (2009). Control chart based on likelihood ratio for monitoring linear profiles. *Computational Statistics and Data Analysis, 53*, 1440–1448.
Zou, C., Qiu, P., & Hawkins, D. (2008). Nonparametric control chart for monitoring profiles using change point formulation and adaptive smoothing. *Statistica Sinica, 19*(3), 1337–1357.
Zou, C., Tsung, F., & Wang, Z. (2007). Monitoring general linear profiles using multivariate exponentially weighted moving average schemes. *Technometrics, 49*(4), 395–408.

Zou, C., Zhang, Y., & Wang, Z. (2006). Control chart based on change-point model for monitoring linear profiles. *IIE Transactions, 38*(12), 1093–1103.
Zou, C., Zhou, C., Wang, Z., & Tsung, F. (2007). A self-starting control chart for linear profiles. *Journal of Quality Technology, 39*(4), 364–375.

## Appendix 1. Appendix for Theorems 1 and 2

In the following, we provide some lemmas to support our main conclusions and the skeleton of their corresponding proofs are list below.

**Lemma 3:** *Assume that* $\eta_i, 1 \le i \le n$ *are independent with* $E(\eta_i) = 0, Var(\eta_i) = \sigma^2, E(\eta_i^{2+\delta}) < \infty, \delta > 0, 1 \le i \le n$ *and the sequence* $\{\mathbf{l}_k = (l_{k1}, \ldots, l_{kn})^T, k = 1, \ldots, p\}$ *satisfy that* $\|\mathbf{l}_k\|^2 = 1, \mathbf{l}_k^T \mathbf{l}_j = 0, k \ne j, \max_j l_{kj} = \mathcal{O}(1/\sqrt{n}), 1 \le k \le p$, *then* $\mathbf{L}\boldsymbol{\eta} \xrightarrow{d} N(\mathbf{0}, \sigma^2 I_p)$ *where* $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T, \mathbf{L} = (l_{ij})_{p\times n} = (\mathbf{l}_1, \ldots, \mathbf{l}_p)^T$.

***Proof*:** Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^T \in R^p$ be a real vector with a unit norm, that is $\|\boldsymbol{\lambda}\| = 1$. It suffices to show that $\boldsymbol{\lambda}^T \mathbf{L}\boldsymbol{\eta} \xrightarrow{d} N(0, \sigma^2)$ by the Cramer–Wold theorem. We define $\boldsymbol{\lambda}^T \mathbf{L}\boldsymbol{\eta} = c_1\eta_1 + \ldots + c_n\eta_n$, then we have $\sum_{j=1}^n c_j^2 = \boldsymbol{\lambda}^T \mathbf{L}\mathbf{L}^T \boldsymbol{\lambda} = 1$ and

$$\begin{aligned}\sum_{j=1}^n |c_j|^{2+\delta} &\le \max_{1\le j\le n} |c_j|^\delta \sum_{j=1}^n c_j^2 = \max_{1\le j\le n} \left|\sum_{i=1}^n \lambda_i l_{ij}\right|^\delta \\ &\le \max_{1\le j\le n} \left(\sum_{i=1}^n \lambda_i^2\right)^{\delta/2} \left(\sum_{i=1}^p l_{ij}^2\right)^{\delta/2} = \mathcal{O}(n^{-\delta/2}).\end{aligned}$$

By the Lyapunov CLT, we obtain $\boldsymbol{\lambda}^T \mathbf{L}\boldsymbol{\eta} \xrightarrow{d} N(0, \sigma^2)$ for an arbitrary unit vector $\lambda$. By the Cramer–Wold theorem, we have $\mathbf{L}\boldsymbol{\eta} \xrightarrow{d} N(\mathbf{0}, \sigma^2 I_p)$. □

**Lemma 4:** $\boldsymbol{\varepsilon}'(0,k)\mathbf{H}_{0,k}\boldsymbol{\varepsilon}(0,k) \xrightarrow{d} \chi^2(p)$ *where* $H_{0,k} = \mathbf{X}_{0,k}\left(\mathbf{X}_{0,k}'\mathbf{X}_{0,k}\right)^{-1}\mathbf{X}_{0,k}'$.

