Applied Soft Computing Journal 89 (2020) 106101

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Hybrid change point detection for time series via support vector regression and CUSUM method

Sangyeol Lee *, Sangjo Lee, Miteum Moon

Department of Statistics, Seoul National University, Seoul, 08826, Korea

ARTICLE INFO

Article history:
Received 26 April 2019
Received in revised form 6 January 2020
Accepted 14 January 2020
Available online 27 January 2020

Keywords:
SVR–ARMA model
ARMA forecasting
Change point test
Hybrid method
LSCUSUM test

ABSTRACT

This study considers the change point testing problem regarding time series based on the location and scale-based cumulative sum (LSCUSUM) test constructed with the residuals obtained from support vector regression (SVR)-autoregressive moving average (ARMA) models. For this, we first estimate the model parameters in SVR–ARMA models from a training time series sample, in which a long AR model is fitted to the data to obtain residuals. We then use these as initial values of the error terms in SVR–ARMA (p, q) models and obtain the forecasting values recursively until the updated error terms converge to a certain limit. Finally, we select an optimal order of p, q with the root mean square error (RMSE) and use the forecasting errors from this selected model as the residuals for constructing the LSCUSUM test. Monte Carlo simulations are performed to evaluate the validity of the test. A real data example is provided for illustration.

1. Introduction

In this study, we consider the change point testing problem for time series based on the location and scale-based cumulative sum (LSCUSUM) test constructed with the support vector regression (SVR)-based autoregressive moving average (ARMA) models. Since [1], the problem of testing for a parameter change has been an important issue in economics, engineering, and medicine, and a multitude of articles have been published in various research areas [2]. Because time series often suffer from structural changes owing to changes in policy and critical social events, the change point test has been viewed as a core issue in this context for several decades. The CUSUM test has been quite popular among many change point tests due to its ease of use and abundant articles exist regarding the CUSUM test for time series. For earlier works, we refer to [3–6] and the papers cited therein, and further, for recent developments, see [7–15] who consider several different types of CUSUM tests and compare their performance.

The conventional estimate-based CUSUM test is designed to compare the discrepancy among sequentially obtained estimators [5]. This estimates-based test generally performs well but suffers from severe size distortions and produces low powers on some occasions, particularly when the underlying model is complicated and has many unknown parameters. Therefore, the residual-based CUSUM test for time series models has been advocated as a remedy [6–8]. However, the residual-based CUSUM test for location-scale models undergoes a severe power loss in location parameter changes. To overcome this drawback, [9] and [10] suggested using the score vector-based CUSUM test for ARMA-generalized autoregressive conditional heteroscedastic (GARCH) models. [11] also proposed a modified residual-based CUSUM test to lessen an effort to deal with the derivatives of the log-likelihood functions in constructing the test of [10] and enhanced the power performance. [12] further improved the test of [11] by introducing a lot handier location and scale-based CUSUM (LSCUSUM) test, demonstrating its validity for ARMA-type models. Because the LSCUSUM test is constructed only with observations and residuals, it has advantages over other CUSUM tests in terms of hybrid capability with other methods that can afford to calculate residuals. Motivated by this, we consider using the SVR–ARMA model in implementing the LSCUSUM test.

In the construction of the LSCUSUM test, an important step is to estimate the residuals accurately. That is, a correct time series prediction is crucial because the residuals are merely the prediction errors. Time series prediction is generally important to forecast the behavior of time series and detect malfunctions or anomalies in statistical process control. In the literature, the most popular time series forecast method is using the classical ARMA model. Conventional linear ARMA models yield an accurate prediction when a time series truly follows them. However, if the time series has significant nonlinear characteristics, the prediction result based on the ARMA models is incorrect and hard to harness for further applications. In this case, practitioners can employ nonparametric prediction methods such as a recurrent neural network (RNN) and support vector regression [16,17]. The

* Corresponding author.
E-mail addresses: sylee@stats.snu.ac.kr (S. Lee), lscho13@snu.ac.kr (S. Lee), rkskf@snu.ac.kr (M. Moon).

https://doi.org/10.1016/j.asoc.2020.106101
1568-4946/

RNN method is well known to outperform the ARMA model in many situations, particularly when time series have, to certain extent, nonlinear and non-stationary features. However, it has some limitations such as the need for a large number of tuning parameters, difficulty in finding a unique global solution owing to a different choice of initial weights, and over-fitting [18]. In contrast, the SVR has flexibility, outstanding forecasting accuracy, and a balance between the training and generalization errors, resulting in better empirical performance than the RNN as well as ARMA models [19,20]. It is well known that the SVR minimizes the structural risk and meets the Structural Risk Minimization Principle, while the RNN minimizes the empirical risk, namely, the error regarding the in-sample estimating data [21]. Motivated by this, we also adopt the SVR method for time series prediction, and based on the obtained ARMA residuals, we construct the LSCUSUM test to test for change points. See [22] for a reference concerning the SVR–ARMA method.

The rest of this paper is organized as follows. Section 2 introduces the LSCUSUM test in the classical ARMA model and outlines its basic principle. Section 3 proposes a forecasting method based on the SVR–ARMA model and describes how to determine an optimal SVR–ARMA model. The residuals are obtained through an optimal SVR–ARMA model to a given training time series sample, which is split into two subseries. A long AR model is fitted to the first subseries to obtain the initial values of residuals, which are used as the error terms in the SVR–ARMA(p, q) model and are recursively updated until the obtained residuals converge to a certain limit. This procedure is applied to each p, q less than a predetermined K. Then, for each estimated SVR–ARMA(p, q) model, we calculate the root mean square errors (RMSEs) based on the second subseries and select an optimal ARMA order with the smallest RMSE. The determined SVR–ARMA(p, q) model is then applied to obtain the prediction errors or residuals, which are finally used to construct the LSCUSUM test. Section 4 performs Monte Carlo simulations to evaluate the LSCUSUM test's validity for various time series models. Section 4 provides a real data example for illustration. Section 5 provides concluding remarks.

2. LSCUSUM test for ARMA models

To develop a CUSUM test in time series models, [11] considered the CUSUM test for the location-scale model of the form $y_t = g_t(\mu) + \sqrt{h_t(\theta)}\eta_t$, where $g_t(\mu)$ and $h_t(\theta)$ are the conditional mean and variance with parameters μ and $\theta = (\mu^T, \lambda^T)^T$, and η_t are iid error terms with mean zero and unit variance. The location-scale model includes a broad class of autoregressive conditional heteroscedastic (ARCH) time series models, covering ARMA-generalized ARCH (GARCH) models. To implement a change point test, they constructed the CUSUM test using the conditional mean and the residuals. [12] recently designed a simpler CUSUM test using only the observations y_t and residuals ϵ_t in ARMA models.

Let us consider the stationary ARMA(p, q) model:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-1} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}, \tag{1}$$

where ϕ_i and θ_j $i = 1, \dots, p$ and $j = 1, \dots, q$ are real numbers, and ϵ_t are iid mean zero random variables with covariance $\sigma^2 > 0$ and a finite fourth moment. Here, our objective is to test whether either conditional mean of y_t over past information $\sum_{i=1}^{p} \phi_i y_{t-1} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$ or the variance of the error terms ϵ_t experience a change point over time t. Since a change occurs if the parameter vector $\vartheta = (\phi_1, \dots \phi_p, \theta_1, \dots \theta_q, \sigma^2)^T$ changes, given observations $y_1, \dots, y_n$, we set up the null and alternative hypotheses:

$$\mathrm{H}_0 : \vartheta \text{ remains constant for } t = 1, \dots, n \quad \text{vs.} \quad \mathrm{H}_1 : \text{ not } \mathrm{H}_0.$$

Under the alternative hypothesis, we actually consider the situation in which $\vartheta = \vartheta_0$ for $t = 1, \dots, [n\tau]$ for some $\tau \in (0, 1)$ and $\vartheta = \vartheta_1$ with $\vartheta_1 \neq \vartheta_0$ for $t = [n\tau] + 1, \dots, n$.

