




Article

Multiscale Change Point Detection for Univariate Time Series Data with Missing Value

Tariku Tesfaye Haile ^{1,2,*} , Fenglin Tian ¹, Ghada AlNemer ³  and Boping Tian ^{1,*} ¹ School of Mathematics, Harbin Institute of Technology, Harbin 150001, China; tfl_hit@126.com² Department of Statistics, College of Natural and Computational Sciences, Arba Minch University, Arba Minch P.O. Box 21, Ethiopia³ Department of Mathematical Sciences, College of Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; gnnemer@pnu.edu.sa

* Correspondence: tariku.tesfaye@amu.edu.et (T.T.H.); bopingt361147@hit.edu.cn (B.T.)

Abstract: This paper studies the autoregressive integrated moving average (ARIMA) state space model combined with Kalman smoothing to impute missing values in a univariate time series before detecting change points. We estimate a scale-dependent time-average variance constant that depends on the length of the data section and is robust to mean shifts under serial dependence. The consistency of the proposed estimator is shown under the assumption allowing heavy tailedness. Integrating the proposed estimator with the moving sum and wild binary segmentation procedures to determine the number and locations of change points is discussed. Furthermore, the performance of the proposed methods is evaluated through extensive simulation studies and applied to the Beijing multi-site air quality dataset to impute missing values and detect mean changes in the data.

Keywords: ARIMA; Kalman smoothing; time-average variance constant; robust estimation; moving sum; wild binary segmentation

MSC: 62F35; 62M10



Citation: Haile, T.T.; Tian, F.; AlNemer, G.; Tian, B. Multiscale Change Point Detection for Univariate Time Series Data with Missing Value. *Mathematics* **2024**, *12*, 3189. <https://doi.org/10.3390/math12203189>

Academic Editor: Davide Valenti

Received: 21 August 2024

Revised: 3 October 2024

Accepted: 10 October 2024

Published: 11 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Change point detection (CPD) has been studied for several decades and remains an active area of research. It has been employed in various fields such as financial analysis, medical condition monitoring, climatology, and air quality. Recent developments in machine learning have also contributed to the advancement of CPD methods, with applications in bioinformatics, speech recognition, image analysis, and monitoring of complex systems have also been motivated. For an extensive review of change point detection, refer to Aminikhanghahi and Cook [1] and Truong et al. [2]. The methods used for detecting and localization of multiple mean shifts in univariate data that search for changes over the data section at various scales are commonly referred to as the canonical change point detection [3]. In such methods, it is typically necessary to estimate the level of noise to distinguish the actual change from the random noise-induced fluctuations, especially when there is a serial correlation.

A brief review of multiple CPDs in scenarios with serial correlation is provided. One research direction aims to adopt techniques originally developed for independent observations to be applicable in the analysis of time series data. This includes the use of test statistics like the cumulative sum (CUSUM) [4] and moving sum (MOSUM) [5] tests. The information criteria of Yao [6] was also adopted for dependent data [3,7] which was initially designed for independent Gaussian random variables. Using specific time series models, such as the autoregressive (AR) model, and concurrently estimating dependence and change points are the topics of another area of study. In line with this, refs. [8–11] utilized the AR model, which is adopted to handle the serial correlations to detect multiple

change points in the mean of AR processes. Furthermore, another research line suggests using long-run variance (LRV) estimators to gauge noise levels in CPD that are robust to multiple mean shifts [12–15]. Estimation of the LRV has been a long-time difficult problem, even without the presence of mean change [16]. The widely used LRV kernel estimator tends to have a downward bias [17,18]. Moreover, sometimes it does not guarantee that the corresponding estimate is always positive unless some underlying assumptions are imposed on the estimators of the LRV [13]. Even some estimators of the LRV can take negative values when the LRV approaches zero [19], which renders their application unsuitable for statistical procedures. This task even becomes more difficult when there are multiple change points and the data exhibit serial dependence.

Missing values are ubiquitous in real-world datasets. They introduce uncertainty during data analysis, which can affect the statistical estimator's properties and reduce efficiency, leading to incorrect conclusions [20,21]. Usually, the deletion of missing observations results in a valuable loss of important information, which biases statistical estimation and inference. The concept of change point analysis with missing values has also been addressed in the literature [22–24]. Another more application-oriented work includes Yang et al. [25], among others, who considered the use of threshold-based regression models, which examine the change based on covariate relationships. The other is Zhao et al. [26] who applied an online missing value imputation algorithm for mixed-type (continuous and categorical) data and detected the change points. The aforementioned literature considers high dimensional data or non-time series data, which depends on the inter-attribute correlation between variables or attributes. Furthermore, most existing univariate change point detection methods, designed for complete datasets, often delete missing values [27,28]. Such an approach may result in the incorrect identification of genuine change point locations, especially when these points coincide with the time points where missing data occur. Given the necessity of replacing missing values with reasonable estimates before applying CPD methods, it is crucial to develop an effective imputation technique to ensure the accurate identification of change points. However, there is a significant gap in addressing change point detection for univariate time series in the presence of missing values.

Therefore, in this paper, we propose the ARIMA state space model combined with Kalman smoothing to impute missing values in univariate time series. This approach leverages the ARIMA model and Kalman smoothing capability to handle time series data with seasonality and trends, making the imputation more accurate and reliable. Moreover, to tackle the challenges associated with using LRV in change point detection, we apply a scale-dependent time-average variance constant (TAVC, Mcgonigle and Cho [27]) estimator, which is resilient to varying mean and serial correlation. For a stationary process $\{\varepsilon_t\}_{t=1}^T$, the scale-dependent TAVC defined as

$$\sigma_D^2 = \text{Var} \left(D^{-\frac{1}{2}} \sum_{t=1}^D \varepsilon_t \right), \quad (1)$$

which depends on scale D , and is different from the earlier LRV estimators. Combining a scale-dependent TAVC estimator with multiscale CPD methods, such as the MOSUM procedure and the wild binary segmentation (WBS) algorithm, effectively addresses the problem of detecting multiple shifts in the series. The performance of our imputation method and robust TAVC estimator is rigorously evaluated through extensive simulations and subsequently applied to the Beijing air quality dataset. Thus, this integrated approach is designed to overcome the challenges posed by missing values and improve the precision of change point detection in real-world time series data. The contributions of this paper are summarized as follows. Firstly, we apply the ARIMA state-space model combined with Kalman smoothing to impute missing values in univariate time series data exhibiting mean shifts. Secondly, we derive the TAVC estimator, and integrate it into multiscale change point detection methods. This approach estimates and locates change points for the imputed time series data, demonstrating robust performance in handling missing values.

The remainder of this paper is structured as follows: Section 2 discusses the imputation of missing values using the ARIMA state space model and Kalman smoothing. Section 3 introduces the concept of TAVC, its robust estimator, and demonstrates its consistency. In Section 4, the application of the TAVC estimator in conjunction with multiscale CPD methods such as MOSUM and WBS has been given in detail. Section 5 evaluates the performance of the proposed approaches and apply to Beijing air quality. Finally, Section 6 summarizes the main findings of the paper.

2. Missing Value Imputation

In this section, we briefly discuss the general framework of the ARIMA state space model and the implementation of Kalman smoothing to estimate missing values. The state-space model offers a versatile method for analyzing time series data due to the recursive behavior of the model, which is particularly useful for computational simplicity of maximum-likelihood estimation and handling of missing values [29–31]. The Kalman filtering method, as part of a state-space model, dynamically adjusts to new data and handles non-stationary series by recursively estimating missing values as new observations become available [32–34]. Moreover, the ARIMA state space and Kalman smoothing method perform exceptionally well for time series data that exhibit strong seasonality with high percentages of missing and large gap sizes under the MAR and MCAR mechanisms [35–37]. The state space model is adaptable for various types of time series data; however, in this paper, we consider the linear time series which follows Gaussian distribution (see, [30,33]).

2.1. State Space Model and Kalman Filter

The state space model comprises a measurement equation that establishes a relationship between the observed data (\mathbf{y}_t) and an m -dimensional state vector α , and a transition equation that describes how the state vector changes over time. For a linear Gaussian state space model, the measurement (observation) equation takes the form

$$\mathbf{y}_t = \mathbf{Z}_t \alpha_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (2)$$

where \mathbf{y}_t is $N \times 1$ vector of observed value, \mathbf{Z}_t is $N \times m$ matrix, and ε_t is $N \times 1$ vector of disturbances, which is normally independently distributed (NID) with mean zero and $N \times N$ covariance matrix \mathbf{H}_t . The elements of the state vector α_t are unobservable and its transition equation is the first-order Markov process

$$\alpha_t = \mathbf{T}_t \alpha_{t-1} + \mathbf{R}_t \eta_t, \quad t = 1, \dots, T, \quad (3)$$

where \mathbf{T}_t is $m \times m$ transition matrix, \mathbf{R}_t is $m \times r$ matrix, and η_t is $r \times 1$ vector of disturbances which is $NID(\mathbf{0}, \mathbf{Q}_t)$.

Assumption 1.

- i. The initial state vector $\alpha_0 \sim N(\mathbf{a}_0, \mathbf{P}_0)$.
- ii. $\mathbb{E}(\varepsilon_t \eta_s') = \mathbf{0}$, $\mathbb{E}(\varepsilon_t \alpha_0') = \mathbf{0}$ and $\mathbb{E}(\eta_t \alpha_0') = \mathbf{0}$ for all $s, t = 1, \dots, T$.

The matrices \mathbf{Z}_t , \mathbf{H}_t , \mathbf{T}_t , \mathbf{R}_t and \mathbf{Q}_t are called the system matrices and contain non-random elements.

Once the model is translated to a state space model, we apply the Kalman filter to estimate state vector parameters. The Kalman filter is a recursive algorithm utilized for the computation of the optimal estimator of the state vector at time t , taking into account all the available information, including the observations y_t [30,32,33]. For the state space model (2) and (3) assume that the initial state α_0 is $N(\mathbf{a}_0, \mathbf{P}_0)$, where \mathbf{a}_0 and \mathbf{P}_0 are known. Let \mathbf{Y}_{t-1} denote the set of past observations $\mathbf{y}_1, \dots, \mathbf{y}_{t-1}$ and \mathbf{a}_{t-1} be the optimal estimators of α_{t-1} based on \mathbf{Y}_{t-1} with the corresponding mean square error (MSE) matrix, \mathbf{P}_{t-1} . Then, the Kalman filter consists of two sets of equations, namely prediction equations and updating equations. Given \mathbf{a}_{t-1} and \mathbf{P}_{t-1} at time $t - 1$, the optimal estimator of α_t and its corresponding MSE matrix are

$$\mathbf{a}_{t|t-1} = \mathbb{E}(\alpha_t | \mathbf{Y}_{t-1}) = \mathbf{T}_t \mathbf{a}_{t-1}, \tag{4}$$

$$\mathbf{P}_{t|t-1} = \mathbb{E}[(\alpha_t - \mathbf{a}_{t|t-1})(\alpha_t - \mathbf{a}_{t|t-1})' | \mathbf{Y}_{t-1}] = \mathbf{T}_t \mathbf{P}_{t-1} \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'. \tag{5}$$

Equations (4) and (5) are known as prediction equations. Additionally, the optimal estimator of \mathbf{y}_t given the information at $t - 1$ is $\hat{\mathbf{y}}_{t|t-1} = \mathbb{E}(\mathbf{y}_t | \mathbf{Y}_{t-1}) = \mathbf{Z}_t \mathbf{a}_{t|t-1}$. The prediction error is computed as $\mathbf{v}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1} = \mathbf{Z}_t(\alpha_t - \mathbf{a}_{t|t-1}) + \varepsilon_t$. Usually \mathbf{v}_t is known as a one-step ahead prediction error and its MSE matrix becomes $Var(\mathbf{v}_t) = \mathbb{E}(\mathbf{v}_t \mathbf{v}_t') = \mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t$. Therefore, the above components are required to derive the prediction error decomposition of the log-likelihood function.

