

---

Article

Not peer-reviewed version

---

# Reinforcing the Moving Linear Model Approach: Theoretical Assessment of Parameter Estimation and Outlier Detection

---

[Koki Kyo](#) \*

Posted Date: 12 May 2025

doi: 10.20944/preprints202505.0768.v1

Keywords: moving linear model approach; evaluation of model results; time series component decomposition; outlier detection; economic time series



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

## Article

# Reinforcing the Moving Linear Model Approach: Theoretical Assessment of Parameter Estimation and Outlier Detection

Koki Kyo

Digital Transformation Center, Gifu Shotoku Gakuen University, 1-1 Takakuwanishi, Yanaizu-cho, Gifu-shi, Gifu 501-6194, Japan; kyo@gifu.shotoku.ac.jp

**Abstract:** This paper presents methodological advancements to enhance the moving linear (ML) model approach for time series analysis, with a particular focus on improving model evaluation and outlier detection. The ML model decomposes time series data into constrained and remaining components, allowing for flexible and effective analysis of economic fluctuations. Building on this framework, the extended moving linear (EML) model introduces a mechanism for outlier detection by treating outliers as parameters estimated via maximum likelihood. However, challenges remain in providing theoretical support for parameter estimation and ensuring the stable identification of outlier locations. To address these issues, we first develop a theoretically grounded evaluation criterion that enables coherent comparison of model outcomes. Second, we propose a new outlier detection method based on maximizing the AIC reduction maximization method, offering a more systematic and effective approach. Empirical analyses using economic time series data demonstrate the improved performance and practical utility of the proposed enhancements.

**Keywords:** moving linear model approach; evaluation metrics; time series component decomposition; outlier detection; economic time series

---

## 1. Introduction

In [1], a moving linear (ML) model approach was proposed for decomposing time series data into two distinct components. Aimed specifically at business cycle analysis, this approach introduces a novel framework called the constrained-remaining components decomposition. In this framework, the time series is separated into two parts: the *constrained component*, which is governed by local linear model restrictions and primarily captures the underlying trend, and the *remaining component*, which is subject to minimal constraints and reflects cyclical movements, irregular fluctuations, and other disturbances in the data.

This separation under mild constraints represents a significant advancement in time series decomposition. While traditional methods typically focus on extracting basic patterns such as trends and cycles, the ML model approach provides a more flexible framework that can be applied across various fields. By employing simple model structures, the ML model achieves both broader applicability and effective handling of diverse data types and analytical objectives. Its ability to decompose time series into constrained and remaining components under flexible restrictions enhances its utility not only in economics but also in other disciplines dealing with temporal data.

A key parameter in the ML model approach is the width of time interval (WTI), which governs the degree of local linearity captured by the constrained component. In [1], WTI is estimated using the maximum likelihood method within a state-space modeling framework, allowing the model to adapt efficiently to local variations in the time series. Accurate estimation of WTI is essential for capturing both short-term fluctuations and long-term trends. In addition, [2] introduced an intuitive, albeit somewhat ad hoc, measure called overall stability to evaluate the performance of the ML model approach, though this measure currently lacks solid theoretical grounding.

Alongside developments in decomposition methods, anomaly detection has recently garnered increasing attention in various research fields, especially in time series analysis (see, [3]). In business cycle analysis in particular, identifying unexpected values or abrupt fluctuations is crucial. Numerous detection methods have been proposed (see e.g., [4]; also references therein), but many are highly specialized. For example, [5] evaluated three statistical outlier detection algorithms for water surface temperature using an unsupervised method. While valuable, these methods often lack general applicability. In the context of ML modeling, [6] proposed a two-step approach combining time series decomposition and anomaly detection. However, this procedure remains somewhat ad hoc and complex.

Building upon the ML model framework, [7] proposed an extension known as the extended moving linear (EML) model, which incorporates a robust method for outlier detection and estimation. In this extended framework, outliers are treated as parameters, with both their number and locations simultaneously estimated using the maximum likelihood method. Model selection is guided by the Akaike Information Criterion (AIC), enhancing the precision of outlier identification. To improve the search for outlier locations, [7] introduced a heuristic but practical procedure, addressing the challenge of accurately identifying irregular points in time series data. This extension significantly improves applicability to real-world economic data, where outliers – such as those from financial crises or global pandemics – can distort analytical results. Nevertheless, while the EML approach provides a well-structured detection method, it remains somewhat ad hoc, and its performance may be unstable due to the neglect of interrelationships among outliers.