To test the above, we recursively obtain the residuals $\hat{\epsilon}_t = y_t - \sum_{i=1}^{p} \hat{\phi}_i Y_{t-i} + \epsilon_t + \sum_{j=1}^{q} \hat{\theta}_j \epsilon_{t-j}$ with $y_t = \epsilon_t = 0$ for all $t \leq 0$, where $\hat{\phi}_i$ and $\hat{\theta}_j$ are consistent estimators. Then, to test for a parameter change in ϑ, [12] constructed the CUSUM tests:

$$\hat{T}_n^{LS} = \max_{1 \leq k \leq n} \left\{ \frac{1}{n\hat{\tau}_{1,n}^2} \left| \sum_{t=1}^{k} (y_t - \hat{\epsilon}_t)\hat{\epsilon}_t - \frac{k}{n} \sum_{t=1}^{n} (y_t - \hat{\epsilon}_t)\hat{\epsilon}_t \right|^2 + \frac{1}{n\hat{\tau}_{2,n}^2} \left| \sum_{t=1}^{k} \hat{\epsilon}_t^2 - \frac{k}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2 \right|^2 \right\},$$

$$\hat{T}_n^{max} = \max_{1 \leq k \leq n} \max \left\{ \frac{1}{\sqrt{n}\hat{\tau}_{1,n}} \left| \sum_{t=1}^{k} (y_t - \hat{\epsilon}_t)\hat{\epsilon}_t - \frac{k}{n} \sum_{t=1}^{n} (y_t - \hat{\epsilon}_t)\hat{\epsilon}_t \right|, \frac{1}{\sqrt{n}\hat{\tau}_{2,n}} \left| \sum_{t=1}^{k} \hat{\epsilon}_t^2 - \frac{k}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2 \right| \right\},$$

where

$$\hat{\tau}_{1,n}^2 = \frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{\epsilon}_t)^2 \hat{\epsilon}_t^2 \quad \text{and} \quad \hat{\tau}_{2,n}^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^4 - \left(\frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2 \right)^2.$$

Then, we reject the null hypothesis of no changes if $\hat{T}_n^{LS} > 2.089$ or $\hat{T}_n^{max} > 1.453$ at the nominal level of 0.05. The critical values of these test are obtained though Monte Carlo simulations using the two-dimensional standard Brownian motion. It is well known that $\hat{T}_n^{LS}$ outperforms $\hat{T}_n^{max}$ when both the ARMA coefficients and the variance of the error terms include changes, whereas the reverse is true when only one of these two include changes [12].

In fact, to check whether the change exists in ARMA coefficients or in the error variance, we can also employ the following tests:

$$\hat{T}_n^{ARMA} = \max_{1 \leq k \leq n} \frac{1}{n\hat{\tau}_{1,n}^2} \left| \sum_{t=1}^{k} (y_t - \hat{\epsilon}_t)\hat{\epsilon}_t - \frac{k}{n} \sum_{t=1}^{n} (Y_t - \hat{\epsilon}_t)\hat{\epsilon}_t \right|^2,$$

$$\hat{T}_n^{VAR} = \max_{1 \leq k \leq n} \frac{1}{n\hat{\tau}_{2,n}^2} \left| \sum_{t=1}^{k} \hat{\epsilon}_t^2 - \frac{k}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2 \right|^2.$$

Because these tests asymptotically behave as the supremum of a standard Brownian bridge [12], we reject the null hypothesis of no changes at the nominal level of 0.05 if they have values larger than 1.358.

Although the ARMA models can cover a broad class of stationary time series, they cannot accommodate all time series with strong nonlinear features. Thus, we use the SVR–ARMA model. That is, we train a training sample using the SVR–ARMA model with specific kernels and finally determine an optimal SVR–ARMA(p_0, q_0) model through Steps 1–4 in Section 4. Then, for any testing sample, we construct the LSCUSUM test based on the trained SVR–ARMA(p_0, q_0) model. The details are presented in the next section.

3. SVR model

Support vector regression (SVR) is a functional tool to approximate various types of functions and make accurate predictions for time series. SVR aims to identify a nonlinear function f that approximates the output y_t within a forecasting error based on given data $\{(\mathrm{x}_t, y_t)\}_{t=1}^{n}$, where $\mathrm{x}_t \in \mathbb{R}^k$ is a k-dimensional input vector and $y_t \in \mathbb{R}$ is a scalar output. More precisely, f has the following form:

$$f(\mathrm{x}_t) = w^T \phi(\mathrm{x}_t) + b,$$

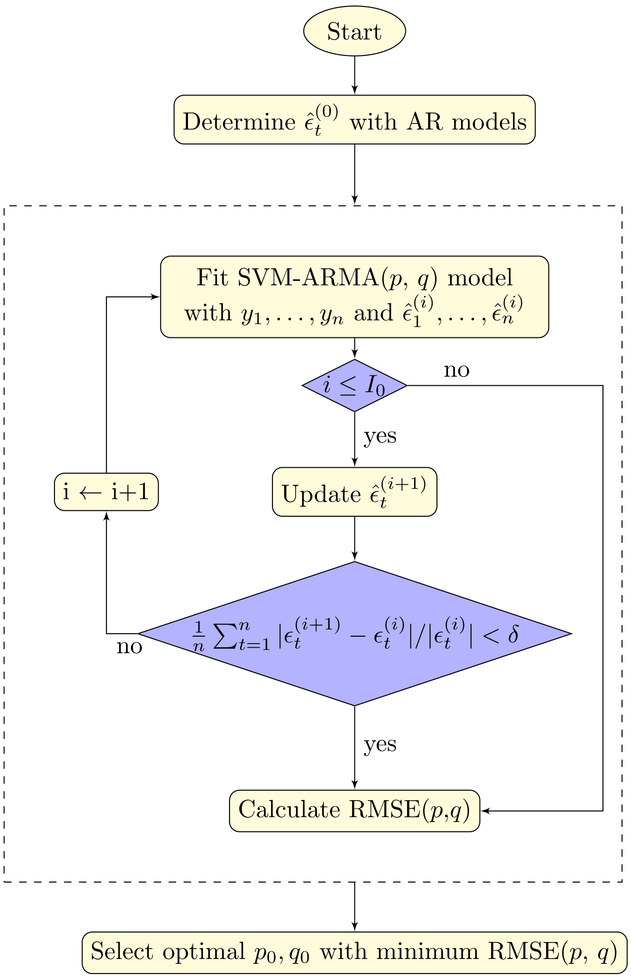

Fig. 1. Steps for obtaining an optimal SVR–ARMA model.

where w and b are regression parameter vectors and $\phi(\cdot)$ is a known nonlinear function. The optimal w^* and b^*, which best predict f, are found by minimizing the following regularized function $R(C)$ [23]:

$$R(C) = \frac{1}{2}\|w\|^2 + C\sum_{t=1}^{n} L_\epsilon(y_t, f(\mathrm{x}_t)), \tag{2}$$

where linear ϵ-insensitive loss function, L_ϵ, is given by:

$$L_\epsilon(y, f(\mathrm{x})) = \begin{cases} |y - f(\mathrm{x})| - \epsilon & \text{if } |y - f(\mathrm{x})| > \epsilon, \\ 0 & \text{otherwise.} \end{cases}$$

This ϵ-insensitive loss function indicates that the observations inside ϵ-tube do not carry any information in the construction of f, and C denotes a trade-off between the function complexity and training error [24].