When additional new observations \mathbf{y}_t become available, the knowledge about the optimal estimator of α_t , $\mathbf{a}_{t|t-1}$ can be updated. The updating equations for α_t given all the available information \mathbf{Y}_t are derived following from Durbin [33]. The optimal predictor $\mathbf{a}_{t|t-1}$ and its MSE matrix are updated using

$$\mathbf{a}_t = \mathbb{E}(\alpha_t | \mathbf{Y}_t) = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{v}_t, \tag{6}$$

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}_{t|t-1}, \tag{7}$$

where $\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t$ and it is assumed to be non-singular. Thus, taken together ((4) and (5)) and ((6) and (7)) make up the Kalman filter. The initial values for the Kalman filter may be specified as \mathbf{a}_0 and \mathbf{P}_0 or $\mathbf{a}_{1|0}$ and $\mathbf{P}_{1|0}$. If the process is stationary and the initial state $\alpha_0 \sim N(\mathbf{a}_0, \mathbf{P}_0)$, the Kalman filter can be initialized as $\mathbf{a}_0 = \mathbf{0}$, \mathbf{P}_0 or $\mathbf{a}_{1|0} = \mathbf{0}$, $\mathbf{P}_{1|0} = \mathbf{P}_0$. For non-stationary transition equations, the initial distribution of α_0 is specified in terms of a diffuse or non-informative prior [30].

Furthermore, the Kalman filter enables a likelihood to be computed via prediction error decomposition which opens the way for the estimation of unknown parameters of the model. The distinguishing feature of a time series model is that the observations are dependent. Thus, the likelihood function is defined using the conditional probability density function as

$$L(\mathbf{y} | \psi) = \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{Y}_{t-1}), \tag{8}$$

where $p(\mathbf{y}_t | \mathbf{Y}_{t-1})$ denotes the distribution of \mathbf{y}_t which is conditional on the observations at time $t - 1$, that is $\mathbf{Y}_{t-1} = \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_1\}$ [30]. For the measurement equation $\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_{t|t-1} + \mathbf{Z}_t(\alpha_t - \mathbf{a}_{t|t-1}) + \varepsilon_t$, its conditional distribution is normal with mean

$$\mathbb{E}(\mathbf{y}_t | \mathbf{Y}_{t-1}) = \hat{\mathbf{y}}_{t|t-1} = \mathbf{Z}_t \mathbf{a}_{t|t-1}$$

and covariance matrix, $\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t$, $t = 1, 2, \dots, T$. Therefore, for the Gaussian model, the log-likelihood function of (8) can be given as

$$\log L = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log |\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^T \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t, \tag{9}$$

where $\mathbf{v}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}$, $t = 1, \dots, T$. Since the conditional mean, $\hat{\mathbf{y}}_{t|t-1}$ of $\mathbf{y}_{t|t-1}$ is also the minimum mean square estimator (MMSE) of \mathbf{y}_t , the $N \times 1$ vector \mathbf{v}_t can be interpreted as a vector of prediction error and (9) is known as prediction error decomposition. To maximize the likelihood function with respect to the unknown parameters ψ , commonly a Newton–Raphson method will be employed. This method iteratively calculates the roots of the function until the parameter values reach a stable state (see, [33,38]).

2.2. State Space Formulation for ARIMA Model

This section presents the translation of the ARIMA model to state space representation and the estimation of the missing value using fixed-point smoothing. Let $\Delta y_t = y_t - y_{t-1}$, $\Delta_s y_t = y_t - y_{t-s}$, where Δ and Δ_s are non-seasonal and seasonal difference

operators, respectively. As suggested by Box et al. [39], differencing is continued until trend and seasonal effects have been eliminated, giving a new variable $y_t^* = \Delta^d \Delta_s^D y_t$ for $d, D = 0, 1, \dots$, which can be modeled as a stationary ARMA(p, q) model.

A non-stationary ARIMA(p, d, q) model can be defined as

$$\phi_p(B)\Delta^d y_t = \theta_q(B)\eta_t, \tag{10}$$

where B is the back-shift operator, $\phi_p(B) = 1 - \sum_{i=1}^p \phi_i B^i$ is an autoregressive operator,

$\theta_q(B) = 1 - \sum_{i=1}^q \theta_i B^i$ is an arbitrary moving average operator, $\Delta = 1 - B$ is the backward

difference operator and $\eta_t \sim NID(0, \sigma^2)$. For ease of simplicity, write $\Delta^d = (1 - B)^d = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_d B^d$. Moreover, the multiplicative seasonal ARIMA(p, d, q) × (P, D, Q)_s is given by

$$\phi_p(B)\Phi_P(B^s)\delta^d(B)\Delta^D(B^s)y_t = \theta_q(B)\Theta_Q(B^s)\eta_t$$

where $\Phi_P(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}$ and $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}$ are seasonal (stationary) AR and MA operators, respectively; $\delta^d(B) = (1 - B)^d$ is non-seasonal differencing and $\Delta^D(B^s) = (1 - B^s)^D$ is seasonal differencing.

For the ARIMA model defined by (10), its state space formulation is presented in accordance with Ansley and Kohn [29]. Let

$$w_t = \Delta^d y_t. \tag{11}$$

Following (10)

$$\phi_p(B)w_t = \theta_q(B)\eta_t. \tag{12}$$

Hence, w_t is a stationary ARMA process; now it is able to be written in the state space form as $w_t = \mathbf{z}'_t \alpha_t$ where $\mathbf{z}'_t = (1, 0, \dots, 0)$ and α_t is $(d + m) \times 1$ state vector with i th element $\alpha_{i,t}$ defined as

$$\begin{aligned} \alpha_{i,t} &= y_{t-i+1} \text{ for } i = 1, \dots, d, \\ \alpha_{i+d,t} &= \sum_{k=m-i+1}^m (\phi_k w_{t+i-1-k} - \theta_{k-1} \eta_{t+i-k}), \quad i = 1, \dots, m \end{aligned} \tag{13}$$

and $m = \max(p, q + 1)$. In (13), $\theta_0 = -1$, $\phi_i = 0$ for $i > p$ and $\theta_i = 0$ for $i > q$. The state transition equation for the state vector α_t will be given as

$$\alpha_t = \mathbf{T}\alpha_{t-1} + R\eta_t, \tag{14}$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_d & \mathbf{T}^* \\ \mathbf{0} & \mathbf{P} \end{bmatrix}, \tag{15}$$

$$\mathbf{T}_d = \begin{bmatrix} \delta_1 & \delta_2 & \dots & \delta_d \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix}, \quad \mathbf{T}^* = \begin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix},$$

$$\Delta^d = (1 - B)^d = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_d B^d,$$

$$\mathbf{P} = \begin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{m-1} & 0 & 0 & \dots & 1 \\ \phi_m & 0 & 0 & \dots & 0 \end{bmatrix}$$

and R is $(m + d) \times 1$ dimensional vector which is defined as $R = (1, 0, \dots, 1, -\theta_1, \dots, -\theta_{m-1})'$ and $\eta_t \sim NID(0, \sigma^2)$. The matrix \mathbf{P} in (15) is similar to the transition matrix for stationary processes. The first d elements of (14) form the state transition equation for the difference (11) while the matrix \mathbf{T}^* provides the link between the stationary and non-stationary parts of the model (11) and (12).

The likelihood function for the ARIMA model with missing observations can be formulated in two different ways: level formulation and difference formulation. Level formulation provides greater flexibility and is recommended for missing observations occurring not at regular pattern [29,40], which is applied in this paper. Let B be the lag operator and $-\delta_j$ be the coefficient of B^j in the expansion of $\Delta^d \Delta^D (B^s) = (1 - B)^d (1 - B^s)^D$. For the stationary ARMA(p, q) process w_t , let $\bar{\alpha}_t$ be the state vector of the process in state space form and define the augmented state vector $\alpha_t = (\bar{\alpha}_t' \ y_{t-1}^{*'})'$, where $y_{t-1}^* = (y_{t-1}, \dots, y_{t-d-sD})$. The transition equation for the augmented state vector is

$$\alpha_t = \begin{bmatrix} \bar{\alpha}_t \\ y_{t-1}^* \end{bmatrix} \left[\begin{array}{cccc|ccc} \mathbf{T} & & & & & \mathbf{0}' & \\ - & - & - & - & - & - & - \\ 1 & 0 & \dots & 0 & \delta_1 & \dots & \delta_k \\ & \mathbf{0}' & & & \mathbf{I}_{k-1} & & 0 \end{array} \right] \begin{bmatrix} \bar{\alpha}_{t-1} \\ y_{t-2}^* \end{bmatrix} + \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \eta_t, \quad t = 1, \dots, T,$$

where \mathbf{T} and \mathbf{R} are matrices associated with state vector $\bar{\alpha}_t$, \mathbf{I}_{k-1} is $k - 1$ dimensional identity matrix and $k = d + sD$. If y_t is observed for all $t = 1, \dots, T$, the measurement equation will be $y_t = (1, \mathbf{0}'_{m-1}, \delta_1, \dots, \delta_k) \alpha_t$, $t = 1, \dots, T$, where $\mathbf{0}_{m-1}$ is an $(m - 1)$ vector of zeros. The initial value for the Kalman filter can be taken at $t = k$ with $\mathbf{a}_{k+1|k} = (\mathbf{0}', y_{k+1}^*)'$ and the covariance matrix

$$\mathbf{P}_{k+1|k} = \begin{bmatrix} \bar{\mathbf{P}}_{1|0} & 0 \\ 0 & 0 \end{bmatrix},$$

where $\bar{\mathbf{P}}_{1|0} = \sigma^{-2} E(\bar{\alpha}_t \bar{\alpha}_t')$. The likelihood function generated by a Kalman filter that is applied to the differenced observations $\Delta^d \Delta_s^D y_t$ is the same as the likelihood function that is obtained by using the Kalman filter for the time series [40].

Estimating Missing Observations

Missing observations can be estimated by utilizing fixed-point smoothing to the level form of the ARIMA model. It proceeds by applying the Kalman filter and augmenting the state vector each time a missing observation is encountered [30]. After processing all observations, the augmented state vector provides the MMSE estimates of the missing data, with the corresponding MSEs derived from the augmented covariance matrix. Let y_t^\dagger be missing at time $t = \tau$. The state vector is augmented by y_τ^\dagger and is given as $\mathbf{z}' \alpha_\tau$. Then the fixed-point smoothing recursions, which is modified from Anderson and Moore [32] leads to

$$\bar{y}_{\tau|t} = \bar{y}_{\tau|t-1} + \mathbf{p}'_{\tau|t-1} \mathbf{z}_t f_t^{-1} v_t, \quad t = \tau, \dots, T, \tag{16}$$

$$\mathbf{p}_{\tau|t} = \mathbf{T}(\mathbf{I} - \mathbf{x}_t \mathbf{z}_t') \mathbf{p}_{\tau|t-1}, \quad t = \tau, \dots, T, \tag{17}$$

where $\mathbf{x}_t = \mathbf{P}_{t|t-1} \mathbf{z}_t f_t^{-1}$ and the initial values are $\bar{y}_{\tau|\tau-1} = \mathbf{z}' \mathbf{a}_{\tau|\tau-1}$ and $\mathbf{p}_{\tau|\tau-1} = \mathbf{P}_{\tau|\tau-1} \mathbf{z}$. The values of v_t , f_t and \mathbf{x}_t are all generated by the Kalman filter for the original state space model and the vector \mathbf{z}_t is being used to define the measurement equation. Whenever a missing observation is encountered at time t , then (16) and (17) lessen to $\bar{y}_{\tau|t} = \bar{y}_{\tau|t-1}$ and $\mathbf{p}_{\tau|t} = \mathbf{T} \mathbf{p}_{\tau|t-1}$, respectively. Moreover, the MSE of $\bar{y}_{\tau|T}$ is given by $\sigma^2 f_{\tau|T}$, where $f_{\tau|T}$ is obtained from the recursion

$$f_{\tau|t} = f_{\tau|t-1} - \mathbf{p}'_{\tau|t-1} \mathbf{z}_t f_t^{-1} \mathbf{z}_t' \mathbf{p}_{\tau|t-1}, \quad t = \tau, \dots, T$$

with $f_{\tau|\tau-1} = \mathbf{z}' \mathbf{P}_{\tau|\tau-1} \mathbf{z}$. The fixed-point smoothing algorithm, which is set up in this way, will result in an extremely efficient estimator for the missing value.