The aim of the present chapter is twofold. First, we focus on reinforcing the ML model approach. While the use of maximum likelihood estimation for determining the WTI presents no particular difficulties, it is inherently model-dependent. Our goal is not merely to justify this approach, but to provide theoretical support for both the modeling framework and its estimation method through a logical and systematic discussion. We also propose a new evaluation criterion that offers a theoretically coherent and practically useful basis for assessing model performance. Second, we aim to refine the EML model approach for outlier detection by proposing a new method based on the maximization of AIC reduction. This approach identifies outliers based on their contribution to AIC improvement, offering a more systematic and effective detection framework.

Through these developments, we seek to enhance the robustness and practical utility of both the ML and EML model approaches. In particular, we aim to improve their performance in analyzing economic time series, where business cycles and outlier events pose significant analytical challenges. By introducing refined methods for model evaluation and anomaly detection, we aim to enable more reliable application of these models to real-world data. To support these methodological advancements, the state-space model-based decomposition approach proposed by [8] serves as a valuable reference point.

The remainder of this paper is organized as follows. Section 2 reviews the ML and EML model approaches. Section 3 introduces new developments aimed at reinforcing these approaches. Section 4 presents empirical examples to illustrate the analytical procedures and demonstrate the effectiveness of the proposed enhancements. Finally, Section 5 provides a summary and discussion.

## 2. A Review of the ML and EML Model Approaches

This section provides a comprehensive review of the Moving Linear (ML) model approach proposed by [1] and its extension, the Extended Moving Linear (EML) model approach introduced by [7]. These approaches serve as the theoretical foundation for the methodology developed in the present study.

### 2.1. The ML Model Approach

We begin by reviewing the ML model approach as presented in [1].

### 2.1.1. The Basic Model

We consider decomposing a time series  $y_t$  into two unobserved components as follows:

$$y_t = s_t + f_t \quad (t = 1, 2, \dots, N), \quad (1)$$

where  $s_t$  and  $f_t$  represent the constrained and remaining components, respectively. Here,  $t$  is the time index, and  $N$  is the sample size. The constrained component  $s_t$  reflects long-term variation, typically smooth, while the remaining component  $f_t$  captures short-term variations, such as cyclical fluctuations.

To ensure smoothness for the constrained component, we use a linear model of time  $t$ :

$$s_t = \alpha_n + \beta_n t, \quad (t = n, n+1, \dots, n+k-1). \quad (2)$$

Here,  $k$  is the width of the time interval (WTI), and  $\alpha_n$  and  $\beta_n$  are the coefficients in this interval. The  $n$ -th time interval is defined as  $[n, n+k-1]$  for  $n = 1, 2, \dots, N-k+1$ . A centered time variable within the  $n$ -th interval is introduced as follows:

$$z_j = t - \bar{t}_n = j - \frac{k+1}{2} \quad (j = 1, 2, \dots, k),$$

where  $\bar{t}_n = n + \frac{k-1}{2}$  is the midpoint of the time interval.

Substituting  $t = z_j + \bar{t}_n$  into the model in Eq. (2) gives:

$$s_t = \mu_n + \beta_n z_j \quad (t = n+j-1; j = 1, 2, \dots, k),$$

where  $\mu_n = \alpha_n + \beta_n \bar{t}_n$ . The model in Eq. (1) then becomes:

$$y_t = \mu_n + \beta_n z_j + f_t \quad (t = n+j-1; j = 1, 2, \dots, k), \quad (3)$$

which is referred to as the *moving linear model* (ML model).

### 2.1.2. State Space Presentation of the ML Model

In the ML model in Eq. (3), the quantities  $\mu_n$ ,  $\beta_n$ , and  $f(n)$  are treated as parameters. To achieve stable parameter estimates, we introduce the following priors for these parameters.