To indicate the errors outside ϵ-insensitive zone, we also introduce slack variables ξ_t and ξ_t' as in [25], and then, equation (2) can be rewritten as:

$$\text{Minimize: } \frac{1}{2}\|w\|^2 + C\sum_{t=1}^{n}(\xi_t + \xi_t'), \tag{3}$$

$$\text{subject to: } \begin{cases} y_t - w^T\phi(\mathrm{x}_t) - b \leq \epsilon + \xi_t, \\ w^T\phi(\mathrm{x}_t) + b - y_t \leq \epsilon + \xi_t', \\ \xi_t, \xi_t' \geq 0. \end{cases}$$

To solve equation (3), we use the Lagrangian multipliers (α_t and α_t') and the Karush–Kuhn–Tucker conditions to transform the problem in its dual form as in [19] as follows:

$$\begin{aligned} \text{Minimize: } & \frac{1}{2}\sum_{s=1}^{n}\sum_{t=1}^{n}(\alpha_s - \alpha_s')(\alpha_t - \alpha_t')\langle\phi(\mathrm{x}_s), \phi(\mathrm{x}_t)\rangle \\ & + \epsilon\sum_{t=1}^{n}(\alpha_t + \alpha_t') - \sum_{t=1}^{n} y_t(\alpha_t' - \alpha_t) \\ \text{subject to } & \begin{cases} \sum_{t=1}^{n}(\alpha_t - \alpha_t') = 0 \\ 0 \leq \alpha_t, \alpha_t' \leq C, \quad t = 1, \ldots, n. \end{cases} \end{aligned} \tag{4}$$

Then, the optimal solution (w^*, b^*) is determined as in [25]:

$$w^* = \sum_{t=1}^{n}(\alpha_t' - \alpha_t)\phi(x_t),$$

$$b^* = \begin{cases} y_s - \sum_{t=1}^{n}(\alpha_t' - \alpha_t)\langle\phi(\mathrm{x}_t), \phi(\mathrm{x}_s)\rangle + \epsilon, & s \in \{t|\alpha_t \in (0, C)\} \\ y_s - \sum_{t=1}^{n}(\alpha_t' - \alpha_t)\langle\phi(\mathrm{x}_t), \phi(\mathrm{x}_s)\rangle - \epsilon, & s \in \{t|\alpha_t' \in (0, C)\}. \end{cases} \tag{5}$$

Thus, the nonlinear function f can be written as in [23]:

$$\begin{aligned} f(\mathrm{x}) &= w^*\phi(\mathrm{x}) + b^* \\ &= \sum_{t=1}^{n}(\alpha_t' - \alpha_t)K(\mathrm{x}_t, \mathrm{x}) + b^*, \end{aligned} \tag{6}$$

where $K(\mathrm{x}_t, \mathrm{x}) = \langle\phi(\mathrm{x}_t), \phi(\mathrm{x})\rangle$ is the inner product kernel function. In this study, we use the Gaussian kernel:

$$K(\mathrm{x}_t, \mathrm{x}) = \exp\left(-\gamma\|\mathrm{x} - \mathrm{x}_t\|^2\right).$$

In implementing the SVR, the coefficients ϵ, C, γ should be determined appropriately. Here, we use $C = 1, \epsilon = 0.1, \gamma = 1$ because this choice gives a satisfactory result in our simulation study. In general, a cross-validation method can be used for the choice of these parameters: for a relevant reference regarding this issue, refer to [26]. Although not reported here, this method (in Step 3 below) yields a result similar to what we obtain in the simulation study.

4. Prediction based on SVR-ARMA model

Suppose that a training sample $y_1, \ldots, y_n, y_{n+1}, \ldots, y_{n+m}$ is given. Here, $y_1, \ldots, y_n$ and $y_{n+1}, \ldots, y_{n+m}$ are also used as validation samples. We assume that the sample is generated from the ARMA model:

$$y_t = f(y_{t-1}, \ldots, y_{t-p}, \epsilon_{t-1}, \ldots, \epsilon_{t-q}) + \epsilon_t, \tag{7}$$

where f is an unknown function to be estimated, p, q are nonnegative integers that should be properly determined, and ϵ_t are iid random variables with zero mean and a finite variance. If the training sample is known to follow an SVR–ARMA(p, q) model with specific orders p and q, we can use this model. However, if it is not the case, we determine the orders from the training sample through the following procedure (see Fig. 1):

Step 1. We fit a long AR(p) model to $y_1, \ldots, y_n$, where $p = p_n$ is a sequence of positive integers with $p_n \to \infty$, $p_n/n \to 0$ when n tends to infinity, for example, $p_n = (\log n)^2$. Then, using the least squares method, we calculate the residuals from this AR(p_n) model, that is, $\hat{\epsilon}_t^{(0)}, t = 1, \ldots, n$ ($\hat{\epsilon}_t^{(0)} = 0, t \leq 0$).

Table 1
Sizes and powers for AR(1) model with $\phi = 0.3, \sigma^2 = 1$ under the null and each alternatives.

$\phi = 0.3, \sigma^2 = 1$			SVR–ARMA	Linear ARMA
Size		$\hat{T}_n^{LS}$	0.058	0.048
		$\hat{T}_n^{max}$	0.052	0.038
Change of ϕ	$\to \phi = 0.5$	$\hat{T}_n^{LS}$	0.474	0.374
		$\hat{T}_n^{max}$	0.520	0.464
	$\to \phi = 0.7$	$\hat{T}_n^{LS}$	0.994	0.982
		$\hat{T}_n^{max}$	0.998	0.994
Change of σ^2	$\to \sigma^2 = 2.0$	$\hat{T}_n^{LS}$	0.992	0.998
		$\hat{T}_n^{max}$	0.996	0.996

Table 2
Sizes and powers for ARMA(1, 1) model with $\phi = 0.3, \theta = 0.3, \sigma^2 = 1$ under the null and each alternatives.

$\phi = 0.3, \theta = 0.3, \sigma^2 = 1$			SVR–ARMA	Linear ARMA
Size		$\hat{T}_n^{LS}$	0.080	0.042
		$\hat{T}_n^{max}$	0.074	0.044
Change of ϕ	$\to \phi = 0.5$	$\hat{T}_n^{LS}$	0.504	0.374
		$\hat{T}_n^{max}$	0.540	0.462
	$\to \phi = 0.7$	$\hat{T}_n^{LS}$	0.992	0.980
		$\hat{T}_n^{max}$	0.996	0.998
Change of θ	$\to \theta = 0.7$	$\hat{T}_n^{LS}$	0.714	0.696
		$\hat{T}_n^{max}$	0.664	0.828
Change of σ^2	$\to \sigma^2 = 2.0$	$\hat{T}_n^{LS}$	0.982	0.980
		$\hat{T}_n^{max}$	0.996	0.996

Table 3
Sizes and powers for threshold ARMA(1, 1) model with $\phi_1 = 0.1, \phi_2 = -0.5, \theta_1 = \theta_2 = 0.5, \sigma^2 = 1$ under the null and each alternatives.

$\phi_1 = 0.1, \phi_2 = -0.5, \theta = 0.5, \sigma^2 = 1$			SVR–ARMA	Linear ARMA
Size		$\hat{T}_n^{LS}$	0.076	0.098
		$\hat{T}_n^{max}$	0.056	0.092
Change of ϕ_1	$\to \phi_1 = 0.5$	$\hat{T}_n^{LS}$	0.836	0.794
		$\hat{T}_n^{max}$	0.886	0.838
	$\to \phi_1 = 0.7$	$\hat{T}_n^{LS}$	0.996	0.994
		$\hat{T}_n^{max}$	0.998	0.998
Change of θ	$\to \theta = -0.5$	$\hat{T}_n^{LS}$	1.000	0.804
		$\hat{T}_n^{max}$	0.998	0.852
Change of σ^2	$\to \sigma^2 = 2.0$	$\hat{T}_n^{LS}$	0.982	0.984
		$\hat{T}_n^{max}$	0.996	0.996

Step 2. Fix $K > 0$. For any $p, q \leq K$, we fit an SVR–ARMA(p, q) model to $y_1, \ldots, y_n$ using the residuals $\hat{\epsilon}_t^{(0)}$ obtained from Step 1. That is, we assume that $\{y_t\}$ approximately satisfies model (7) as follows:

$$y_t \simeq f(y_{t-1}, \ldots, y_{t-p}, \hat{\epsilon}_{t-1}^{(0)}, \ldots, \hat{\epsilon}_{t-q}^{(0)}) + \epsilon_t. \tag{8}$$