3. TAVC Estimation

3.1. Multiscale Change Point Detection in the Mean

Change point detectability depends on the interrelationship between the magnitude of changes and their distance to neighboring change points. Large changes are usually detectable even when close to others, but smaller changes need more distance to be identified. When data shows repeated large fluctuations and minor changes in data over prolonged stable periods, it demands a multiscale method to accurately detect all change points [3].

The change point model, which allows for multiple changes in the mean, is defined by

$$Y_t = f_t + \varepsilon_t = f_0 + \sum_{i=1}^p \mu_i \cdot \mathbb{I}(t \geq \tau_i + 1) + \varepsilon_t, \quad t = 1, \dots, T, \tag{18}$$

where f_t is piecewise deterministic signal with p change points at locations τ_i , $i = 1, \dots, p$, with $\tau_0 = 0$ and $\tau_{p+1} = T$. μ_i is mean at τ_i . The error term $\{\varepsilon_t\}_{t=1}^T$ is assumed to be a (weakly) stationary time series satisfying $\mathbb{E}(\varepsilon_t) = 0$ with finite LRV $\sigma^2 = \lim_{N \rightarrow \infty} \text{Var} \left(N^{-\frac{1}{2}} \sum_{t=1}^N \varepsilon_t \right) \in (0, \infty)$. These errors can be serially correlated and heavy-tailed. The objective is to detect changes in the mean and accurately estimate both the number and locations of change points.

To address this problem, we use a local testing method for data segmentation. This method applies a local test for a single change point adopted to detect and estimate multiple change points [3]. The procedure involves evaluating a test statistic of the form $\widehat{\sigma}_{s,e}^{-1} |\mathcal{T}_{s,k,e}|$ to a threshold, say Γ . A change point detector $\mathcal{T}_{s,k,e}(Y)$ is computed on the segment $\{Y_t, s < t \leq e\}$ as

$$\mathcal{T}_{s,k,e}(Y) = \sqrt{\frac{(k-s)(e-k)}{e-s}} (\bar{Y}_{s:k} - \bar{Y}_{k:e}) \tag{19}$$

at some location $0 \leq s < k < e \leq T$, which is determined in a method-specific way, usually it can either be fixed, random or data-dependent and $\bar{Y}_{a:b} = \frac{1}{b-a} \sum_{t=a+1}^b Y_t$. Here, $\sigma_{s,e}^2$ denotes variability in $\{Y_t\}_{t=s+1}^e$, $\widehat{\sigma}_{s,e}^2$ is its estimator. If the error term $\{\varepsilon_t\}_{t=1}^T$ is assumed to be stationarity, then the appropriate choice of the variance will be $\sigma_{s,e}^2 = \sigma_{e-s}^2$ is the TAVC, which depends on the scale $D = e - s$ is given in (1). As the result, if $\widehat{\sigma}_{s,e}^{-1} |\mathcal{T}_{s,k,e}| > \Gamma$, then it indicates that the data segment $\{Y_t\}_{t=s+1}^e$ contains a change point within the interval; otherwise, it signals no such change point has been detected in the given data section [3].

However, in the presence of serial correlation, distinguishing a true mean shift from random fluctuations becomes challenging. Consequently, it is imperative to have a robust variance estimator which effectively handles the variability in the data section under consideration. For such a multiscale procedure, using an estimator of the global LRV instead of $\sigma_{s,e}^2$, regardless of the interval length used to compute the detector statistic, can lead to spurious change point estimations. Therefore, a variance that systematically captures the variability in the data section under serial correlation for standardization of multiscale change point detectors will be σ_D^2 estimator [27].

3.2. Robust Estimation of Multiscale TAVC

Here, the derivation σ_D^2 estimator is presented. For notational convenience, assume that the data length D is an even number. To account for temporal dependency, group the dataset into blocks of the same size $G = \frac{D}{2}$. Then, for some starting point $h \in (0, \dots, G - 1)$ with number of blocks $N(h) = \lfloor \frac{(T-h-G)}{G} \rfloor$, define

$$\bar{Y}_{j,h} = \frac{1}{G} \sum_{t=jG+h+1}^{(j+1)G+h} Y_t, \text{ and } \pi_{j,h} = \frac{G(\bar{Y}_{j,h} - \bar{Y}_{j-1,h})^2}{2} \text{ for } j = 1, \dots, N(h).$$

Similarly, define

$$\bar{\varepsilon}_{j,h} = \frac{1}{G} \sum_{t=jG+h+1}^{(j+1)G+h} \varepsilon_t \text{ and } \tilde{\pi}_{j,h} = \frac{G(\bar{\varepsilon}_{j,h} - \bar{\varepsilon}_{j-1,h})^2}{2}.$$

Then, the mean of $\tilde{\pi}_{j,h}$ expressed as

$$\hat{\sigma}_D^2 = \frac{1}{N(h)} \sum_{j=1}^{N(h)} \tilde{\pi}_{j,h} \tag{20}$$

considers time dependency in local data sections of length $D = 2G$. In addition, we have $\mathbb{E}(\hat{\sigma}_D^2) - \sigma_D^2 = o(1)$ for $D \rightarrow \infty$ (see Theorem 1 below). The natural estimate of TAVC is $N(h)^{-1} \sum_{j=1}^{N(h)} \pi_{j,h}$ but it is usually contaminated by jumps or the mean shifts. Furthermore, its counterpart $\hat{\sigma}_D^2$ is indicative of the level of variability but being inaccessible (as it is defined with $\bar{\varepsilon}_{j,h}$ in place of $\bar{Y}_{j,h}$).

Thus, the M -estimation technique discussed in [41] is adopted to achieve the required estimator as detailed in [42]. Consider a non-decreasing influence function $\rho : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\rho(x) = \begin{cases} -\log(2), & x \leq -1, \\ \log(1 + x + \frac{x^2}{2}), & -1 \leq x \leq 0, \\ -\log(1 - x + \frac{x^2}{2}), & 0 \leq x \leq 1, \\ \log(2), & x \geq 1. \end{cases} \tag{21}$$

Then, the estimator of TAVC at D and h , $(\hat{\sigma}_{D,h}^2)$, is defined as the solution of the M -estimation equation

$$J_{D,h}(\theta) = \frac{1}{N(h)} \sum_{j=1}^{N(h)} \rho_v(\pi_{j,h} - \theta) = 0, \tag{22}$$

where $\rho_v(x) = v^{-1}\rho(vx)$ for some $v > 0$. The condition for v will be set later under Section 5.1. Equation (22) may have more than one solution; in such a case, any of them can be used to define $\hat{\sigma}_{D,h}^2$.

3.3. Consistency of the Scale-Dependent TAVC Estimator

Assumption 2.

- (i) The error process is assumed to be linear, i.e., $\varepsilon_t = \sum_{k=0}^{\infty} w_k \eta_{t-k}$ where $\{\eta_t\}_{t \in \mathbb{Z}}$ is a sequence of iid random variables and $|w_k| \leq \Xi(k+1)^{-\beta}$ for some constants $\beta > 2.5$ and $\Xi > 0$ for all $k \geq 0$.
- (ii) $\exists \sigma > 0$ s.t. $\sigma^2 = (\sum_{k \geq 0} w_k)^2 < \infty$, where $\sigma^2 \geq c_\sigma$.
- (iii) On the distribution of $\{\eta_t\}_{t \in \mathbb{Z}}$, we work under any one of the two following scenarios.
 - a. For $q > 4$, $\|\eta_1\|_q = (\mathbb{E}(|\eta_1|^q))^{\frac{1}{q}} < \infty$.
 - b. For $R_\eta > 0$ and $\omega > 0$, $\|\eta_1\|_q \leq R_\eta q^\omega$ for all $q \geq 1$.

The assumptions stated above have the following underlying basic reasons. Linearity of the error process $\{\varepsilon_t\}_{t=1}^T$ assumed in Assumption 2 (i) enables controlling of the functional dependence in $\{\pi_{j,h}\}_{j=1}^{N(h)}$. This allows the temporal dependence of the error process to decay at an algebraic rate. Whereas Assumption 2 (ii) declared to secure positivity of the LRV in such a way that it is well-defined while computing change point detector statistics. Assumption 2 (iii) (a) characterizes the heavy-tailedness of the error process $\{\varepsilon_t\}_{t=1}^T$. While Assumption 2 (iii) (b) claims a stronger condition that requires sub-Weibull tail behavior on

$\{\eta_t\}_{t \in \mathbb{Z}}$. This includes sub-Gaussian distribution for $\omega = \frac{1}{2}$ such that $(\mathbb{E}(|\eta_1|^q))^{\frac{1}{q}} \leq R_\eta q^{\frac{1}{2}}$ and sub-exponential distribution for $\omega = 1$, where $(\mathbb{E}(|\eta_1|^q))^{\frac{1}{q}} \leq R_\eta q$ which are defined based on the growth behavior of moment [43].

Definition 1. Let $\{a_n\}$ and $\{b_n\}$ be positive sequences, then $a_n \asymp b_n$ if there exists $\kappa_1, \kappa_2 > 0$ such that $\kappa_1 \leq a_n/b_n \leq \kappa_2$ as $n \rightarrow \infty$; $a_n \ll b_n$ or $a_n = o(b_n)$ means $a_n/b_n \rightarrow 0$; $a_n \lesssim b_n$ or $a_n = \mathcal{O}(b_n)$ if there exists a constant $C > 0$ such that $a_n/b_n \leq C$ as $n \rightarrow \infty$.

Theorem 1. Under Assumption 2. Let TAVC be defined as

$$\tilde{\sigma}_D^2 = \text{Var} \left(\frac{1}{\sqrt{D}} \left(\sum_{t=1}^G \varepsilon_t - \sum_{t=G+1}^D \varepsilon_t \right) \right) \text{ for } D = 2G.$$

Then provided that $v \asymp \sqrt{\frac{pG}{T}}$, under Assumption 2 (iii) (a) the estimator $\hat{\sigma}_{D,h}^2$ satisfies

$$|\hat{\sigma}_{D,h}^2 - \tilde{\sigma}_D^2| = \mathcal{O}_P \left(\sqrt{\frac{pD}{T}} + \max \left\{ \left(\frac{D}{T} \right)^{\frac{q-2}{q+2}}, \sqrt{\frac{D \log T}{T}} \right\} \right) \tag{23}$$

and under Assumption 2 (iii) (b),

$$|\hat{\sigma}_{D,h}^2 - \tilde{\sigma}_D^2| = \mathcal{O}_P \left(\sqrt{\frac{pD}{T}} + \sqrt{\frac{D \log^{4\omega+3} T}{T}} \right) \tag{24}$$

for any fixed $h \in \{0, \dots, G - 1\}$. In addition, $\tilde{\sigma}_D^2$ satisfies

$$|\tilde{\sigma}_D^2 - \sigma_D^2| = \mathcal{O}(D^{-1}) \tag{25}$$

Proof. The concept of the proof is similar to the one discussed in [27]. \square

Theorem 1 illustrates that the proposed robust estimator consistently approximates the TAVC at scale D . The estimation error is decomposed into two components: the error from approximating σ_D^2 by $\tilde{\sigma}_D^2$ in (25), and the error in estimating $\tilde{\sigma}_D^2$ by $\hat{\sigma}_{D,h}^2$. The second error accounts explicitly for the influence of multiple mean shifts on the estimator through the term $\sqrt{\frac{pD}{T}}$ in (23) and (24), while the remaining terms capture the effect of the error distribution. Consequently, Theorem 1 establishes that

$$|\tilde{\sigma}_D^2 - \sigma^2| = \mathcal{O} \left(\frac{1}{D} \right) \tag{26}$$

therefore, as D increases, the scale- D TAVC approximates the global LRV as expected.