***Proof*:** Firstly, by the spectral decomposition theorem we can write $H_{0,k}$ as $H_{0,k} = \sum_{i=1}^p \mathbf{l}_i\mathbf{l}_i'$ with $p$ orthogonal unit eigenvectors $\{\mathbf{l}_i, i = 1, \ldots, p\}$ and correspondingly $\boldsymbol{\varepsilon}'(0,k)\mathbf{H}_{0,k}\boldsymbol{\varepsilon}(0,k) = \sum_{i=1}^p \left(\boldsymbol{\varepsilon}'(0,k)\mathbf{l}_i\right)^2$. By Proposition 8.2 in Puntanen, Styan, and Isotalo (2011), we can derive $\max_j l_{ij} = \mathcal{O}(1/\sqrt{n})$. Based on Lemma 3, we know that $\boldsymbol{\varepsilon}'(0,k)\mathbf{H}_{0,k}\boldsymbol{\varepsilon}(0,k)$ is asymptotically chi-square distributed with $p$ degrees of freedom. □

**Lemma 5:** *Under assumptions A1,A2 and A3 and the null case of no change-points, for* $1 \le k \le t-1, t \ge 2$ *we have*

$$\begin{aligned}\left(\mathbf{X}_{0,k}^T\mathbf{X}_{0,k}\right)^{1/2}\sigma_0^{-1}\left(\hat{\beta}(0,k) - \beta_0\right) &= k^{-1/2}\sum_{j=1}^k \xi_{j,n_j}^{(1)} + o_p(1), \\ \left(\mathbf{X}_{k,t}^T\mathbf{X}_{k,t}\right)^{1/2}\sigma_0^{-1}\left(\hat{\beta}(k,t) - \beta_0\right) &= (t-k)^{-1/2}\sum_{j=k+1}^t \xi_{j,n_j}^{(2)} + o_p(1), \\ \left(\sum_{j=1}^k n_j\right)^{1/2} v^{-1}\left[\hat{\sigma}^2(0,k) - \sigma_0^2\right] &= k^{-1/2}\sum_{j=1}^k \xi_{j,n_j}^{(2)} + o_p(1), \\ \left(\sum_{j=k+1}^t n_j\right)^{1/2} v^{-1}\left[\hat{\sigma}^2(k,t) - \sigma_0^2\right] &= (t-k)^{-1/2}\sum_{j=k+1}^t \xi_{j,n_j}^{(2)} + o_p(1),\end{aligned}$$

*where* $\xi_{j,n_j}^{(1)} = n_j^{-1/2}\sigma_0^{-1}\mathbf{Q}^{-1/2}\mathbf{X}_j\boldsymbol{\varepsilon}_j, \xi_{j,n_j}^{(2)} = n_j^{-1/2}v^{-1}\sum_{i=1}^{n_j}(\sigma_0^2\varepsilon_{ij}^2 - \sigma_0^2)$.

***Proof*:** Let $\boldsymbol{\varepsilon}(0,k) = \left(\boldsymbol{\varepsilon}_1^T, \boldsymbol{\varepsilon}_2^T, \ldots, \boldsymbol{\varepsilon}_k^T\right)^T$, then by the assumption A1, we obtain

$$\begin{aligned}\left(\mathbf{X}_{0,k}^T\mathbf{X}_{0,k}\right)^{1/2}\sigma_0^{-1}\left(\hat{\beta}(0,k) - \beta_0\right) &= \sigma_0^{-1}\left(\mathbf{X}_{0,k}^T\mathbf{X}_{0,k}\right)^{-1/2}\mathbf{X}_{0,k}^T\boldsymbol{\varepsilon}(0,k) \\ &= k^{-1/2}\sum_{j=1}^k \xi_{j,n_j}^{(1)} + o_p(1), \\ \left(\mathbf{X}_{k,t}^T\mathbf{X}_{k,t}\right)^{1/2}\sigma_0^{-1}\left(\hat{\beta}(k,t) - \beta_0\right) &= (t-k)^{-1/2}\sum_{j=k+1}^t \xi_{j,n_j}^{(2)} + o_p(1).\end{aligned}$$