Firstly, it is assumed that  $\mu_n$  and  $\beta_n$  vary smoothly with respect to  $n$  as follows:

$$\mu_n = \mu_{n-1} + \varepsilon_n, \quad \beta_n = \beta_{n-1} + \phi_n \quad (4)$$

under diffuse priors for  $\varepsilon_n$  and  $\phi_n$ . Specifically, it is assumed that  $\varepsilon_n \sim \mathcal{N}(0, \eta^2)$  and  $\phi_n \sim \mathcal{N}(0, \eta^2)$ , where the variance  $\eta^2$  is set to a sufficiently large value.

Moreover, to uniquely identify the remaining component  $f_t$  locally, the following prior distribution for  $f_1^{(n)}$  and  $f_2^{(n)}$  is introduced to probabilistically maintain the orthogonality relationships between  $f_t$  and  $\mu_n$ , as well as between  $f_t$  and  $z_j$  in the  $n$ -th time interval:

$$f_1^{(n)} = - \sum_{j=1}^{k-1} f_j^{(n-1)} + e_{n+k-1}, \quad f_2^{(n)} = - \sum_{j=2}^k \frac{z_{k-j}}{z_k} f_j^{(n-1)} + e_{n+k-2}. \quad (5)$$

To express the transition relationships, the priors for the other remaining component terms are given as:

$$f_j^{(n)} = f_{j-1}^{(n-1)} + e_{n+k-j} \quad (j = 3, 4, \dots, k). \quad (6)$$

In these equations,  $e_n, e_{n-1}, \dots, e_{n-k+1}$  represent a set of system noises, where it is assumed that  $e_i \sim \mathcal{N}(0, \sigma^2)$  i.i.d., with  $n = 1, 2, \dots, N-k+1$ , and  $\sigma^2$  is an unknown parameter.

Bayesian form of the ML model was constructed based on the above settings.

Then, under the settings

$$\mathbf{x}_n = \begin{bmatrix} \mu_n \\ \beta_n \\ f_1^{(n)} \\ f_2^{(n)} \\ \vdots \\ f_k^{(n)} \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & -1 & -1 & \cdots & \cdots & -1 & 0 \\ 0 & 0 & 0 & -\frac{z_{k-1}}{z_k} & -\frac{z_{k-2}}{z_k} & \cdots & \cdots & -\frac{z_1}{z_k} \\ 0 & 0 & 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \quad (7)$$

$$\mathbf{v}_n = \begin{bmatrix} \varepsilon_n \\ \phi_n \\ e_{n+k-1} \\ e_{n+k-2} \\ \vdots \\ e_n \end{bmatrix}, \quad \mathbf{y}(n) = \begin{bmatrix} y_{n+k-1} \\ y_{n+k-2} \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{G} = \mathbf{I}_{k+2}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{1}_k & \mathbf{z} & \mathbf{I}_k \end{bmatrix}, \quad (8)$$

a state-space representation for the moving linear model can be defined as

$$\mathbf{x}_n = \mathbf{F}\mathbf{x}_{n-1} + \mathbf{G}\mathbf{v}_n, \quad (9)$$

$$\mathbf{y}(n) = \mathbf{H}\mathbf{x}_n \quad (10)$$

with  $\mathbf{v}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ , where  $\mathbf{Q} = \text{diag}(\eta^2, \eta^2, \sigma^2, \dots, \sigma^2)$ .

Therefore, given the initial conditions  $\mathbf{x}_{0|0}$  and  $\mathbf{V}_{0|0}$ , and the observations  $Y_{1:N} = \{y_1, y_2, \dots, y_N\}$ , we can compute the mean vectors and covariance matrices of the state variable  $\mathbf{x}_n$  for  $n = 1, 2, \dots, N - k + 1$  using the Kalman filter and fixed-interval smoothing algorithms, as described below (for details, see [1,9]):

#### [Kalman filter]

$$\begin{aligned} \mathbf{x}_{n|n-1} &= \mathbf{F}\mathbf{x}_{n-1|n-1}, \\ \mathbf{V}_{n|n-1} &= \mathbf{F}\mathbf{V}_{n-1|n-1}\mathbf{F}^T + \mathbf{G}\mathbf{Q}\mathbf{G}^T, \\ \mathbf{K}_n &= \mathbf{V}_{n|n-1}\mathbf{H}^T(\mathbf{H}\mathbf{V}_{n|n-1}\mathbf{H}^T)^{-1}, \\ \mathbf{x}_{n|n} &= \mathbf{x}_{n|n-1} + \mathbf{K}_n(\mathbf{y}(n) - \mathbf{H}\mathbf{x}_{n|n-1}), \\ \mathbf{V}_{n|n} &= (\mathbf{I}_{k+2} - \mathbf{K}_n\mathbf{H})\mathbf{V}_{n|n-1}. \end{aligned}$$