Then, using y_t and $x_t = (y_{t-1}, \ldots, y_{t-p}, \hat{\epsilon}_{t-1}^{(0)}, \ldots, \hat{\epsilon}_{t-q}^{(0)})$, we apply the SVR prediction method used in Section 3 to obtain an updated residual $\hat{\epsilon}_t^{(1)}$. Recursively, we obtain $\hat{\epsilon}_t^{(i+1)}$ from model

$$y_t \simeq f(y_{t-1}, \ldots, y_{t-p}, \hat{\epsilon}_{t-1}^{(i)}, \ldots, \hat{\epsilon}_{t-q}^{(i)}) + \epsilon_t. \tag{9}$$

We repeat this procedure until we find $i_0 := i_0(p, q) = \min\left\{i; \frac{1}{n}\sum_{t=1}^{n} |\epsilon_t^{(i+1)} - \epsilon_t^{(i)}|/|\epsilon_t^{(i)}| < \delta\right\} \wedge I_0$ with a preassigned positive integer I_0, 1000, and $\delta > 0$, 0.001, for example.

Step 3. For each $p, q \leq K$, we assume that the $y_1, \ldots, y_n$ approximately satisfies

$$y_t \simeq f(y_{t-1}, \ldots, y_{t-p}, \epsilon_{t-1}^{(i_0)}, \ldots, \epsilon_{t-q}^{(i_0)}) + \epsilon_t, \tag{10}$$

where i_0 is the one obtained in Step 2, and the SVR–ARMA(p, q) model is fitted with y_t and $x_t = (y_{t-1}, \ldots, y_{t-p}, \epsilon_{t-1}^{(i_0)}, \ldots, \epsilon_{t-q}^{(i_0)})$ as in Section 3. Then, applying the obtained SVR–ARMA(p, q) model to $y_{n+1}, \ldots, y_{n+m}$, we calculate the prediction errors $\tilde{\epsilon}_t$, $t = 1, \ldots, m$, and the corresponding RMSE=: RMSE$(p, q) = \left\{\sum_{t=1}^{m} \tilde{\epsilon}_t^2/m\right\}^{1/2}$. We then select the optimal order (p_0, q_0) as the (p, q) that minimizes the RMSEs.

Step 4. We train the SVR–ARMA(p_0, q_0) model based on the whole training sample $y_1, \ldots, y_{n+m}$. Then, for any testing samples, we use the estimated SVR–ARMA(p_0, q_0) model in calculating the prediction errors $\hat{\epsilon}_t$. The obtained $\hat{\epsilon}_t$ are then used in the construction of the LSCUSUM tests $\hat{T}_n^{LS}$ and $\hat{T}_n^{max}$ in Section 2, called the "SVR-based LSCUSUM test".

Table 4
Sizes and powers for threshold ARMA(1, 1) model with $\phi_1 = 0.3, \phi_2 = -0.5, \theta_1 = \theta_2 = 0.5, \sigma^2 = 1$ under the null and each alternatives.

$\phi_1 = 0.3, \phi_2 = -0.5, \theta = 0.5, \sigma^2 = 1$			SVR–ARMA	Linear ARMA
Size		$\hat{T}_n^{LS}$	0.080	0.098
		$\hat{T}_n^{max}$	0.052	0.092
Change of ϕ_1	$\to \phi_1 = 0.5$	$\hat{T}_n^{LS}$	0.414	0.438
		$\hat{T}_n^{max}$	0.408	0.492
	$\to \phi_1 = 0.7$	$\hat{T}_n^{LS}$	0.970	0.946
		$\hat{T}_n^{max}$	0.962	0.964
Change of θ	$\to \theta = -0.5$	$\hat{T}_n^{LS}$	1.000	0.244
		$\hat{T}_n^{max}$	1.000	0.280
Change of σ^2	$\to \sigma^2 = 2.0$	$\hat{T}_n^{LS}$	0.980	0.984
		$\hat{T}_n^{max}$	0.998	0.994

Remark 1. In Step 3, if one wants to select p_0 and q_0 more accurately, we change the role of the training and evaluation datasets, that is, $y_{n+1}, \ldots, y_{n+m}$ is the training sample and $y_1, \ldots, y_n$ is the evaluation sample. Then, if the optimal order (p_0^*, q_0^*) obtained from this set up coincides with the (p_0, q_0), we select (p_0, q_0) as an optimal order; otherwise, between (p_0^*, q_0^*) and (p_0, q_0), we choose the one that yields a smaller RMSE value, calculated from the full sample $y_1, \ldots, y_{n+m}$.

5. Simulation study

In this section, we evaluate the performance of the SVR-based LSCUSUM tests $\hat{T}_n^{LS}$ and $\hat{T}_n^{max}$ for ARMA, threshold ARMA, and time-varying AR models. Each simulation is conducted with a sample size of 500 at the nominal level of 0.050. The sizes and powers are calculated as the rejection number of the null of no changes out of 500 repetitions. Under alternatives, the change is assumed to occur in the middle of the testing sample. The SVR-based LSCUSUM tests are compared with the ARMA-based LSCUSUM tests. For the SVM, we use the R package e1071.

(1) linear ARMA model

To observe the performance of the SVR-based LSCUSUM tests, we generate time series from the ARMA (1, 1) model:

$$y_t = \phi y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t$$

where ϵ_t are from $N(0, \sigma^2)$. Under the null hypothesis, we consider the two cases:

(i) $\phi = 0.3, \theta = 0.0$, and $\sigma^2 = 1.0$
(ii) $\phi = 0.3, \theta = 0.3$, and $\sigma^2 = 1.0$

Under the alternative, we consider the cases that only one of ϕ, θ, and σ^2 changes and the others remain the same. Table 1

Table 5
Sizes and powers for PAR(1) model with $\phi_1 = 0.3, \phi_2 = 0.5, T = 2$ under the null and each alternatives.

$\phi_1 = 0.3, \phi_2 = 0.5, T = 2$			SVR–ARMA	Linear ARMA
Size		$\hat{T}_n^{LS}$	0.058	0.044
		$\hat{T}_n^{max}$	0.052	0.048
Change of ϕ_1	$\to \phi_1 = 0.5$	$\hat{T}_n^{LS}$	0.194	0.144
		$\hat{T}_n^{max}$	0.192	0.130
	$\to \phi_1 = 0.7$	$\hat{T}_n^{LS}$	0.562	0.410
		$\hat{T}_n^{max}$	0.616	0.496
Change of ϕ_2	$\to \phi_2 = 0.7$	$\hat{T}_n^{LS}$	0.152	0.100
		$\hat{T}_n^{max}$	0.138	0.082
	$\to \phi_2 = -0.3$	$\hat{T}_n^{LS}$	0.886	0.884
		$\hat{T}_n^{max}$	0.920	0.936

Table 6
Sizes and powers for PAR(1) model with $\phi_1 = -0.3, \phi_2 = -0.5, T = 2$ under the null and each alternatives.

$\phi_1 = -0.3, \phi_2 = -0.5, T = 2$			SVR–ARMA	Linear ARMA
Size		$\hat{T}_n^{LS}$	0.054	0.046
		$\hat{T}_n^{max}$	0.050	0.042
Change of ϕ_1	$\to \phi_1 = -0.5$	$\hat{T}_n^{LS}$	0.194	0.138
		$\hat{T}_n^{max}$	0.198	0.158
	$\to \phi_1 = -0.7$	$\hat{T}_n^{LS}$	0.564	0.384
		$\hat{T}_n^{max}$	0.624	0.456
Change of ϕ_2	$\to \phi_2 = -0.7$	$\hat{T}_n^{LS}$	0.144	0.072
		$\hat{T}_n^{max}$	0.146	0.076
	$\to \phi_2 = 0.3$	$\hat{T}_n^{LS}$	0.916	0.870
		$\hat{T}_n^{max}$	0.932	0.928

Table 7
Non-linearity tests for training and test sets of Nikkei225.