In the following, we analyze the structure of $\hat{\sigma}^2(0,k), \hat{\sigma}^2(k+1,t)$

$$\begin{aligned}
&\left(\sum_{j=1}^{k} n_j\right)^{1/2} v^{-1}\left[\hat{\sigma}^2(0,k)-\sigma_0^2\right] \\
&= \left(\sum_{j=1}^{k} n_j\right)^{1/2} v^{-1}\left[\sigma_0^2\,\boldsymbol{\varepsilon}'(0,k)\left(\mathbf{I}-\mathbf{H}_{0,k}\right)\boldsymbol{\varepsilon}(0,k)\Big/\sum_{j=1}^{k} n_j-\sigma_0^2\right] \\
&= \left(v^{-1}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\sigma_0^2\varepsilon_{ij}^2-\sigma_0^2\right)-v^{-1}\sigma_0^2\boldsymbol{\varepsilon}'(0,k)\mathbf{H}_{0,k}\boldsymbol{\varepsilon}(0,k)\right)\Big/\sqrt{\sum_{j=1}^{k} n_j} \\
&= k^{-1/2}\sum_{j=1}^{k}\xi_{j,n_j}^{(2)}+o_p(1)
\end{aligned} \tag{A1}$$

where $H_{0,k}=\mathbf{X}_{0,k}\left(\mathbf{X}_{0,k}'\mathbf{X}_{0,k}\right)^{-1}\mathbf{X}_{0,k}'$ represents the so-called 'hat matrix'. The last equality can be induced by Lemma 4, $\boldsymbol{\varepsilon}'(0,k)\mathbf{H}_{0,k}\boldsymbol{\varepsilon}(0,k)=\mathcal{O}_p(1)$. Similarly, $\left(\sum_{j=k+1}^{t} n_j\right)^{1/2}\left[\hat{\sigma}^2(k,t)-\sigma_0^2\right]=(t-k)^{-1/2}\sum_{j=k+1}^{t}\xi_{j,n_j}^{(2)}+o_p(1)$. □

**Lemma 6:** *Let* $\xi_{j,n_j}^{(1)}=n_j^{-1/2}\sigma_0^{-1}\mathbf{Q}^{-1/2}\mathbf{X}_j\boldsymbol{\varepsilon}_j, \xi_{j,n_j}^{(2)}=n_j^{-1/2}v^{-1}\sum_{i=1}^{n_j}(\sigma_0^2\varepsilon_{ij}^2-\sigma_0^2), \xi_{j,n_j}=\begin{pmatrix}\xi_{j,n_j}^{(1)}\\ \xi_{j,n_j}^{(2)}\end{pmatrix}$. *Under assumptions* A1, A2 *and* A3 *and the null case of no change-points, we have established that the series* $\{\xi_{j,n_j}, j=1,2,\ldots\}$ *are independent and*

$$\{\xi_{j,n_j}, j=1,2,\ldots\} \Longrightarrow \{\xi_j, j=1,2,\ldots\}, \tag{A2}$$

*where* $\{\xi_j, j=1,2,\ldots\}$ *are i.i.d.,* $\xi_j \sim N(\mathbf{0}, \mathbf{I}_{p+1})$.

***Proof*:** For an arbitrary $\lambda \in^{R^{p+1}}$, $\|\lambda\|=1$, we consider $\lambda^T\xi_{j,n_j}$. It is easy to verify that the Lindeberg condition works and the conclusion $\lambda^T\xi_{j,n_j}\xrightarrow{d}N(0,1)$ is true by the Lindeberg–Feller Theorem. Based on Wold's Theorem, we arrive at our final conclusion. □

**Lemma 7:** *Under assumptions* A1, A2 *and* A3 *and the null case of no change-points, we have*

$$\hat{\sigma}^2(0,k,t)\xrightarrow{p}\sigma_0^2, \quad \hat{v}^2(0,k,t)\xrightarrow{p}v^2. \tag{A3}$$