3.3.1. Setting the Maximum Time-Scale for TAVC Estimation

In this section, we briefly discuss the maximum time scale or length required for D in the data section. As shown in (25), the error resulting from approximating σ_D^2 with $\tilde{\sigma}_D^2$ diminishes as D increases. Conversely, the error in estimating $\tilde{\sigma}_D^2$ with $\hat{\sigma}_{D,h}^2$ increases with D , as given in (23) and (24). This increase is attributed to the growing effect of mean shifts with an increase in D , coupled with the decrease in the number of available blocks. To balance these two errors, Mcgonigle and Cho [27] proposed setting a maximum time-scale, denoted as M , for use with a multiscale CPD algorithm. Particularly when the change point detector $\mathcal{T}_{s,k,e}$ has $e - s \leq M$, we scale the detector using the estimator of σ_{e-s}^2 , the TAVC at scale $D = e - s$. However, if $e - s > M$, it is recommended to scale the

detector using the estimator of σ_M^2 , the TAVC at the maximum time-scale M , ensuring that $|\sigma_{e-s}^2 - \sigma_M^2| = \mathcal{O}(M^{-1})$.

4. Applications of Robust TAVC Estimator

For multiple CPD, the suggested TAVC estimator can be used with the MOSUM method and WBS algorithms that scan the data using detector statistics of the type (19).

4.1. Multiscale MOSUM Procedure

The MOSUM procedure is utilized for simultaneously detecting and locating points where the various changes have occurred [5,13,44]. The MOSUM statistic used for detecting a change in mean at time point k for a given bandwidth G is defined as

$$T_G = \max_{G \leq k \leq T-G} \frac{|T_G(k)|}{\hat{\sigma}_k^2} \tag{27}$$

with MOSUM detector

$$T_G(k) = T_G(k; Y) = \frac{1}{\sqrt{2G}} \left(\sum_{t=k+1}^{k+G} Y_t - \sum_{t=k-G+1}^k Y_t \right), \quad k = G, \dots, T - G$$

and $\hat{\sigma}_k^2$ is the estimator of error variance.

As the exact probability distribution of the MOSUM statistic is not well known, we use an asymptotic result to develop a MOSUM-based test for changes in the mean [13,45]. By utilizing the asymptotic result of a MOSUM-based test with an asymptotic significance level $\alpha \in (0, 1)$, we reject the null hypothesis $H_0 : p = 0$ in favor of the alternative $H_1 : p \geq 1$ when the MOSUM statistic T_G , as indicated in (27), surpasses the asymptotic critical value given by

$$\Gamma_T(G; \alpha) = \frac{b_{G,T} + c_\alpha}{a_{G,T}} \text{ with } c_\alpha = -\log \log \frac{1}{\sqrt{1 - \alpha}}.$$

Here, $a_{G,T}$ and $b_{G,T}$ are properly chosen scaling and shifting factors that depend on the data length T and the bandwidth G (see [13]). In practical situations, the performance of the methodology is significantly determined by choice of the bandwidth G . Employing of mono-bandwidth MOSUM procedure yields consistent estimators with optimal localization rate only under the assumption that $\min_{0 \leq i \leq p} (\tau_{i+1} - \tau_i) > 2G$, $\min_{1 \leq i \leq p} \mu_i^2 G / \log(T/G) \rightarrow \infty$ as $T \rightarrow \infty$ (see, Theorem 3.2 of [13] and Corollary D1 of [46]). When the change points are heterogeneous, i.e., if the data sequence contains both large changes over short intervals and small changes over a long stretches, the mono-bandwidth MOSUM procedure fails to achieve consistency in the estimation of change points. One way to address this issue is by utilizing a multiscale MOSUM procedure that incorporates bottom-up merging [47] and localized pruning [48] techniques to identify a set of candidate change point estimators. A multiscale MOSUM procedure with bottom-up merging has low computational complexity [45] hence, we will employ this procedure in our study.

Let $\mathcal{G} = \{G_\iota, 1 \leq \iota \leq H : G_1 < \dots < G_H\}$ represent a range of bandwidths, and let $\chi(G)$ indicate the set of estimators identified with a particular bandwidth $G \in \mathcal{G}$. Next, check if there is a previously accepted change-point \hat{k}_1 in the final set of estimators χ obtained using the finest bandwidth G_1 , which is in proximity to the current candidate \hat{k}_2 detected with bandwidth G_2 . Proceed to accept the change point $\hat{k}_2 \in \chi(G_2)$ only if $|\hat{k}_2 - \hat{k}_1| > \eta G_2$; otherwise, reject \hat{k}_2 as it has already been identified by \hat{k}_1 . Repeat this process sequentially for $\iota = 3, \dots, H$. The recommended value of η is 0.4 to have a good estimate of the change point and its localization [45].

Then, we apply the robust TAVC estimator to the multiscale MOSUM procedure with bottom-up merging to standardize the MOSUM detector. For each $G_\iota \in \mathcal{G}$, the TAVC at scale $D = 2G_\iota$ is estimated by $\hat{\sigma}_{2G_\iota}^2$ solving (22), under the condition that $2G_\iota \leq M$, where M represents the maximum scale, as explained in Section 3.3.1. Subsequently, the global

estimator $\hat{\sigma}_k$ in (27) will be substituted with $\hat{\sigma}_{2G_i}$ to standardize the MOSUM detector $T_{G_i}(k)$, while for $2G_i > M$, we will use $\hat{\sigma}_M$ instead of $\hat{\sigma}_{2G_i}$ to scale the MOSUM detector. Such a way of standardizing will correctly reflect the extent of variation over the sliding window and accurately estimate the change points. The algorithm for the multiscale MOSUM method combined with the TAVC estimator can be found in [27].

4.2. WBS2 Procedure

Binary segmentation, introduced by [49], is a generic technique that provides a simple framework to extend single change point detection to multiple changes. It initially scans the entire dataset for a change point using a CUSUM statistic as defined in (19). If a significant change point is found, the data are split into two segments, and the process is repeated recursively until no change point is detected. It has various extensions like circular binary segmentation [50], wild binary segmentation (WBS) [51], and WBS2 [52], each proposed to address the shortcomings associated with the earlier versions. WBS2 is designed to overcome the issues of lack of computational adaptivity and incomplete solution path associated with WBS. We combine the WBS2 algorithm with the proposed TAVC estimator for multiple change point detection in the mean under model (18).

Suppose that $\mathcal{B}_{s,e} = \{(l,r) \in \mathbb{Z}^2 : s \leq l < r \leq e, r - l > 1\}$ represents the collection of all possible intervals within $\{s + 1, \dots, e\}$ for some $0 \leq s < e \leq T$. Let $\mathcal{H}_{s,e} \subset \mathcal{B}_{s,e}$ selected either randomly or deterministically (we use a deterministic grid selection approach) with $|\mathcal{H}_{s,e}| = \min(\mathcal{M}, |\mathcal{B}_{s,e}|)$ for some given sub-samples $\mathcal{M} \leq \frac{T(T-1)}{2}$ drawn from $\{1, \dots, T\}$. Starting with $(s, e) = (0, T)$, we identify the candidate change point that maximizes the absolute CUSUM statistic as

$$(s_0, k_0, e_0) = \arg \max_{\substack{(l,k,r): l < k < r \\ (l,r) \in \mathcal{H}_{s,e}}} \frac{|\mathcal{T}_{l,k,r}|}{\hat{\sigma}_{r-l}} \text{ with } \frac{|\mathcal{T}_{s_0,k_0,e_0}|}{\hat{\sigma}_{e_0-s_0}} > \Gamma \tag{28}$$

where $\hat{\sigma}_{r-l}^2$ denotes the estimator of the TAVC with length $D = r - l$ (when $r - l$ is odd, we use $\hat{\sigma}_{r-l-1}^2$ instead) and Γ is the threshold. Γ is determined as $\Gamma = C\sqrt{2 \log T}$ where C is a constant taken based on the recommendation of Fryzlewicz [51]. As described in Section 3.3.1, for standardizing of CUSUM statistic, we use $\hat{\sigma}_M$ for any interval of length greater than the maximum scale M . WBS2 marks the first candidate change point (s_0, k_0, e_0) which satisfies Equation (28) and then splits the entire dataset into $\{Y_t\}_{t=s+1}^{k_0}$ and $\{Y_t\}_{t=k_0+1}^e$ on which the same step be repeated on the partitioned segments. If there is no such (s_0, k_0, e_0) found or the minimum segment length defined by the user has been attained, the search for change point will end up on $\{Y_t\}_{t=s+1}^e$ and no change point is returned as the final result. An algorithm for WBS2 combined with TAVC estimator can be found in [27].

5. Numerical Results

This section outlines simulation and real data applications to validate theoretical findings.

5.1. Parameter Tuning for CPD

The tuning parameter ν in (22) is defined as $\nu = \nu_h = \bar{\pi}_h^{-1} \sqrt{G/T}$ where $\bar{\pi}_h$ is a fixed constant for a given starting value $h \in \{0, 1, \dots, G - 1\}$. For the choice of $\bar{\pi}_h$, we employ both the trimmed mean of $\{\pi_{j,h}\}_{j=1}^{N(h)}$ proposed by Chen et al. [42] and scaled median value of $\{\pi_{j,h}\}_{j=1}^{N(h)}$ suggested by Mcgonigle et al. [53]. A trimmed mean is defined as

$$\bar{\pi}_{[1],h} = \frac{1}{[3N(h)/4] - \lceil N(h)/4 \rceil + 1} \sum_{j=\lceil N(h)/4 \rceil}^{\lfloor 3N(h)/4 \rfloor} \pi_{(j),h}, \tag{29}$$

where $\pi_{(1),h} \leq \dots \leq \pi_{(N(h)),h}$ are ordered $\{\pi_{j,h}\}_{j=1}^{N(h)}$. Whereas an appropriately scaled median value of $\{\pi_{j,h}\}_{j=1}^{N(h)}$ that ensures unbiasedness is given as

$$\bar{\pi}_{[2],h} = \Delta \cdot \text{Median}\{\pi_{1,h}, \dots, \pi_{N(h),h}\}, \tag{30}$$

where Δ is a scaling constant and it is taken to be $\Delta = 2.125$. Hence, we employ $\bar{\pi}_{[i],h}, i = 1, 2$, in setting the parameter ν_h in order to observe their effect on the simulation result. The other is the maximum time scale M , recommended by Mcgonigle and Cho [27] at which the TAVC to be estimated is $M = \lfloor 2.5\sqrt{T} \rfloor$. For tuning the parameters of the multiscale MOSUM procedure, we follow [27,46]. Moreover, we use the R package ‘mosum’ available on CRAN [45] suggested default settings to define additional tuning parameters, and set $\eta = 0.4$ and $\alpha = 0.05$. Whereas, for tuning the parameters of the WBS2 algorithm, we followed the approaches discussed at [3,27,51].