Test			Tsay's F-test	Keenan test	Ljung–Box test
Daily	Training	Statistics	2.021	4.791	98.548
		p-value	0.028	0.029	< 0.001
	Testing	Statistics	5.349	4.632	34.468
		p-value	0.021	0.032	< 0.001
Weekly	Training	Statistics	6.045	7.023	0.773
		p-value	0.015	0.009	0.379
	Testing	Statistics	0.878	1.282	0.005
		p-value	0.454	0.259	0.941

shows that in the AR (1) model, both the SVR- and ARMA-based LSCUSUM tests have no size distortions and produce reasonably good powers. However, Table 2 shows that in the ARMA (1, 1) model, the SVR-based LSCUSUM tests have some size distortions. In both cases, the SVR-based $\hat{T}_n^{\max}$ appears to outperform the SVR-based $\hat{T}_n^{LS}$ in terms of stability and power, and the former is therefore preferred to the latter.

(2) Threshold AR model

We generate time series from the model:

$$y_t = \phi_1 y_{t-1}^{+} + \phi_2 y_{t-1}^{-} + \theta \epsilon_{t-1} + \epsilon_t.$$

Under the null hypothesis, we consider the two cases as in Tables 3 and 4:

(i) $\phi_1 = 0.1, \phi_2 = -0.5, \theta = 0.5$, and $\sigma^2 = 1.0$
(ii) $\phi_1 = 0.3, \phi_2 = -0.5, \theta = 0.5$, and $\sigma^2 = 1.0$

We consider the cases that only one of ϕ_1, ϕ_2, θ, and σ^2 changes and the others remain the same. Tables 3 and 4 show that the ARMA-based LSCUSUM tests have severe size distortions.

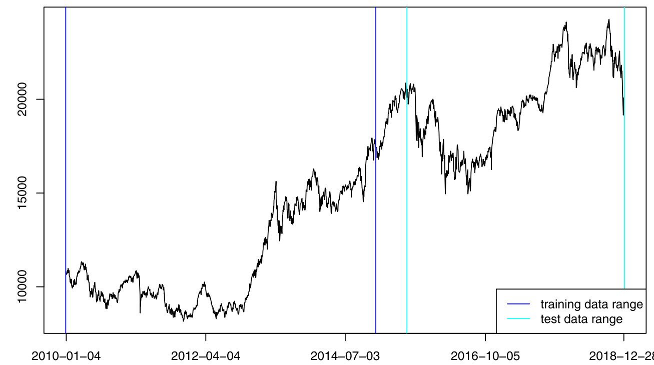

Fig. 2. Daily price of Nikkei 225.

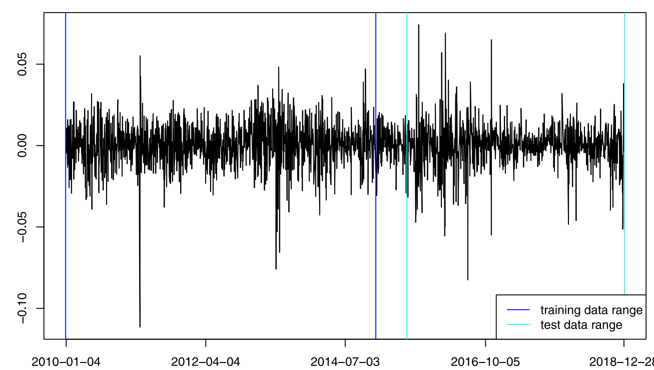

Fig. 3. Daily log-return of Nikkei 225.

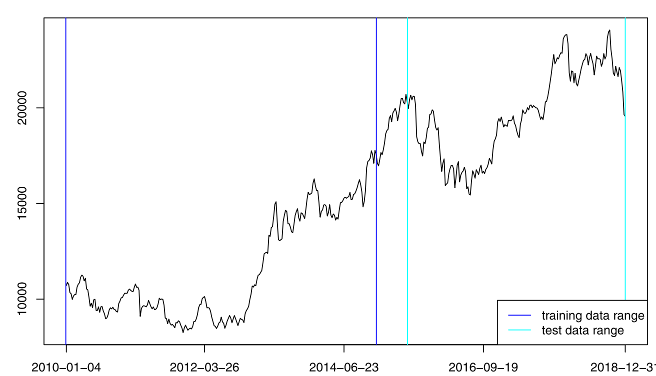

Fig. 4. Weekly price of Nikkei 225.

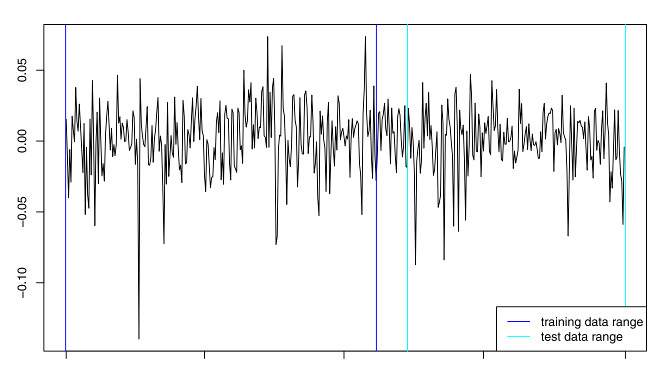

Fig. 5. Weekly log-return of Nikkei 225.

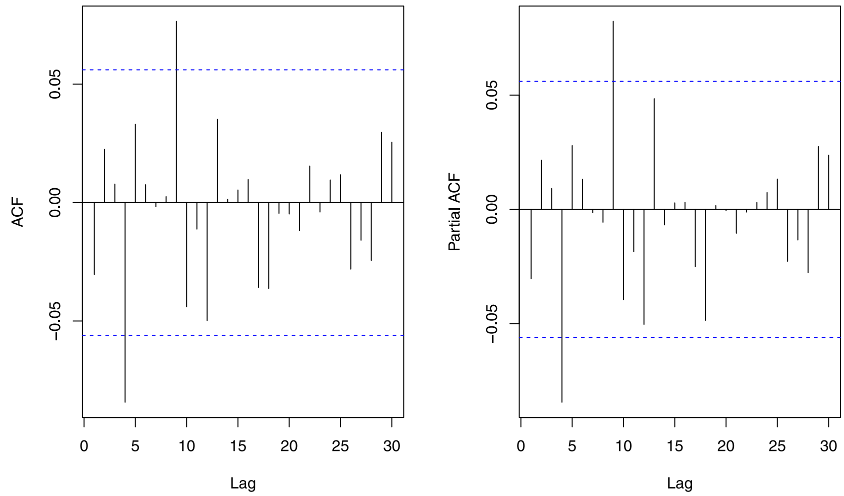

Fig. 6. ACF and PACF of Daily Nikkei 225.

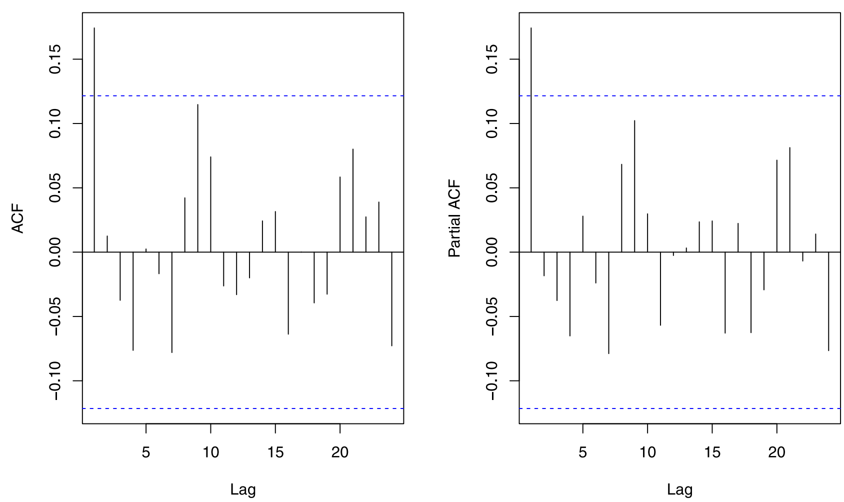

Fig. 7. ACF and PACF of Weekly Nikkei 225.