#### [Fixed-interval smoothing]

$$\begin{aligned} \mathbf{A}_n &= \mathbf{V}_{n|n}\mathbf{F}^T\mathbf{V}_{n+1|n}^{-1}, \\ \mathbf{x}_{n|N} &= \mathbf{x}_{n|n} + \mathbf{A}_n(\mathbf{x}_{n+1|N} - \mathbf{x}_{n+1|n}), \\ \mathbf{V}_{n|N} &= \mathbf{V}_{n|n} + \mathbf{A}_n(\mathbf{V}_{n+1|N} - \mathbf{V}_{n+1|n})\mathbf{A}_n^T. \end{aligned}$$

Using the above procedure, component decomposition can be performed. In particular, by incorporating  $f_t$  for  $t = 1, 2, \dots, N$  into the state vector  $\mathbf{x}_n$  (for  $n = 1, 2, \dots, N - k + 1$ ), the remaining component can be estimated within this model framework. The posterior distribution of the state vector  $\mathbf{x}_n$  obtained from fixed-interval smoothing is Gaussian, with mean  $\mathbf{x}_{n|N}$  and covariance matrix  $\mathbf{V}_{n|N}$ . Accordingly, the estimate  $\hat{f}_t$  of  $f_t$  can be extracted from  $\mathbf{x}_{n|N}$  for  $n = 1, 2, \dots, N - k + 1$ , and the estimate  $\hat{s}_t$  of the constrained component  $s_t$  is obtained as  $\hat{s}_t = y_t - \hat{f}_t$ .

### 2.1.3. Method for Estimating the Parameters

In the setup described above, the WTI,  $k$ , and the variance of the system noise  $\sigma^2$  are the key parameters that characterize the ML model.

According to [8], based on the Kalman filter, the density function of the predictive distribution of  $\mathbf{y}(n)$  is expressed in the form of a normal distribution:

$$f(\mathbf{y}(n)|\sigma^2, k) = \frac{1}{\sqrt{(2\pi)^k |\mathbf{U}_{n|n-1}|}} \exp\left(-\frac{S^2(n)}{2}\right)$$

with

$$S^2(n) = (\mathbf{y}(n) - \mathbf{y}_{n|n-1})^t \mathbf{U}_{n|n-1}^{-1} (\mathbf{y}(n) - \mathbf{y}_{n|n-1}),$$

where  $\mathbf{y}_{n|n-1}$  and  $\mathbf{U}_{n|n-1}$  are the mean and covariance matrix of the predictive distribution of  $\mathbf{y}(n)$ , given by:

$$\mathbf{y}_{n|n-1} = \mathbf{H}\mathbf{x}_{n|n-1}, \quad \mathbf{U}_{n|n-1} = \mathbf{H}\mathbf{V}_{n|n-1}\mathbf{H}^T$$

Here,  $\mathbf{x}_{n|n-1}$  and  $\mathbf{V}_{n|n-1}$  are the mean and covariance matrix of the state  $\mathbf{x}_n$  in the prediction step of the Kalman filter.

To define the likelihood as the product of conditional probability distributions, it is necessary to structure the vectors and matrices related to the observation vector  $\mathbf{y}_n$  as follows. Specifically,  $\mathbf{y}(n)$ ,  $\mathbf{y}_{n|n-1}$ ,  $\mathbf{V}_{n|n-1}$ , and  $\mathbf{H}$  are expressed as:

$$\mathbf{y}(n) = \begin{bmatrix} \mathbf{y}_{n+k-1} \\ \mathbf{y}^{(1)}(n) \end{bmatrix}, \quad \mathbf{y}_{n|n-1} = \begin{bmatrix} \hat{\mathbf{y}}_{n+k-1} \\ \mathbf{y}_{n|n-1}^{(1)} \end{bmatrix}, \quad \mathbf{V}_{n|n-1} = \begin{bmatrix} a & \mathbf{b}^t \\ \mathbf{b} & \mathbf{V}_{n|n-1}^{(1)} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \end{bmatrix}$$

with  $a$  being a scalar and  $\mathbf{b}$  a vector of appropriate dimension.