5.2. Simulation Results

In this section, we conduct an extensive simulation study comparing the performance of the proposed ARIMA state space and Kalman smoothing with the various competitive univariate time series imputation methods for the series with mean shift. Moreover, our proposed CPD approach is compared against that of various competing methods that consider serial dependence under (18) and are easily accessible in R, with different scenarios for generating serially correlated $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. We generate 1000 realizations with $T = 1000$ under each setting by introducing a mean shift in the data where $u_t \stackrel{iid}{\sim} N(0,1)$ unless otherwise stated.

5.2.1. Set-Up

This section first presents settings to examine the performance of missing value imputation methods. The data are generated under different scenarios including change points (mean shift). Once the data are generated, we artificially deleted varying percentages (20%, 30%, and 40%) of values with missing completely at random (MCAR) mechanism using ‘missMethods’ R package available on CRAN [54]. Then, the missing values for simulated data are imputed using various univariate time series imputation methods, namely: interpolation (linear, spline and Stineman), mean imputation, last observation carried forward (LOCF), and moving average (simple (SMA), linear weighted (LWMA), and exponential weighted (EWMA)) methods. We employ ‘imputeTS’ R package available on CRAN [55] to estimate the missing values. To show the robustness of the proposed ARIMA state space and Kalman smoothing in estimating missing values in the presence of change point (mean shift), we calculate the proportion of change points estimated using MOSUM CPD method scaled by the TAVC estimator.

- (M1*) f_t undergoes $p = 2$ with change points at $(\tau_1, \tau_2) = (200, 600)$ and $(f_0, f_1, f_2) = (0, 10, 20)$, and the model follows seasonal ARIMA $(0, 1, -0.4)(0, 1, -0.56)_{12}$.
- (M2*) f_t undergoes $p = 2$ with change points at $(\tau_1, \tau_2) = (300, 700)$ and $(f_0, f_1, f_2) = (0, 30, 10)$, and the model follows seasonal ARIMA $(0.5, 1, -0.8)(0, 1, 0)_{12}$.
- (M3*) f_t undergoes $p = 4$ with change points at $(\tau_1, \tau_1, \tau_2, \tau_4) = (200, 400, 600, 800)$ and $f_i = \lceil T * i / 16 \rceil, i = 1, \dots, 4$, and the model follows seasonal ARIMA $(0.9, 0, 0)(0, 1, 0)_{12}$.

Models (M1*) and (M2*) represent nonstationary seasonal series scenarios and model (M3*) is strongly autocorrelated seasonal series. Model (M1*) is adopted based on modeling air passenger time series data, which is commonly used in much literature [39,40,56] with $\sigma = 0.0025$. While model (M2*) is adopted from Dette et al. [14] with modification of introducing nonstationarity to make the model a nonstationary seasonal ARIMA model with $\sigma = 1$. Finally, model (M3*) is taken from Cho and Fryzlewicz [28] and we added seasonality to produce strongly autocorrelated seasonal $AR(1)$ series with $\sigma = 1$. In model (M1*), (M2*), and (M3*), we injected 20%, 40% and 30% of the missing values, respectively. For the detection of change points, we apply the MOSUM procedure scaled with the TAVC estimator by choosing the tuning parameter ν in (22) using (29), the bandwidth (G) is set to be 60 and $\alpha = 0.05$.

Furthermore, we investigate the performance of MOSUM and WBS2 procedures scaled with the TAVC estimator as described in Section 4 on simulated datasets, in comparison

with various methods when there is no change ($p = 0$) and there is at least one change ($p \geq 1$) under different error structures. Moreover, we study the condition where $f_t = 0$ to examine the performance in size control.

- (M1) $\varepsilon_t = u_t$, and f_t undergoes $p = 4$ with change points at $(\tau_1, \tau_2, \tau_3, \tau_4) = (200, 400, 600, 800)$ and $(f_0, f_1, f_2, f_3, f_4) = (0, 2, -2, 2, -2)$.
- (M2) $\varepsilon_t = u_t$, where u_t is an independent random variable following t_5 distribution, and f_t and τ_i are same as in (M1).
- (M3) $\varepsilon_t = u_t + 0.9\varepsilon_{t-1}$ follows AR(1) process with $\sigma_u = 0.4359$, and f_t undergoes $p = 4$ with change points τ_i as in (M2) and $(f_0, f_1, f_2, f_3, f_4) = (0, 2, -1, 2, -1)$.
- (M4) $\varepsilon_t = u_t + 0.5\varepsilon_{t-1} + 0.3\varepsilon_{t-2}$ follows AR(2) process with $\sigma_u = 0.6676$, and f_t and τ_i undergo as in (M2).
- (M5) $\varepsilon_t = u_t - 0.9u_{t-1}$ follows MA(1) process and f_t and τ_i undergo as in (M2).
- (M6) $\varepsilon_t = \sigma_t^2 u_t$ with $\sigma_t^2 = 0.5 + 0.4\varepsilon_{t-1}^2$ follows ARCH(1) model, and f_t and τ_i undergo as in (M5).

Model (M1) is widely used in various kinds of literature, and (M2) is taken from [46] to evaluate how well the proposed approach performs other than Gaussian error. In models (M3) and (M4), ε_t has a high degree of auto-correlations, which is adapted from Cho and Fryzlewicz [28]. For model (M5), long-run variance approaches zero, and thus it is difficult to obtain an accurate estimate and may affect the performance of the methods. Model (M6) is adopted from Mcgonigle and Cho [27] to examine whether a method works well for non-linear processes. We apply the proposed TAVC robust estimator to both multiscale MOSUM and WBS2 procedures, denoted as MOSUM.TAVC[i] and WBS2.TAVC[i], respectively. Here, i indicates the selection of tuning parameter $\bar{\pi}_{[i],h}$ related to the parameter ν ; refer to (29) and (30). The selection of tuning parameters is discussed in Section 5.1.

For comparison, we include DeCAFS [9], DepSMUCE [14], and WCM.gSa [28]. DeCAFS is adaptable to the problem of detecting multiple change points in the mean under the assumption that the error term follows the stationary AR(1) model. DepSMUCE extends the SMUCE procedure of [57] to dependent data by estimating the long-run variance (LRV) using a difference-type estimator. The WCM.gSa is adapted to detect multiple change points in the mean of an otherwise stationary, autocorrelated, linear time series. It combines solution path generation based on the wild contrast maximization principle, and an information criterion-based model selection strategy termed the gappy Schwarz algorithm. The significance levels $\alpha \in \{0.05, 0.1\}$ are considered for the DepSMUCE method. The tuning parameters not specified herein are implemented according to the authors' recommendations.

5.2.2. Results

The comparative simulation study results on missing value imputation, which encompassed 1000 realizations for $T = 1000$, generated according to models (M1*)–(M3*) with $p = 2, 4$, are succinctly summarized in Table 1. We report the average mean absolute percentage error (MAPE), and root mean square error (root MSE) over 1000 realizations to evaluate the performance of various missing value imputation methods. In addition, we also report the distribution of the estimated number of change points and the average covering metric (CM, defined later) for the imputed data returned by applying the MOSUM procedure scaled with the TAVC estimator. Methods that achieve the best performance based on the evaluation criteria for each scenario are shown in bold.

MAPE and root MSE between the actual time series value y_t and the corresponding imputed value \hat{y}_t can be defined as:

$$MAPE(y_t, \hat{y}_t) = \frac{\sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{|y_t|} * 100\%}{T}, \quad Root\ MSE(y_t, \hat{y}_t) = \sqrt{\frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{T}}$$

where T is the time period (number of observations).

Table 1. Performance comparisons of various univariate imputation methods for model (M1*)–(M3*) over 1000 realizations for $T = 1000$.

Model	Method	MAPE	Root MSE	$(\hat{p} - p)$						CM
				≤ -2	-1	0	1	2	≥ 3	
(M1*)	ARIMA.Kalman	0.539	3.940	0.000	0.000	0.957	0.041	0.002	0.000	0.992
	Linear interpolation	0.581	3.808	0.000	0.000	0.938	0.057	0.005	0.000	0.989
	Spline interpolation	0.603	4.075	0.000	0.000	0.914	0.078	0.006	0.002	0.986
	Stineman interpolation	0.570	3.727	0.000	0.000	0.941	0.055	0.004	0.000	0.990
	Mean imputation	6.051	36.109	0.040	0.209	0.355	0.244	0.108	0.044	0.571
	LOCF	0.911	5.817	0.000	0.000	0.804	0.167	0.026	0.003	0.969
	SMA	0.915	6.055	0.000	0.000	0.811	0.170	0.017	0.002	0.970
	LWMA	0.800	5.368	0.000	0.000	0.861	0.126	0.011	0.002	0.978
	EWMA	0.709	4.785	0.000	0.000	0.895	0.092	0.013	0.000	0.982
(M2*)	ARIMA.Kalman	160.729	4.430	0.000	0.000	0.933	0.060	0.007	0.000	0.987
	Linear interpolation	199.100	5.888	0.000	0.009	0.896	0.084	0.011	0.000	0.979
	Spline interpolation	251.705	7.987	0.021	0.095	0.716	0.129	0.032	0.007	0.916
	Stineman interpolation	206.664	6.025	0.000	0.010	0.896	0.083	0.011	0.000	0.978
	Mean imputation	566.688	6.512	0.010	0.149	0.777	0.056	0.008	0.000	0.924
	LOCF	235.137	6.560	0.005	0.023	0.844	0.113	0.013	0.002	0.967
	SMA	161.952	5.055	0.000	0.002	0.919	0.073	0.006	0.000	0.985
	LWMA	165.832	5.166	0.000	0.002	0.917	0.073	0.008	0.000	0.984
	EWMA	173.513	5.369	0.000	0.003	0.909	0.081	0.007	0.000	0.983
(M3*)	ARIMA.Kalman	37.006	3.454	0.000	0.000	0.887	0.112	0.001	0.000	0.988
	Linear interpolation	16.695	3.966	0.000	0.000	0.892	0.107	0.001	0.000	0.989
	Spline interpolation	19.823	4.279	0.000	0.000	0.895	0.103	0.002	0.000	0.989
	Stineman interpolation	15.966	3.484	0.000	0.000	0.889	0.110	0.001	0.000	0.988
	Mean imputation	1438.254	49.049	0.032	0.219	0.655	0.087	0.006	0.001	0.913
	LOCF	25.292	4.805	0.000	0.000	0.901	0.098	0.001	0.000	0.989
	SMA	18.608	4.317	0.000	0.000	0.890	0.108	0.002	0.000	0.988
	LWMA	16.948	3.960	0.000	0.000	0.892	0.106	0.002	0.000	0.988
	EWMA	15.783	3.713	0.000	0.000	0.889	0.109	0.002	0.000	0.988

Table 1 shows that ARIMA state space and Kalman smoothing imputation outperformed other methods in most scenarios, as indicated by their lower MAPE and root MSE values. Among the other methods, interpolation and moving average methods provided more accurate approximations than mean imputation and LOCF methods. Conversely, mean imputation, which relies on a single measure (mean, median, or mode) to fill in all missing values, exhibited the poorest performance. Moreover, the number of change points is estimated with great accuracy for the data imputed using the ARIMA state space and Kalman smoothing method, as indicated by the model selection accuracy, $|\hat{p} - p|$ and the covering metric. The proportion of the correctly estimated number of change points and the covering metric for the data imputed by using ARIMA state space and Kalman smoothing almost approaches one in many of the cases. Furthermore, the proposed imputation method has not resulted in under-detecting change points, but a very small percent of spurious change points are observed. Overall, the proposed approach outperformed most other univariate imputation methods in the majority of scenarios. Imputation by interpolation and moving average methods yields a relatively good estimate of the missing value for the data in the presence of a mean shift. Nevertheless, the worst false positive and false negative change points are detected for the data imputed by using the mean imputation method. Therefore, we will employ ARIMA state space and the Kalman smoothing method to estimate and impute missing values for the real data application and determine change points by applying multiscale CPD methods scaled with TAVC as discussed in Section 4.