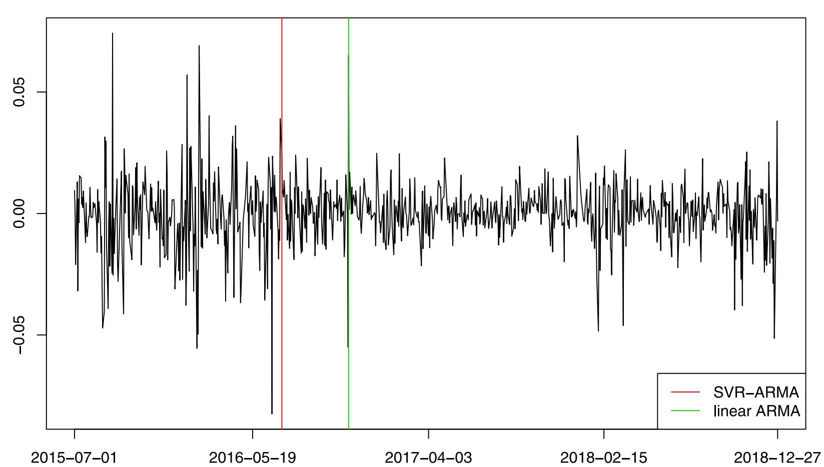

Fig. 8. Change points for Nikkei 225 daily log-returns.

In particular, the SVR-based $\hat{T}_n^{max}$ mostly outperforms the other tests.

(3) Time-varying AR model

We generate time series from the model:

$$y_t = \phi_t y_{t-1} + \epsilon_t$$

where ϕ_t is periodic with period $t = 1, 2$ [27]. Here, we consider the following cases:

(i) $\phi_1 = 0.3$, $\phi_2 = 0.5$, and $\sigma^2 = 1.0$
(ii) $\phi_1 = -0.3$, $\phi_2 = -0.5$, and $\sigma^2 = 1.0$

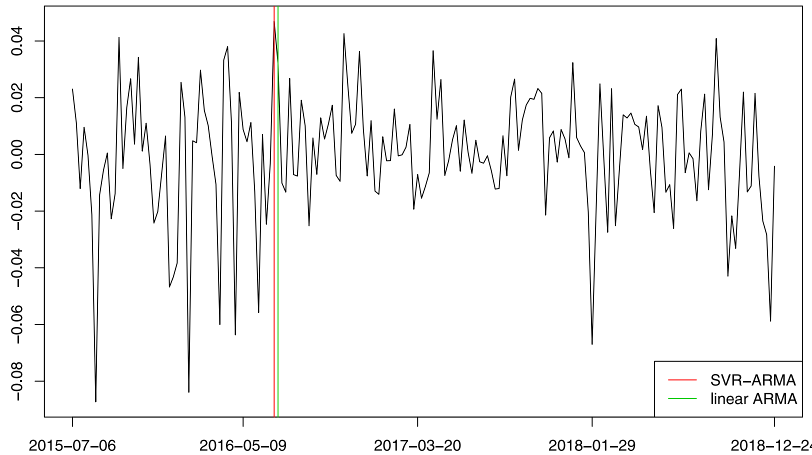

Fig. 9. Change points of Nikkei 225 weekly log-returns.

Tables 5 and 6 show that all the tests have no size distortions, and the SVR-based LSCUSUM test outperforms the ARMA-based LSCUSUM tests. As in Case (2), the SVR-based $\hat{T}_n^{max}$ appears to outperform the other tests. Upon the suggestion of a referee, we applied the fitted SVR model obtained from ARMA(1, 1) time series in **(1)** to 1000 newly generated ARMA(1, 1) time series datasets and examined the size and power behavior of the LSCUSUM test, and subsequently, had outcomes with a similar pattern to Tables 1 and 2 (not reported here). Similar results were also obtained for models **(2)** and **(3)**.

Moreover, we conducted the same experiments using the SVR model with other kernels such as polynomial and sigmoid kernels. However, the results for these kernels were not satisfactory. In fact, it is well known that the Gaussian kernel function is effective in capturing the local characteristic of the sample and the learning capability is quite strong compared to polynomial and sigmoid kernels, so is widely used for forecast. Overall, our findings demonstrate the validity of the SVR-based LSCUSUM tests and the superiority of the SVR-based $\hat{T}_n^{max}$ to other tests.

Remark 2. One referee suggested considering a different SVR method, such as the twin SVR (TSVR) method [28,29], in the construction of the CUSUM test. In its implementation, a reduced number of observations (10% of the whole observations) are used to deal with the high dimensional matrix inverse problem as discussed in (39)-(42) of [30]. As reported in [30], the TSVR is shown to reduce the CPU time markedly and have a good forecasting performance. Its CPU time from Steps 1–4 is measured as about 15 s, which is at least 16 times faster than our method with CPU time of 242 s on the average, when implemented in R on Windows 10 running on a PC with Intel i7-3770 processor (3.4 GHz) with 8 GB of RAM. Further, although not reported here in details, it is seen that the TSVR-based CUSUM test compares well with our test in terms of stability, that is, no severe size distortions are observed. However, it is also revealed that the TSVR method produces less powers than our method (in particular, the gap between the two methods is somewhat large in Case (3) time-varying AR model). This is not strange, though, because the high forecasting accuracy does not necessarily guarantee a high performance of the CUSUM test in general, possibly due to over-fitting. Our current work is only a first attempt to make a hybrid between the CUSUM and SVR methods, and cannot cover all different SVR methods for comparison. We believe that more efforts must be invested to refine our method via employing more advanced methodologies, which will be pursued in our future project.

6. Real data analysis

In this section, we apply the SVR-based LSCUSUM method to the Nikkei daily 225 data. We analyze 100*log-returns of daily Nikkei225 prices from January 4 2010 to December 31 2018. We split the dataset into the training dataset from January 4 2010 to December 30 2014 and the testing dataset from July 1 2015 to December 31 2018. Figs. 2 and 3 plot daily and weekly datasets and Figs. 4 and 5 plot daily and weekly log-return datasets.

Figs. 6 and 7 show that both autocorrelation function (ACF) and partial ACF (PACF) plots have irregular peaks and falls, and Table 7 shows that both Tsay's and Keenan's linearity tests [31] are quite against the linearity assumption, which strongly suggests the inadequacy of the ARMA-fitting to our datasets. Moreover, the augmented Dickey–Fuller's test and Phillips–Perron's test [32] indicate the existence of no unit roots. The Ljung–Box test [31] applied to the squares of residuals reveals that the daily data has an ARCH effect, but the weekly data does not.

We apply the SVR-based $\hat{T}_n^{max}$ to the testing datasets. Subsequently, we obtain $\hat{T}_n^{max} = 2.9163$ for the daily dataset and $\hat{T}_n^{max} = 1.4661$ for the weekly dataset. Since they are larger than 1.453, we reject the null hypothesis of no changes at the level of 0.05. Moreover, because the $\hat{T}_n^{max}$ is maximized at $t = 252$ for the daily dataset and $t = 53$ for the weekly dataset, we conclude that the change points lie at July 11 2016 for the daily dataset and July 4 2016 for the weekly dataset. See Figs. 8 and 9, wherein the red vertical lines denote the change points. This result makes sense in that the two estimated change points are quite close to each other.