In this case, the density function of  $\mathbf{y}^{(1)}(n)$  is expressed as:

$$f_1(\mathbf{y}^{(1)}(n)|\sigma^2, k) = \frac{1}{\sqrt{(2\pi)^{k-1} |\mathbf{U}_{n|n-1}^{(1)}|}} \exp\left(-\frac{S_1^2(n)}{2}\right)$$

with

$$S_1^2(n) = (\mathbf{y}^{(1)}(n) - \mathbf{y}_{n|n-1}^{(1)})^t (\mathbf{U}_{n|n-1}^{(1)})^{-1} (\mathbf{y}^{(1)}(n) - \mathbf{y}_{n|n-1}^{(1)}),$$

where

$$\mathbf{U}_{n|n-1}^{(1)} = \mathbf{H}_1 \mathbf{V}_{n|n-1}^{(1)} \mathbf{H}_1^T.$$

Thus, for  $n = 1$ , the joint density function of  $Y_{1:k} = \mathbf{y}(1)$  is given by  $f(\mathbf{y}(1)|\sigma^2, k)$ . For  $n = 2, 3, \dots, N - k + 1$ , the conditional density function of  $\mathbf{y}_{n+k-1}$  given  $\mathbf{y}^{(1)}(n)$  is expressed as:

$$f_2(\mathbf{y}_{n+k-1}|\mathbf{y}^{(1)}(n), \sigma^2, k) = \frac{f(\mathbf{y}(n)|\sigma^2, k)}{f_1(\mathbf{y}^{(1)}(n)|\sigma^2, k)} \quad (n = 2, 3, \dots, N - k + 1)$$

Moreover, the joint density function of  $Y_{1:N}$  is expressed as:

$$L(\sigma^2, k|Y_{1:N}) = f(\mathbf{y}(1)|\sigma^2, k) \prod_{n=2}^{N-k+1} f_2(\mathbf{y}_{n+k-1}|\mathbf{y}^{(1)}(n), \sigma^2, k)$$

When the observed values of  $Y_{1:N}$  are given, the function  $L(\sigma^2, k | Y_{1:N})$  becomes the likelihood function for  $\sigma^2$  and  $k$ . Thus, the log-likelihood function is given by:

$$\begin{aligned} \text{LL}(\sigma^2, k | Y_{1:N}) &= \log(L(\sigma^2, k | Y_{1:N})) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \left( \log |\mathbf{U}_{1|0}| + S^2(1) \right) \\ &\quad - \frac{1}{2} \sum_{n=2}^{N-k+1} \left( \log |\mathbf{U}_{n|n-1}| - \log |\mathbf{U}_{n|n-1}^{(1)}| \right) - \frac{1}{2} \sum_{n=2}^{N-k+1} \left( S^2(n) - S_1^2(n) \right) \end{aligned} \quad (11)$$

Therefore, when  $k$  is given, the log-likelihood function  $\text{LL}(\sigma^2, k | Y_{1:N})$  defined in Eq. (11) can be maximized to obtain the estimate  $\hat{\sigma}^2$  of  $\sigma^2$ . However, this optimization problem is difficult to solve analytically and generally requires numerical optimization methods.

Further, if  $\sigma^2$  is temporarily set to 1 in the Kalman filter, the maximum likelihood estimate of  $\sigma^2$  can be obtained analytically as:

$$\hat{\sigma}^2 = \frac{1}{N} \left\{ S^2(1) + \sum_{n=2}^{N-k+1} (S^2(n) - S_1^2(n)) \right\}.$$

Finally, the conditional log-likelihood function with respect to  $k$ , denoted as  $\text{LL}(\hat{\sigma}^2, k | Y_{1:N})$ , is a function of  $k$  alone. Therefore, the estimate of  $k$  can be obtained by maximizing  $\text{LL}(\hat{\sigma}^2, k | Y_{1:N})$ .