While this method improves the handling of missing data by offering the nearest estimated value, it exhibits a high computational complexity in comparison to other univariate imputation techniques, particularly when applied to large datasets. This complexity is due

to the algorithm’s iterative nature, which updates the estimate of the state vector as new observations become available [37,56].

Table 2 summarizes the results of the simulation study for CPD, showing the performance of different methods for $T = 1000$ generated over 1000 realizations as in (M1)–(M6) with $p \in \{0, 4\}$. We report the proportion of realizations of $\hat{p} \geq 1$ when $p = 0$ (in ‘size’ column) and the summary of estimated change points based on the distribution of \hat{p} . We also present the averaged CM and relative mean squared error (RMSE) over 1000 realizations when $p = 4$. The modal value of the distribution of the estimated number of change points with the highest value for each scenario is shown in bold. Based on the evaluation criteria, the methods that attained the highest performance level when $p = 4$ are underlined. From the theoretical point of view, RMSE is one the most common error metrics that is applied in various literatures [27,28] to evaluate the performance of change point detection methods, which measures the deviation between the estimated change point positions and the true change point locations. It assigns a higher weight to larger errors due to its squaring effect, making it sensitive to substantial deviations in change point detection, which is essential in the context of time series analysis. Moreover, RMSE serves to evaluate the performance of change point detection methods in the presence of missing data, illustrating how accurately the detected change points correspond to the true points, even in light of imputation errors. It also provides insight into the impact of imputed data on the accuracy of change point estimates. If imputation is flawed, RMSE will increase, reflecting greater discrepancies between estimated and actual change points. RMSE is defined as:

$$RMSE = \frac{\sum_{t=1}^T (f_t - \hat{f}_t)^2}{\sum_{t=1}^T (f_t - \hat{f}_t^*)^2}$$

where \hat{f}_t is the piecewise constant signal constructed with the set of estimated change point locations $\hat{\tau}$, and \hat{f}_t^* is an oracle estimator constructed with the true change point locations, τ_i .

Covering metric is another evaluation metric, whose concept is a type of clustering metric that aligns with the aim that CPD fundamentally seeks to partition the time series into separate segments characterized by a stable data-generating process [58]. It measures the overlap between predicted region segments produced from the estimated change point and the ground truth regions obtained from the true change points [59]. Moreover, the covering metric captures how well the method detects structural changes in time series data, even when some points are imputed. Thus, it is recommended as an evaluation metric to measure the performance of CPD algorithms. Let \mathcal{S} be a segmentation of the interval $\{1, 2, \dots, T\}$ into disjoint sets R_i s.t. R_i is the segment $\{\tau_{i-1} + 1, \dots, \tau_i\}$ for the true change locations. In the same way, the estimated change locations $\{\hat{\tau}_i\}_{i=1}^{\hat{p}}$ generates a partition \mathcal{S}' of segments R' . Then, we define CM as

$$CM(\mathcal{S}, \mathcal{S}') = \frac{1}{T} \sum_{R \in \mathcal{S}} |R| \max_{R' \in \mathcal{S}'} \frac{|R \cap R'|}{|R \cup R'|}$$

The value of CM lies [0, 1]. Together, RMSE and the covering metric provide a balanced evaluation of the performance, accounting for both exactness (RMSE) and the method’s ability to cover true structural changes (covering metric), making them valuable tools for evaluating CPD in datasets with missing values.

Empirical simulation results presented in Table 2 show that, overall, MOSUM.TAVC tends to have more false positives in terms of size control compared to WBS2.TAVC. This is because the MOSUM procedure accepts all the candidate change point estimators obtained from the smallest bandwidth. In addition, MOSUM.TAVC outperforms some of its competitors, including DeCAFS and DepSMUCE, in terms of size control. Apart from this, regarding the choice of the tuning parameter ν employed in (22), the multiscale procedures corre-

sponding to subscript 2, obtained by applying the median given in (30) for TAVC estimation, return better size control. On the contrary, when $p \geq 1$, the utilization of the trimmed mean (represented by subscript 1) defined in Equation (29) exhibits improved power. This suggests adopting an approach that combines both choices of ν , thereby achieving a more balanced performance that takes into account the trade-off between power and size.

Table 2. Performance comparisons for the competing methods for (M1)–(M6) for $T = 1000$ over 1000 realizations.

Model	Method	$(\hat{p} - p)$					CM	RMSE	
		Size	≤ -2	-1	0	1			≥ 2
(M1)	MOSUM.TAVC[1]	0.104	0.000	0.000	1.000	0.000	0.000	0.997	3.007
	MOSUM.TAVC[2]	0.066	0.000	0.000	0.999	0.001	0.000	0.997	3.007
	WBS2.TAVC[1]	0.035	0.000	0.000	1.000	0.000	0.000	<u>0.999</u>	1.943
	WBS2.TAVC[2]	0.035	0.000	0.000	1.000	0.000	0.000	<u>0.999</u>	<u>1.894</u>
	DeCAFS	0.007	0.000	0.325	0.671	0.002	0.002	0.934	7.839
	DepSMUCE(0.05)	0.003	0.000	0.000	0.893	0.107	0.000	0.821	4.820
	DepSMUCE(0.1)	0.020	0.000	0.000	0.711	0.289	0.000	0.857	5.811
	WCM.gSa	0.006	0.000	0.000	0.984	0.015	0.001	0.998	1.941
(M2)	MOSUM.TAVC[1]	0.116	0.000	0.000	0.998	0.002	0.000	0.995	3.373
	MOSUM.TAVC[2]	0.070	0.000	0.000	1.000	0.000	0.000	0.995	3.380
	WBS2.TAVC[1]	0.049	0.000	0.000	1.000	0.000	0.000	0.997	<u>2.402</u>
	WBS2.TAVC[2]	0.024	0.000	0.000	1.000	0.000	0.000	0.997	<u>2.402</u>
	DeCAFS	0.010	0.000	0.087	0.092	0.150	0.671	0.847	21.360
	DepSMUCE(0.05)	0.340	0.000	0.008	0.673	0.108	0.211	0.914	11.538
	DepSMUCE(0.1)	0.476	0.000	0.010	0.640	0.129	0.221	0.943	12.202
	WCM.gSa	0.010	0.000	0.000	0.991	0.006	0.003	<u>0.998</u>	2.429
(M3)	MOSUM.TAVC[1]	0.161	0.000	0.052	0.802	0.129	0.110	0.940	2.930
	MOSUM.TAVC[2]	0.084	0.001	0.078	0.812	0.106	0.003	0.935	3.023
	WBS2.TAVC[1]	0.107	0.000	0.145	0.853	0.002	0.000	0.950	<u>2.509</u>
	WBS2.TAVC[2]	0.062	0.001	0.175	0.824	0.000	0.000	0.938	2.785
	DeCAFS	0.344	0.000	0.129	0.661	0.180	0.030	<u>0.956</u>	4.130
	DepSMUCE(0.05)	1.000	0.000	0.001	0.749	0.237	0.013	<u>0.926</u>	2.547
	DepSMUCE(0.1)	1.000	0.000	0.000	0.713	0.261	0.026	0.939	2.633
	WCM.gSa	0.044	0.006	0.068	0.793	0.065	0.068	0.943	2.608
(M4)	MOSUM.TAVC[1]	0.122	0.000	0.040	0.940	0.015	0.002	<u>0.974</u>	2.491
	MOSUM.TAVC[2]	0.015	0.000	0.056	0.923	0.019	0.002	0.970	2.374
	WBS2.TAVC[1]	0.055	0.000	0.079	0.921	0.000	0.000	0.973	2.126
	WBS2.TAVC[2]	0.037	0.000	0.170	0.830	0.000	0.000	0.956	2.742
	DeCAFS	0.370	0.000	0.325	0.671	0.002	0.002	0.883	6.776
	DepSMUCE(0.05)	1.000	0.000	0.004	0.821	0.068	0.107	0.950	2.941
	DepSMUCE(0.1)	1.000	0.000	0.001	0.778	0.043	0.178	0.964	6.776
	WCM.gSa	0.032	0.000	0.000	0.984	0.015	0.001	0.967	2.042
(M5)	MOSUM.TAVC[1]	0.027	0.000	0.000	1.000	0.000	0.000	<u>0.998</u>	37.926
	MOSUM.TAVC[2]	0.016	0.000	0.000	1.000	0.000	0.000	<u>0.998</u>	37.926
	WBS2.TAVC[1]	0.069	0.000	0.000	1.000	0.000	0.000	<u>0.998</u>	25.809
	WBS2.TAVC[2]	0.033	0.000	0.000	1.000	0.000	0.000	<u>0.998</u>	25.809
	DeCAFS	0.000	0.001	0.006	0.991	0.002	0.000	0.995	70.859
	DepSMUCE(0.05)	0.058	0.000	0.005	0.527	0.016	0.452	0.871	492.296
	DepSMUCE(0.1)	0.119	0.000	0.000	0.462	0.024	0.514	0.919	512.047
	WCM.gSa	0.000	0.000	0.000	1.000	0.000	0.000	<u>0.998</u>	<u>24.396</u>
(M6)	MOSUM.TAVC[1]	0.154	0.001	0.000	0.999	0.001	0.000	0.998	1.196
	MOSUM.TAVC[2]	0.107	0.000	0.000	0.995	0.004	0.001	0.998	1.198
	WBS2.TAVC[1]	0.091	0.000	0.000	0.997	0.003	0.000	<u>0.999</u>	<u>0.815</u>
	WBS2.TAVC[2]	0.042	0.000	0.000	0.999	0.001	0.000	<u>0.999</u>	<u>0.815</u>
	DeCAFS	0.774	0.000	0.055	0.215	0.071	0.659	0.898	16.904
	DepSMUCE(0.05)	0.417	0.000	0.010	0.701	0.193	0.096	0.880	7.476
	DepSMUCE(0.1)	0.498	0.000	0.012	0.681	0.182	0.125	0.906	8.363
	WCM.gSa	0.017	0.000	0.000	0.972	0.021	0.007	0.997	2.867

Moreover, when $p \geq 1$, WBS2.TAVC performs well in most scenarios across different evaluation criteria, such as model selection accuracy measured by $|\hat{p} - p|$, RMSE, or CM. In particular, WBS2.TAVC corresponding to the subscript 1 belonging to the trimmed mean seems to perform exceptionally well according to the evaluation criteria. DepSMUCE may encounter calibration problems. To inhibit the detection of spurious change points, it is necessary to use a conservative value for α . Conversely, for better detection power, a higher α value is preferable. In addition, the LRV estimator proposed in DepSMUCE tends to underestimate the LRV when it approaches zero, as observed in (M5), leading to a substantial number of falsely detected change points. DeCAFS works based on ε_t follows an AR(1) process. As a result, it is utilized under a model that is not correctly specified in certain scenarios, yet it continues to show satisfactory performance. However, it returns spurious change points and fails to adequately control size, even when applied to the correctly specified model (M3). Based on the evaluation criteria, in most scenarios, WCM.gSa generally demonstrates strong performance in maintaining size control and accurately determining the number of change points when $p \geq 1$.