Next, we apply the (linear) ARMA-based LSCUSUM test to the testing datasets. We here expect this method to misbehave due to the nonlinearity of the datasets. Using the training datasets, we obtain the ARMA (5, 5) model as a best fit for the daily dataset, and the ARMA (5, 4) model for the weekly dataset. Based on these models, we detect a change point at November 9 2016 for the daily dataset and July 11 2016 for the weekly dataset. Figs. 8 and 9, wherein the green vertical lines denote the change points, show that the change point for the daily dataset is quite away from the previously obtained one, July 11 2016, while that for the weekly dataset is close to the previous one, July 4 2016. This result demonstrates that the ARMA-based LSCUSUM can be severely undermined when the given time series sample has a strong nonlinear characteristic.

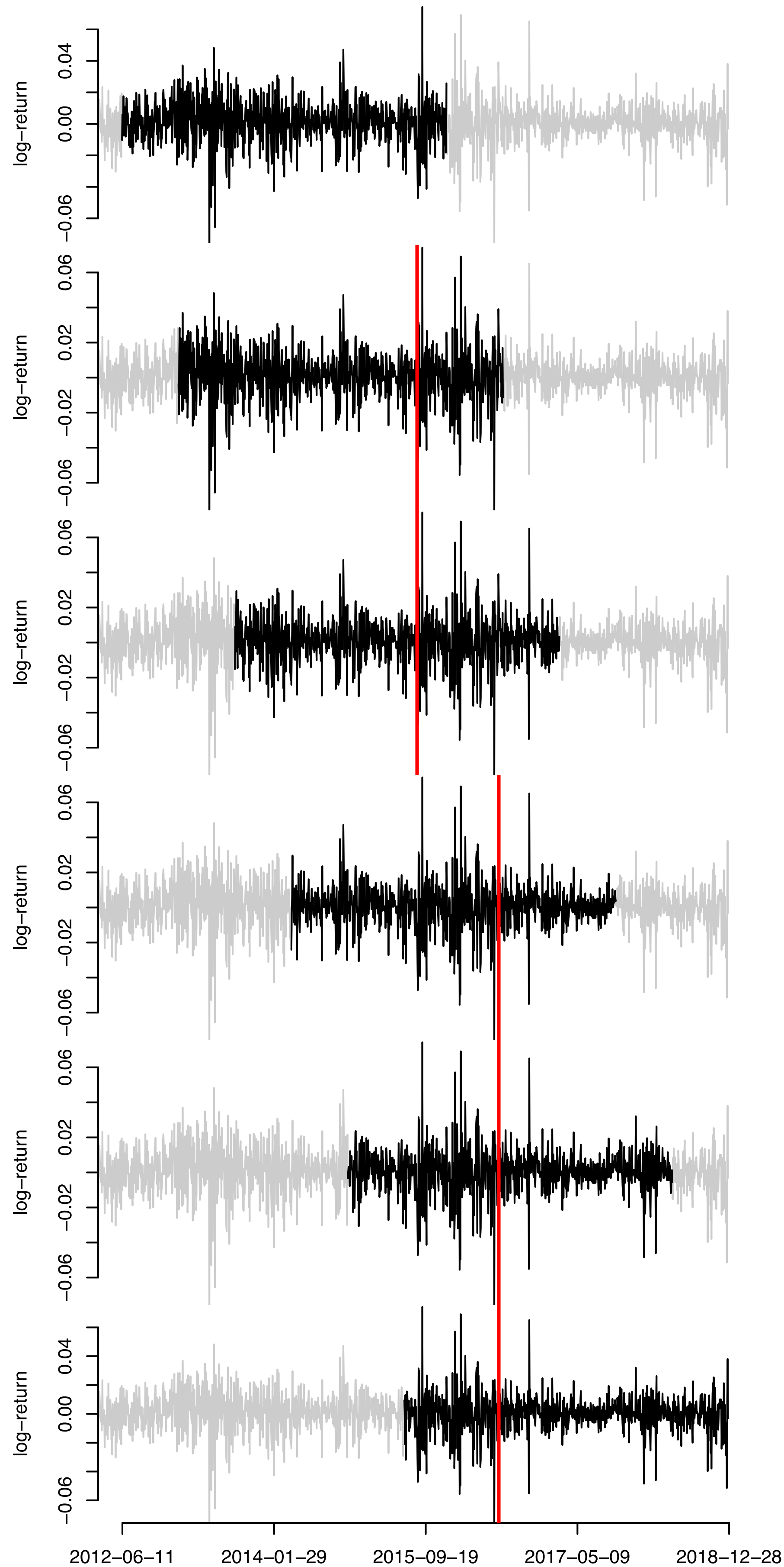

Fig. 10. Change points of moving Nikkei 225 daily log-returns.

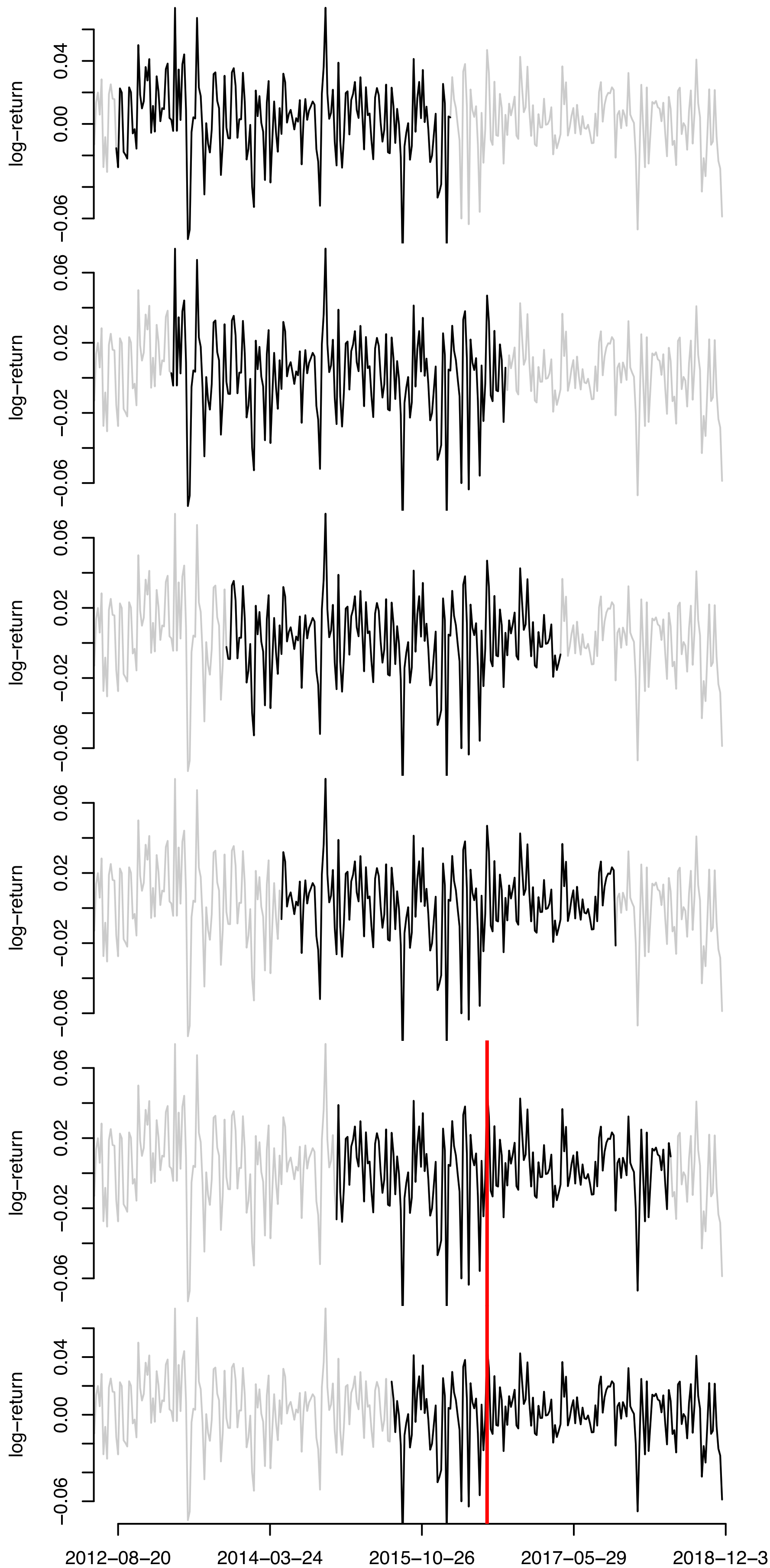

Fig. 11. Change points of moving Nikkei 225 weekly log-returns.