***Proof*:** Similar to (A1), $\hat{\sigma}^2(0,k,t)$ can be written equivalently as

$$\begin{aligned}
&\hat{\sigma}^2(0,k,t) \\
&= \left(\sum_{j=1}^{k} n_j\right)\left(\sum_{j=1}^{t} n_j\right)^{-1}\hat{\sigma}^2(0,k)+\left(\sum_{j=k+1}^{t} n_j\right)\left(\sum_{j=1}^{t} n_j\right)^{-1}\hat{\sigma}^2(k,t) \\
&= \left(\sum_{j=1}^{t} n_j\right)^{-1}\sum_{j=1}^{t}\sum_{i=1}^{n_j}\varepsilon_{ij}^2-\left(\sum_{j=1}^{t} n_j\right)^{-1}\left[\boldsymbol{\varepsilon}'(0,k)\mathbf{H}_{0,k}\boldsymbol{\varepsilon}(0,k)+\boldsymbol{\varepsilon}'(k,t)\mathbf{H}_{k,t}\boldsymbol{\varepsilon}(k,t)\right] \\
&\xrightarrow{p}\sigma_0^2
\end{aligned} \tag{A4}$$

where $H_{k,t}=\mathbf{X}_{k,t}\left(\mathbf{X}_{k,t}'\mathbf{X}_{k,t}\right)^{-1}\mathbf{X}_{k,t}'$ and $\boldsymbol{\varepsilon}(k,t)=\left(\boldsymbol{\varepsilon}_{k+1}^T,\ldots,\boldsymbol{\varepsilon}_t^T\right)^T$ correspondingly. By assumptions A1, A2 and A3 and Lemma 6, we obtain

$$\begin{aligned}
&\hat{v}^2(0,k) \\
&= \left\|\mathbf{e}^2(0,k)-\hat{\sigma}^2(0,k)\mathbf{1}_k\right\|^2\Big/\sum_{j=1}^{k} n_j \\
&\frac{1}{\sum_{j=1}^{k} n_j}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\hat{\varepsilon}_{ij}^2-\hat{\sigma}^2(0,k)\right)^2=\frac{1}{\sum_{j=1}^{k} n_j}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\hat{\varepsilon}_{ij}^2-\sigma_0^2\right)^2+\left(\hat{\sigma}^2(0,k)-\sigma_0^2\right)^2
\end{aligned}$$

$$
\begin{aligned}
&\frac{1}{\sum_{j=1}^{k} n_j} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left[ \left( \sigma_0 \varepsilon_{ij} + \mathbf{x}'_{ij} \left( \beta_0 - \hat{\beta}(0,k) \right) \right)^2 - \sigma_0^2 \right]^2 + o_p(1) \\
&\frac{1}{\sum_{j=1}^{k} n_j} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left[ \left( \sigma_0^2 \varepsilon_{ij}^2 - \sigma_0^2 \right) + \left( \mathbf{x}'_{ij} \left( \beta_0 - \hat{\beta}(0,k) \right) \right)^2 + 2\sigma_0 \varepsilon_{ij} \mathbf{x}'_{ij} \left( \beta_0 - \hat{\beta}(0,k) \right) \right]^2 + o_p(1) \\
&\frac{1}{\sum_{j=1}^{k} n_j} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left( \sigma_0^2 \varepsilon_{ij}^2 - \sigma_0^2 \right)^2 + \frac{1}{\sum_{j=1}^{k} n_j} \sum_{j=1}^{k} \sum_{i=1}^{n_j} 2\sigma_0 \varepsilon_{ij} \mathbf{x}'_{ij} \left( \beta_0 - \hat{\beta}(0,k) \right) \left( \sigma_0^2 \varepsilon_{ij}^2 - \sigma_0^2 \right) + o_p(1) \\
&\frac{1}{\sum_{j=1}^{k} n_j} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left( \sigma_0^2 \varepsilon_{ij}^2 - \sigma_0^2 \right)^2 + \left( \beta_0 - \hat{\beta}(0,k) \right)' \frac{2\sigma_0^3}{\sum_{j=1}^{k} n_j} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left( \varepsilon_{ij}^3 - \varepsilon_{ij} \right) \mathbf{x}_{ij} + o_p(1) \\
=&\frac{1}{\sum_{j=1}^{k} n_j} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left( \sigma_0^2 \varepsilon_{ij}^2 - \sigma_0^2 \right)^2 + o_p \left( \left( \sum_{j=1}^{k} n_j \right)^{-1/2} \right) O_p(1) + o_p(1) \xrightarrow{p} v^2.
\end{aligned} \tag{A5}
$$