Furthermore, Figure 1 shows histogram plots of the detected change point locations from different methods for model (M6) applied to 1000 realizations of data. The multiscale MOSUM.TAVC and WBS2.TAVC are highly precise and reliable, showing strong peaks at the true change points (200, 400, 600, and 800) with minimal spurious detections. DeCAFS and DepSMUCE(0.1) exhibit more spurious change points, indicating lower precision compared to the MOSUM.TAVC and WBS2.TAVC methods. DepSMUCE(0.05) shows better precision than DepSMUCE(0.1) but still has some spurious detections. WCM.gSa demonstrates high precision, similar to MOSUM.TAVC and WBS2.TAVC methods, making it a reliable choice as well. Overall, for applications where minimising false positives is crucial, MOSUM.TAVC and WBS2.TAVC are the best choices.

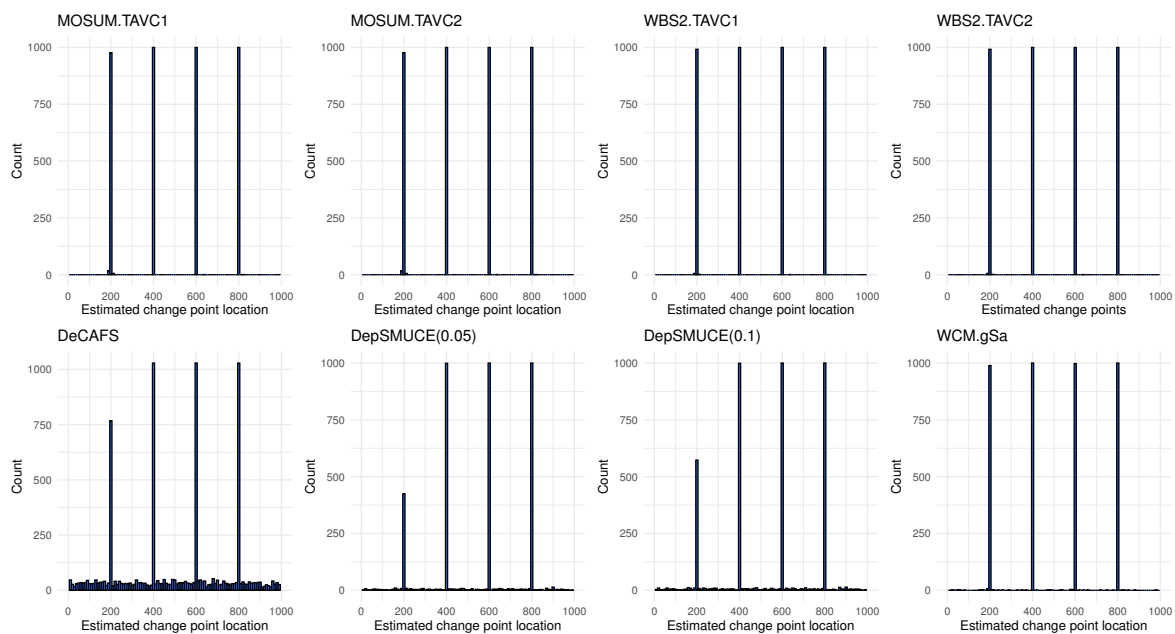


Figure 1. Histogram plots of estimated change points obtained from various methods under model (6), with true change points at 200, 400, 600, and 800.

5.3. Real Data Application: Fine Particulate Matter 2.5 (PM2.5)

Different studies revealed that fine particulate matter, particularly PM2.5, is the main air pollutant, among others. The main anthropogenic sources of PM2.5 and its acute and chronic health effects have been well-presented in [60,61]. In this study, we consider the Beijing multi-site air quality dataset collected from 1 March 2013 to 28 February 2017, particularly the daily average concentration of PM2.5. The dataset consists of six air pollutants, including PM2.5, and six meteorological variables collected from twelve monitoring sta-

tions in Beijing. It has 35,064 data points, which were recorded hourly at each station from which the daily measure is taken, and consists of randomly missing values. The dataset was retrieved from the University of California Irvine machine learning repository and is available from [62].

We have estimated the hourly missing records of PM2.5 by using the ARIMA state space model and Kalman smoothing. After estimating and imputing the hourly missing data, the daily average PM2.5 concentration data are the results of averaging the site according to the hourly data of the environmental protection station recorded that day (<https://www.aqistudy.cn>, accessed on 15 February 2024). The data show seasonality, which may be attributed to heating during the winter season as the primary contributor to PM2.5. To mitigate this effect, we apply a square root transformation to the data and eliminate seasonality. We analyze the transformed imputed time series from PM2.5 concentrations for change points in the level using the proposed MOSUM.TAVC and WBS2.TAVC. We select the parameter ν in (22) using (29) and choose the other tuning parameters as recommended in Section 5.1. The daily average PM2.5 concentration along with the detected change points by MOSUM.TAVC and WBS2.TAVC are plotted in Figure 2. For comparison, we also present the change points estimated by DeCAFS, DepSMUCE(0.05), and DepSMUCE(0.1) in Table 3. As WCM.gSa did not detect any change points, it has been excluded from the report.

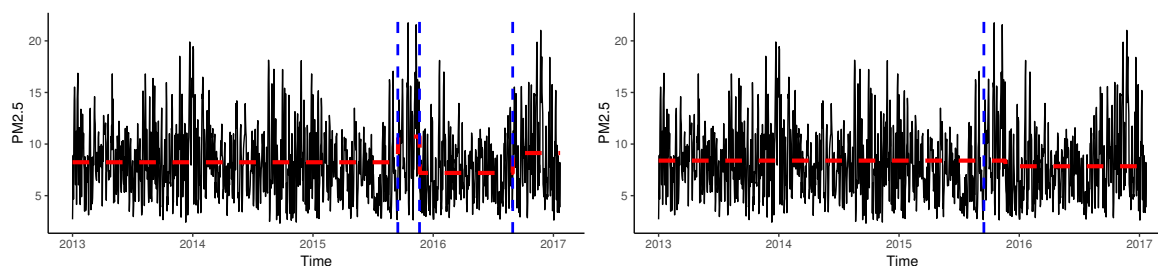


Figure 2. Daily average concentrations of PM2.5, plotted together with the change points detected by MOSUM.TAVC (left) and WBS2.TAVC (right) are denoted by blue vertical lines and estimated means are given by red horizontal lines.

Table 3. Change points detected from the daily average concentrations of PM2.5.

Method	Detected Change Point Location (Time)
MOSUM.TAVC	2015-10-31, 2016-01-03, 2016-10-09
WBS2.TAVC	2016-01-03
DeCAFS	2014-02-12, 2014-02-24, 2014-02-26, 2014-10-07, 2014-10-11, 2015-11-26, 2015-11-30, 2015-12-01, 2015-12-24, 2015-12-26, 2016-01-03, 2016-03-02, 2016-03-04, 2016-09-21, 2016-12-19, 2016-12-21, 2016-12-29, 2017-01-07, 2017-01-27, 2017-01-28
DepSMUCE(0.05)	2014-02-12, 2014-02-26, 2015-11-26, 2016-01-03, 2016-12-16, 2017-01-07
DepSMUCE(0.1)	2014-02-12, 2014-02-26, 2015-05-01, 2015-11-26, 2015-12-01, 2016-01-03, 2016-12-16, 2017-01-07

MOSUM.TAVC and WBS2.TAVC detect three and one change point, respectively, some of which can be linked to weather conditions and policy changes likely affecting the levels of air pollutants. Adverse meteorological conditions in November and December 2015 caused several heavy pollution events in North China, where daily average PM2.5 concentrations exceeded $150 \mu\text{g}/\text{m}^3$. At the same time, the Beijing–Tianjin–Hebei region experienced two major pollution episodes, each persisting for more than five consecutive days [61]. This change more likely corresponds to 31 October 2015, which was detected by MOSUM.TAVC. On the other hand, China implemented the Action Plan for Prevention and Control of Air Pollution, along with a reinforced version for the Beijing–Tianjin–Hebei Region (2016–2017). Measures included optimising the energy structure, switching from coal to electricity or gas, eliminating small coal-fired boilers, and implementing ultra-low emission technical

reforms. Furthermore, old vehicles were phased out, and new emission standards for light-duty vehicles and ship engines were established [63]. The implementation of the policy has significantly reduced the level of PM_{2.5}, which can be linked to the change on 3 January 2016 and it is detected by all methods.

6. Conclusions

In this paper, we proposed the ARIMA state space model and Kalman smoothing to impute missing values for a univariate time series. By utilizing fixed-point smoothing on the Kalman filter, we have shown that ARIMA state space and Kalman smoothing yield a minimum mean square estimate for the missing observation. It has also been shown that the proposed imputation method has better approximated the missing observation and provides good adaptivity to the mean shift in the series. For detecting change points in imputed time series, we have applied a robust scale-dependent TAVC estimator and showed its consistency under heavy tails and serial dependence circumstances. Moreover, the MOSUM and WBS2 procedures scaled with the proposed TAVC estimator have better estimated the number of change points and their corresponding locations. Specifically, WBS2 combined with the TAVC estimator has accurately detected the change in mean shift. Notably, the method also successfully identified change points in time series with seasonal variations, without the need for data transformation, highlighting its practical applicability for real-world datasets and indicating a promising avenue for future research. Furthermore, our approach has the potential to be applied to local stationarity, providing a promising direction for future research.

In addition, for detecting change points in a time series exhibiting periodic fluctuations such as seasonal and cyclic variation, investigating periodic models [64,65] to separate cyclical changes from genuine shifts. Furthermore, integrating Markov switching models provides a promising avenue for detecting regime shifts, where the system changes from one state to another [66]. Finally, threshold models could be explored to address structural breaks in time series when variables exceed certain critical levels, offering a deeper understanding of nonlinear dynamics and abrupt shifts in real-world phenomena [67]. By addressing these open problems, future research could extend the applicability and robustness of the proposed methods, particularly for complex time series with structural or periodic changes in diverse fields such as economics, finance, and environmental science.

Author Contributions: Conceptualization, T.T.H., F.T., G.A. and B.T.; Formal analysis, T.T.H., F.T., G.A. and B.T.; Supervision, B.T.; Validation, T.T.H., F.T., G.A. and B.T.; Visualization, T.T.H., F.T., G.A. and B.T.; Writing—original draft, T.T.H.; Writing—review and editing, T.T.H., F.T., G.A. and B.T.; Investigation, T.T.H., F.T., G.A. and B.T.; Methodology, T.T.H., F.T., G.A. and B.T.; Funding acquisition, G.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R45), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Data Availability Statement: The data applied in this paper are openly available at the University of California, Irvine Machine Learning Repository, accessible through <https://doi.org/10.24432/C5RK5G> (accessed on 20 July 2023).