We also conduct the LSCUSUM test applying the SVR model fitted from the training data to six time series of size 862 (daily data) starting from August 20 2011 to July 1 2015, denoted by $S_{d,i}$, $i = 1, \ldots, 6$, and another six time series of size 183 (weekly data) starting from June 11 2012 to July 6 2015, denoted by $S_{w,i}$, $i = 1, \ldots, 6$, with their initial dates increased by 150 days and 30 weeks, respectively. Figs. 10 and 11 illustrate the plots of $S_{d,i}$ and $S_{w,i}$ (in dark black) with change points denoted by red vertical lines, showing that the LSCUSUM test does not detect a change point for the first few time series sets because a larger portion of the training data is included in those time series, but detected a change point for the rest of time series. Note that the LSCUSUM test detects a change point for $S_{d,i}$, $i = 2, 3$, almost one year earlier than the original change point July 11, 2016, which is not unreasonable, though, because high volatilities can affect change point tests to a great degree, whereas it only detects a change for $S_{w,i}$, $i = 4, 5, 6$, at the original change point July 4, 2016. The result suggests that the LSCUSUM test for ARMA time series can easily misidentify the change point in the presence of high volatilities, which is generally well known even for model-based CUSUM tests. In general, the location of change point depends upon the period of time series, and thus, the selection of the training time series would be an important issue in practical usage. In the selection process, a preliminary test could be conducted using parametric ARMA or GARCH models, which, however, is not completely reliable due to a possibility of model misspecification. As the daily time series can have significant volatilities, the LSCUSUM test should be redesigned based on GARCH-SVR models to capture the GARCH effect [11]. We leave this issue for our future project as it is beyond the scope of this study.

7. Concluding remarks

In this study, we consider the SVR-based LSCUSUM test to detect a change point for time series. Our simulation study confirms the validity of our method and shows that the SVR-LSCUSUM test outperforms the ARMA-based LSCUSUM test when the underlying model is nonlinear. For illustration, a data analysis was conducted using a Nikkei225 dataset, which also supports the practicality of the SVR-based LSCUSUM test. We plan to extend our work to time series with high volatility in our future project.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2020.106101.

CRediT authorship contribution statement

Sangyeol Lee: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing - original draft. **Sangjo Lee:** Data curation, Formal analysis, Methodology, Software, Validation, Visualization. **Miteum Moon:** Methodology, Software, Validation.

Acknowledgments

We would like to thank the Editor, an AE, and three anonymous referees for their valuable comments. This research is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (No. 2018R1A2A2A05019433).

References

[1] E.S. Page, A test for a change in a parameter occurring at an unknown point, Biometrika 42 (1955) 523–527.
[2] M. Csörgő, L. Horváth, Limit Theorems in Change-Point Analysis, John Wiley & Sons Inc., New York, 1997.
[3] C. Inclán, G. Tiao, Use of cumulative sums of squares for retrospective detection of changes of variance, J. Amer. Statist. Assoc. 89 (1994) 913–923.
[4] S. Kim, S. Cho, S. Lee, On the CUSUM test for parameter changes in GARCH (1, 1) models, Comm. Statist. Theory Methods 29 (2000) 445–462.
[5] S. Lee, O. Na, S. Na, On the CUSUM of squares test for variance change in nonstationary and nonparametric time series models, Ann. Inst. Statist. Math. 55 (2003) 467–485.
[6] S. Lee, Y. Tokutsu, K. Maekawa, The CUSUM test for parameter change in regression models with ARCH errors, J. Japan Statist. Soc. 34 (2004) 173–188.
[7] O. Na, J. Lee, S. Lee, Change point detection in SCOMDY models, AStA Adv. Stat. Anal. 97 (2013) 215–238.
[8] J. Lee, S. Lee, Parameter change test for nonlinear time series models with GARCH type errors, J. Korean Math. Soc. 52 (2015) 503–522.
[9] E. Gombay, Change detection in autoregressive time series, J. Multivariate Anal. 99 (2008) 451–464.
[10] H. Oh, S. Lee, On score vector- and residual-based CUSUM tests in ARMA-GARCH models, Stat. Methods Appl. 27 (2018) 385–406.
[11] H. Oh, S. Lee, Modified residual CUSUM test for location-scale time series models with heteroscedasticity, Ann. Inst. Statist. Math. 71 (2019) 1059–1091.
[12] S. Lee, Location and scale-based CUSUM test with application to autoregressive models, 2019, submitted for publication.
[13] C.W.S. Chen, S. Lee, Generalized autoregressive Poisson models for time series of counts, Comput. Statist. Data Anal. 99 (2016) 51–67.
[14] M.C. Dong, C.W.S. Chen, S. Lee, S. Sriboonchitta, How strong is the relationship among gold and USD exchange rates? Analytics based on structural change models, Comput. Econ. 53 (2016) 343–366.
[15] H. Oh, S. Lee, Parameter change test for location-scale time series models with heteroscedasticity based on bootstrap, Appl. Stoch. Models Bus. (2019) Online published.
[16] G.S. Atsalakis, K.P. Valavanis, Surveying stock market forecasting techniques Part II: Soft computing methods, Expert Syst. Appl. 36 (2009) 5932–5941.
[17] K.R. Müller, A.J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, V. Vapnik, Predicting time series with support vector machines, in: International Conference on Artificial Neural Networks, Springer, Berlin, Heidelberg, 1997, pp. 999–1004.
[18] F.E. Tay, L. Cao, Application of support vector machines in financial time series forecasting, Omega 29 (2001) 309–317.
[19] N.I. Sapankevych, R. Sankar, Time series prediction using support vector machines: a survey, IEEE Comput. Intell. Mag. 4 (2009) 24–38.
[20] R.C. Cavalcante, R.C. Brasileiro, V.L. Souza, J.P. Nobrega, A.L. Oliveira, Computational intelligence and financial markets: A survey and future directions, Expert Syst. Appl. 55 (2016) 194–211.
[21] L.J. Cao, F.E.H. Tay, Support vector machine with adaptive parameters in financial time series forecasting, IEEE Trans. Neural Netw. 14 (2003) 1506–1518.
[22] S. Chen, K. Jeong, W.K. Härdle, Recurrent support vector regression for a non-linear ARMA model with applications to forecasting financial returns, Comput. Statist. 30 (2015) 821–843.
[23] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 2000.
[24] G. Sermpinis, C. Stasinakis, K. Theofilatos, A. Karathanasopoulos, Inflation and unemployment forecasting with genetic support vector regression, J. Forecast. 33 (2014) 471–487.
[25] B. Schölkopf, A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond, MIT press, New York, 2001.
[26] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Netw. 17 (2004) 113–126.
[27] E. Broszkiewicz-Suwaj, A. Makagon, R. Weron, A. Wyłomańska, On detecting and modeling periodic correlation in financial data, Physica A 336 (2004) 196–205.
[28] L. Tang, Y. Tian, C. Yang, Nonparallel support vector regression model and its SMO-type solver, Neural Netw. 105 (2018) 431–446.
[29] M. Carrasco, J. López, S. Maldonado, Epsilon-nonparallel support vector regression, Appl. Intell. 49 (2019) 4223–4236.
[30] X. Peng, TSVR: an efficient twin support vector machine for regression, Neural Netw. 23 (2010) 365–372.
[31] W.K. Li, Diagnostic Checks in Time Series, Chapman & Hall/CRC, New York, 2004.
[32] G. Kirchgässner, J. Wolters, U. Hassler, Introduction to Modern Time Series Analysis, second ed., Springer, Berlin, 2013.