Similarly, $\hat{v}^2(k,t) \xrightarrow{p} v^2$ and together with (A5), we have

$$
\begin{aligned}
\hat{v}^2(0,k,t) &= \hat{v}^2(0,k) \sum_{j=1}^{k} n_j / \sum_{j=1}^{t} n_j + \hat{v}^2(k,t) \sum_{j=k+1}^{t} n_j / \sum_{j=1}^{t} n_j \\
&\xrightarrow{p} \frac{k}{t} v^2 + \frac{t-k}{t} v^2 = v^2.
\end{aligned}
$$

□

**Lemma 8:** *Under assumptions* A1, A2 *and* A3 *and the null case of no change-points, we have*

$$
(CW_{1,t}, CW_{2,t}, \ldots, CW_{t-1,t})^T \Longrightarrow \left( \frac{B_{t,1}^T B_{t,1}}{\frac{1}{t}\left(1-\frac{1}{t}\right)}, \frac{B_{t,2}^T B_{t,2}}{\frac{2}{t}\left(1-\frac{2}{t}\right)}, \ldots, \frac{B_{t,t-1}^T B_{t,t-1}}{\frac{t-1}{t}\left(1-\frac{t-1}{t}\right)} \right)^T
$$

*where* $B_{t,k} = W_t\left(\frac{k}{t}\right) - \frac{k}{t} W_t(1), k = 1, \ldots, t-1$ *are some discrete values of a standard* $(p+1)-$*dimensional Brownian bridge at finite points in time* $1/t, \ldots, (t-1)/t$ *and* $W_t(k/t)$ *correspond to some discrete values of a standard* $(p+1)-$*dimensional Brownian motion at finite time points* $1/t, \ldots, (t-1)/t$ *such that* $W_t(k/t) = 1/\sqrt{t} \sum_{j=1}^{k} \xi_j$, $\xi_j \sim N(\mathbf{0}, \mathbf{I}_{p+1})$ *does not depend on* $t$.

***Proof*:** Let

$$
\begin{aligned}
\widehat{\mathrm{DC}} &= W_{1k}^{1/2} \hat{\sigma}^{-1}(1,k,t) \left( \hat{\beta}(0,k) - \hat{\beta}(k,t) \right), \\
\widehat{\mathrm{DV}} &= W_{2k}^{1/2} \hat{v}^{-1}(1,k,t) \left( \hat{\sigma}^2(1,k) - \hat{\sigma}^2(k+1,t) \right),
\end{aligned}
$$

then $\mathrm{CW}_{k,t} = \|\widehat{\mathrm{DC}}\|^2 + \|\widehat{\mathrm{DV}}\|^2$. By Lemmas 5, 7 and assumption A1, we can obtain

$$
\begin{aligned}
\widehat{\mathrm{DC}} &= W_{1k}^{1/2} \hat{\sigma}^{-1}(1,k,t) \left( \hat{\beta}(0,k) - \beta_0 \right) - W_{1k}^{1/2} \hat{\sigma}^{-1}(1,k,t) \left( \hat{\beta}(k,t) - \beta_0 \right) \\
&= \left[ \frac{t}{t-k} + o(1) \right]^{-1/2} (\sigma_0 + o_p(1))^{-1} \left( \mathbf{X}_{0,k}^T \mathbf{X}_{0,k} \right)^{1/2} \left( \hat{\beta}(0,k) - \beta_0 \right) \\
&= -\left[ \frac{t}{k} + o(1) \right]^{-1/2} (\sigma_0 + o_p(1))^{-1} \left( \mathbf{X}_{k,t}^T \mathbf{X}_{k,t} \right)^{1/2} \left( \hat{\beta}(k,t) - \beta_0 \right) \\
&= \left[ \frac{t}{t-k} + o(1) \right]^{-1/2} \left( k^{-1/2} \sum_{j=1}^{k} \xi_{j,n_j}^{(1)} + o_p(1) \right) \\
&= -\left[ \frac{t}{k} + o(1) \right]^{-1/2} \left( (t-k)^{-1/2} \sum_{j=k+1}^{t} \xi_{j,n_j}^{(1)} + o_p(1) \right) \\
&= \left[ \frac{k}{t}\left(1-\frac{k}{t}\right) \right]^{-1/2} \left[ \frac{1}{\sqrt{t}} \sum_{j=1}^{k} \xi_{j,n_j}^{(1)} - \frac{k}{t} \frac{1}{\sqrt{t}} \sum_{j=1}^{t} \xi_{j,n_j}^{(1)} \right] + o_p(1)
\end{aligned} \tag{A6}
$$