Acknowledgments: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R45), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aminikhanghahi, S.; Cook, D.J. A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **2017**, *51*, 339–367. [[CrossRef](#)] [[PubMed](#)]
2. Truong, C.; Oudre, L.; Vayatis, N. Selective review of offline change point detection methods. *Signal Process.* **2020**, *167*, 107299. [[CrossRef](#)]
3. Cho, H.; Kirch, C. Data segmentation algorithms: Univariate mean change and beyond. *Econom. Stat.* **2024**, *30*, 76–95. [[CrossRef](#)]

4. Csorgo, M.; Horváth, L. *Limit Theorems in Change-Point Analysis*; Wiley: New York, NY, USA, 1997.
5. Hušková, M.; Slaby, A. Permutation tests for multiple changes. *Kybernetika* **2001**, *37*, 605–622.
6. Yao, Y.C. Estimating the number of change-points via Schwarz' criterion. *Stat. Probab. Lett.* **1988**, *6*, 181–189. [[CrossRef](#)]
7. Lavielle, M.; Moulines, E. Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal.* **2000**, *21*, 33–59. [[CrossRef](#)]
8. Chakar, S.; Lebarbier, E.; Lévy-Leduc, C.; Robin, S. A robust approach for estimating change-points in the mean of an AR(1) process. *Bernoulli* **2017**, *23*, 1408–1447. [[CrossRef](#)]
9. Romano, G.; Rigai, G.; Runge, V.; Fearnhead, P. Detecting abrupt changes in the presence of local fluctuations and autocorrelated noise. *J. Am. Stat. Assoc.* **2022**, *117*, 2147–2162. [[CrossRef](#)]
10. Lu, Q.; Lund, R.; Lee, T.C.M. An MDL approach to the climate segmentation problem. *Ann. Appl. Stat.* **2010**, *4*, 299–319. [[CrossRef](#)]
11. Gallagher, C.; Killick, R.; Lund, R.; Shi, X. Autocovariance Estimation in the Presence of Changepoints. *J. Korean Stat. Soc.* **2022**, *51*, 1021–1040. [[CrossRef](#)]
12. Tecuapetla-Gómez, I.; Munk, A. Autocovariance estimation in regression with a discontinuous signal and m-dependent errors: A difference-based approach. *Scand. J. Stat.* **2017**, *44*, 346–368. [[CrossRef](#)]
13. Eichinger, B.; Kirch, C. A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* **2018**, *24*, 526–564. [[CrossRef](#)]
14. Dette, H.; Eckle, T.; Vetter, M. Multiscale change point detection for dependent data. *Scand. J. Stat.* **2020**, *47*, 1243–1274. [[CrossRef](#)]
15. Chan, K.W. Optimal difference-based variance estimators in time series: A general framework. *Ann. Stat.* **2022**, *50*, 1376–1400. [[CrossRef](#)]
16. Robbins, M.; Gallagher, C.; Lund, R.; Aue, A. Mean shift testing in correlated data. *J. Time Ser. Anal.* **2011**, *32*, 498–511. [[CrossRef](#)]
17. den Haan, W.J.; Levin, A.T. 12 A practitioner's guide to robust covariance matrix estimation. In *Handbook of Statistics*; Elsevier: Amsterdam, The Netherlands, 1997; Volume 15, pp. 299–342.
18. Chan, K.W.; Yau, C.Y. High-order corrected Estimator of Asymptotic Variance with Optimal Bandwidth. *Scand. J. Stat.* **2017**, *44*, 866–898. [[CrossRef](#)]
19. Hušková, M.; Kirch, C. A note on studentized confidence intervals for the change-point. *Comput. Stat.* **2010**, *25*, 269–289. [[CrossRef](#)]
20. Schmitt, P.; Mandel, J.; Guedj, M. A comparison of six methods for missing data imputation. *J. Biom. Biostat.* **2015**, *6*, 1.
21. Somasundaram, R.; Nedunchezian, R. Evaluation of three simple imputation methods for enhancing preprocessing of data with missing values. *Int. J. Comput. Appl.* **2011**, *21*, 14–19. [[CrossRef](#)]
22. Xie, Y.; Huang, J.; Willett, R. Change-point detection for high-dimensional time series with missing data. *IEEE J. Sel. Top. Signal Process.* **2012**, *7*, 12–27. [[CrossRef](#)]
23. Follain, B.; Wang, T.; Samworth, R.J. High-dimensional changepoint estimation with heterogeneous missingness. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2021**, *84*, 1023–1055. [[CrossRef](#)]
24. Londschien, M.; Kovács, S.; Bühlmann, P. Change-point detection for graphical models in the presence of missing values. *J. Comput. Graph. Stat.* **2021**, *30*, 768–779. [[CrossRef](#)]
25. Li, S.; Hu, T.; Wang, L.; McMahan, C.S.; Tebb, J.M. Regression analysis of group-tested current status data. *Biometrika* **2024**, *111*, 1047–1061. [[CrossRef](#)]
26. Zhao, Y.; Landgrebe, E.; Shekhtman, E.; Udell, M. Online missing value imputation and change point detection with the gaussian copula. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Philadelphia, PA, USA, 27 February–2 March 2022; Volume 36, pp. 9199–9207.
27. McGonigle, E.T.; Cho, H. Robust multiscale estimation of time-average variance for time series segmentation. *Comput. Stat. Data Anal.* **2023**, *179*, 107648. [[CrossRef](#)]
28. Cho, H.; Fryzlewicz, P. Multiple change point detection under serial dependence: Wild contrast maximisation and gappy Schwarz algorithm. *J. Time Ser. Anal.* **2024**, *45*, 479–494. [[CrossRef](#)]
29. Ansley, C.F.; Kohn, R. On the estimation of ARIMA models with missing values. In *Time Series Analysis of Irregularly Observed Data: Proceedings of a Symposium Held at Texas A & M University, College Station, TX, USA, 10–13 February 1983*; Springer: Berlin/Heidelberg, Germany, 1984; pp. 9–37.
30. Harvey, A.C. *Forecasting, Structural Time Series Models and the Kalman Filter*; Cambridge University Press: New York, NY, USA, 1989.
31. Tsay, R.S. *Analysis of Financial Time Series*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2010.
32. Anderson, B.; Moore, J.B. *Optimal Filtering*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, USA, 1979.
33. Durbin, J.; Koopman, S.J. *Time Series Analysis by State Space Methods*, 2nd ed.; Oxford University Press: Oxford, UK, 2012.
34. Welch, G.; Bishop, G. *An Introduction to the Kalman Filter*; University of North Carolina: Chapel Hill, NC, USA, 1997.
35. Wijesekara, L.; Liyanage, L. Mind the Large Gap: Novel Algorithm Using Seasonal Decomposition and Elastic Net Regression to Impute Large Intervals of Missing Data in Air Quality Data. *Atmosphere* **2023**, *14*, 355. [[CrossRef](#)]
36. Wijesekara, W.M.L.K.N.; Liyanage, L. Comparison of Imputation Methods for Missing Values in Air Pollution Data: Case Study on Sydney Air Quality Index. In *Advances in Information and Communication*; Arai, K., Kapoor, S., Bhatia, R., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 257–269.
37. Moritz, S.; Bartz-Beielstein, T. imputeTS: Time series missing value imputation in R. *R J.* **2017**, *9*, 207. [[CrossRef](#)]

38. Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; Wiley-Interscience: Hoboken, NJ, USA, 1987.
39. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2015.
40. Harvey, A.C.; Pierse, R.G. Estimating missing observations in economic time series. *J. Am. Stat. Assoc.* **1984**, *79*, 125–131. [[CrossRef](#)]
41. Catoni, O. Challenging the empirical mean and empirical variance: A deviation study. *Ann. de l'IHP Probabilités et Stat.* **2012**, *48*, 1148–1185. [[CrossRef](#)]
42. Likai Chen, W.W.; Wu, W.B. Inference of Breakpoints in High-dimensional Time Series. *J. Am. Stat. Assoc.* **2022**, *117*, 1951–1963. [[CrossRef](#)]
43. Wong, K.C.; Li, Z.; Tewari, A. Lasso guarantees for β -mixing heavy-tailed time series. *Ann. Stat.* **2020**, *48*, 1124–1142. [[CrossRef](#)]
44. Chu, C.S.J.; Hornik, K.; Kaun, C.M. MOSUM tests for parameter constancy. *Biometrika* **1995**, *82*, 603–617. [[CrossRef](#)]
45. Meier, A.; Kirch, C.; Cho, H. mosum: A package for moving sums in change-point analysis. *J. Stat. Softw.* **2021**, *97*, 1–42. [[CrossRef](#)]
46. Cho, H.; Kirch, C. Two-stage data segmentation permitting multiscale change points, heavy tails and dependence. *Ann. Inst. Stat. Math.* **2022**, *74*, 653–684 [[CrossRef](#)]
47. Messer, M.; Kirchner, M.; Schiemann, J.; Roeper, J.; Neining, R.; Schneider, G. A multiple filter test for the detection of rate changes in renewal processes with varying variance. *Ann. Appl. Stat.* **2014**, *8*, 2027–2067. [[CrossRef](#)]
48. Niu, Y.S.; Zhang, H. The screening and ranking algorithm to detect DNA copy number variations. *Ann. Appl. Stat.* **2012**, *6*, 1306. [[CrossRef](#)]
49. Scott, A.J.; Knott, M. A cluster analysis method for grouping means in the analysis of variance. *Biometrics* **1974**, *30*, 507–512. [[CrossRef](#)]
50. Olshen, A.B.; Venkatraman, E.S.; Lucito, R.; Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **2004**, *5*, 557–572. [[CrossRef](#)]
51. Fryzlewicz, P. Wild binary segmentation for multiple change-point detection. *Ann. Stat.* **2014**, *42*, 2243–2281. [[CrossRef](#)]
52. Fryzlewicz, P. Detecting possibly frequent change-points: Wild Binary Segmentation 2 and steepest-drop model selection. *J. Korean Stat. Soc.* **2020**, *49*, 1027–1070. [[CrossRef](#)]
53. McGonigle, E.T.; Killick, R.; Nunes, M.A. Detecting changes in mean in the presence of time-varying autocovariance. *Stat* **2021**, *10*, e351. [[CrossRef](#)]
54. Rockel, T. *R Package*, version 0.4.0.; missMethods: Methods for Missing Data; R Foundation for Statistical Computing: Vienna, Austria, 2022.
55. Moritz, S.; Gatscha, S. *R Package*, version 3.3; imputeTS: Time Series Missing Value Imputation in R; R Foundation for Statistical Computing: Vienna, Austria, 2021.
56. Moritz, S.; Sardá, A.; Bartz-Beielstein, T.; Zaefferer, M.; Stork, J. Comparison of different methods for univariate time series imputation in R. *arXiv* **2015**, arXiv:1510.03924.
57. Frick, K.; Munk, A.; Sieling, H. Multiscale change point inference. *J. R. Stat. Society. Ser. B (Stat. Methodol.)* **2014**, *76*, 495–580. [[CrossRef](#)]
58. Van den Burg, G.J.; Williams, C.K. An evaluation of change point detection algorithms. *arXiv* **2020**, arXiv:2003.06222.
59. Arbeláez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour Detection and Hierarchical Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 898–916. [[CrossRef](#)]
60. Liang, X.; Zou, T.; Guo, B.; Li, S.; Zhang, H.; Zhang, S.; Huang, H.; Chen, S.X. Assessing Beijing's PM2.5 pollution: Severity, weather impact, APEC and winter heating. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2015**, *471*, 20150257.
61. Dai, Q.; Dai, T.; Hou, L.; Li, L.; Bi, X.; Zhang, Y.; Feng, Y. Quantifying the impacts of emissions and meteorology on the interannual variations of air pollutants in major Chinese cities from 2015 to 2021. *Sci. China Earth Sci.* **2023**, *66*, 1725–1737. [[CrossRef](#)]
62. Chen, S. Beijing Multi-Site Air-Quality Data. UCI Machine Learning Repository 2019. Available online: <https://archive.ics.uci.edu/dataset/501/beijing+multi+site+air+quality+data> (accessed on 20 July 2023).
63. Ministry of Environmental Protection. 2016 Report on the State of the Environment in China. 31 May 2017. Available online: <https://english.mee.gov.cn/Resources/Reports/soe/> (accessed on 13 May 2024).
64. Boubacar Maïnassara, Y.; Ilmi Amir, A. Portmanteau tests for periodic ARMA models with dependent errors. *J. Time Ser. Anal.* **2024**, *45*, 164–188. [[CrossRef](#)]
65. Ghezal, A.; Alzeley, O. Probabilistic properties and estimation methods for periodic threshold autoregressive stochastic volatility. *AIMS Math.* **2024**, *9*, 11805–11832. [[CrossRef](#)]
66. Cavicchioli, M. A matrix unified framework for deriving various impulse responses in Markov switching VAR: Evidence from oil and gas markets. *J. Econ. Asymmetries* **2024**, *29*, e00349. [[CrossRef](#)]
67. Tong, H. *Threshold Models in Non-Linear Time Series Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 21.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.