and

$$\begin{aligned}
\widehat{\mathrm{DV}} &= W_{2k}^{1/2}\hat{\nu}^{-1}(1,k,t)\left(\hat{\sigma}^2(1,k)-\sigma_0^2\right) - W_{2k}^{1/2}\hat{\nu}^{-1}(1,k,t)\left(\hat{\sigma}^2(k+1,t)-\sigma_0^2\right)\\
&= \left[\frac{t}{t-k}+o(1)\right]^{-1/2}(\nu+o_p(1))^{-1}\left(\sum_{i=1}^{k}n_i\right)^{1/2}\left(\hat{\sigma}^2(1,k)-\sigma_0^2\right)\\
&\quad -\left[\frac{t}{k}+o(1)\right]^{-1/2}(\nu+o_p(1))^{-1}\left(\sum_{i=k+1}^{t}n_i\right)^{1/2}\left(\hat{\sigma}^2(k+1,t)-\sigma_0^2\right)\\
&= \left[\frac{t}{t-k}+o(1)\right]^{-1/2}\left(k^{-1/2}\sum_{j=1}^{k}\xi_{j,n_j}^{(2)}+o_p(1)\right)\\
&\quad -\left[\frac{t}{k}+o(1)\right]^{-1/2}\left((t-k)^{-1/2}\sum_{j=k+1}^{t}\xi_{j,n_j}^{(2)}+o_p(1)\right)\\
&= \left[\frac{k}{t}\left(1-\frac{k}{t}\right)\right]^{-1/2}\left[\frac{1}{\sqrt{t}}\sum_{j=1}^{k}\xi_{j,n_j}^{(2)}-\frac{k}{t}\frac{1}{\sqrt{t}}\sum_{j=1}^{t}\xi_{j,n_j}^{(2)}\right]+o_p(1).
\end{aligned} \tag{A7}$$

Combining (A6) and (A7) together, by Lemma 6 and Slutsky's theorem we get

$$\begin{aligned}
\begin{pmatrix}\widehat{\mathrm{DC}}\\ \widehat{\mathrm{DV}}\end{pmatrix} &= \left[\frac{k}{t}\left(1-\frac{k}{t}\right)\right]^{-1/2}\left[\frac{1}{\sqrt{t}}\sum_{j=1}^{k}\begin{pmatrix}\xi_{j,n_j}^{(1)}\\ \xi_{j,n_j}^{(2)}\end{pmatrix}-\frac{k}{t}\frac{1}{\sqrt{t}}\sum_{j=1}^{t}\begin{pmatrix}\xi_{j,n_j}^{(1)}\\ \xi_{j,n_j}^{(2)}\end{pmatrix}\right]+o_p(1)\\
&= \left[\frac{k}{t}\left(1-\frac{k}{t}\right)\right]^{-1/2}\left[\frac{1}{\sqrt{t}}\sum_{j=1}^{k}\xi_{j,n_j}-\frac{k}{t}\frac{1}{\sqrt{t}}\sum_{j=1}^{t}\xi_{j,n_j}\right]+o_p(1)\\
\Longrightarrow &\left[\frac{k}{t}\left(1-\frac{k}{t}\right)\right]^{-1/2}B_{t,k}, k=1,2,\ldots,t-1
\end{aligned}$$

where

$$B_{t,k} = W_t\left(\tfrac{k}{t}\right)-\tfrac{k}{t}W_t(1), k=1,\ldots,t-1$$

are some discrete values of a standard $(p+1)$-dimensional Brownian bridge at finite points in time $1/t,\ldots,(t-1)/t$ and $W_t(k/t)$ correspond to discrete values of a standard $(p+1)$-dimensional Brownian motion in finite times points $1/t,\ldots,(t-1)/t$ such that $W_t(k/t) = 1/\sqrt{t}\sum_{j=1}^{k}\xi_j$, $\xi_j \sim N(\mathbf{0},\mathbf{I}_{p+1})$ does not depend on $t$. By the continuous mapping theorem, we can obtain the final conclusion. □

**Proof of Theorem 1–2:** By the Continuous Mapping Theorem and Lemma 8, we easily obtain the conclusions. □