## 2.2. The EML Model Approach for Outlier Detection

Below, we provide another review of the EML model approach, based on [7].

### 2.2.1. The Basic Model

Within the framework of the ML model approach, the following model is introduced to handle cases in which the time series  $y_t$  contains outliers:

$$y_t = s_t + f_t + u_t \quad (t = 1, 2, \dots, N), \quad (12)$$

where  $u_t$  represents the unusually varying component of the time series, which may include outliers. The other variables and associated assumptions in Eq. (12) are consistent with those in Eq. (1).

As a basic assumption, we suppose that the full dataset  $Y_{1:N} = \{y_1, y_2, \dots, y_N\}$  contains  $m$  outliers, denoted by  $\{\delta_1, \delta_2, \dots, \delta_m\}$ , with  $m \ll N$ . If an outlier is present at time  $t$ , the corresponding unusually varying component  $u_t$  is assigned one of the values from  $\{\delta_1, \delta_2, \dots, \delta_m\}$ . If no outlier is present at time  $t$ ,  $u_t$  is set to zero. In other words, within the set  $\{u_1, u_2, \dots, u_N\}$ , exactly  $m$  elements correspond one-to-one to the elements in the outlier set  $\{\delta_1, \delta_2, \dots, \delta_m\}$ , while the remaining elements are zero.

Therefore, a simple setting for handling outliers can be employed (see [7]). Specifically, let  $\delta = (\delta_1, \delta_2, \dots, \delta_m)^T$  denote the vector of outliers. We consider  $\{u_1, u_2, \dots, u_N\}$  as a set of functions of the outliers  $\delta$ , where each  $\delta_i$  is treated as an unknown constant. By defining  $y_t^* = y_t - u_t$ , the transformed dataset  $Y_{1:N}^*(\delta) = \{y_1^*, y_2^*, \dots, y_N^*\}$  becomes a version of the data with outliers removed. Applying  $Y_{1:N}^*(\delta)$  in place of  $Y_{1:N}$  to the ML model approach, the log-likelihood  $\text{LL}(\hat{\sigma}^2, k | Y_{1:N}^*(\delta))$  defined in Eq. (11) becomes a function of  $\delta$ , enabling estimation via the maximum likelihood method.

While this setting is simple and intuitive, when  $m$  is large, it leads to high computational costs and may compromise the reliability of the estimation results.

### 2.2.2. Bayesian Approach to Outlier Estimation

To address the presence of multiple outliers in a more general framework, a Bayesian approach is introduced as follows.

In this framework, outliers are formally treated as time-varying random variables. Specifically, each outlier  $\delta_j$  is represented as  $\delta_{jn}$ , and a transition equation is defined as

$$\delta_{jn} = \delta_{j(n-1)} \quad (j = 1, 2, \dots, m; n = 1, 2, \dots, N),$$

which assumes that each outlier remains constant over time at step  $n$ .

By incorporating Eq.(4), (5), and (6), a Bayesian model is constructed for the EML model, with Eq.(7) and (8) are redefined as follows:

$$\mathbf{x}_n = \begin{bmatrix} \mu_n \\ \beta_n \\ f_1^{(n)} \\ f_2^{(n)} \\ \vdots \\ f_k^{(n)} \\ \delta_{1n} \\ \delta_{2n} \\ \vdots \\ \delta_{mn} \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & -1 & -1 & \cdots & \cdots & -1 & 0 \\ 0 & 0 & 0 & -\frac{z_{k-1}}{z_k} & -\frac{z_{k-2}}{z_k} & \cdots & \cdots & -\frac{z_1}{z_k} \\ 0 & 0 & 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \mathbf{O},$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_{k+2} \\ \mathbf{O} \end{bmatrix}, \quad \mathbf{H}_n = \begin{bmatrix} \psi_{n+k-1} & & & \\ & \psi_{n+k-2} & & \\ & & \ddots & \\ & & & \ddots \\ & & & & \psi_n \end{bmatrix},$$

where  $\mathbf{I}_m$  denotes an  $m$ -dimensional identity matrix,  $\mathbf{O}$  denotes a zero matrix with appropriate dimensions, and  $\psi_i$  is a concatenation function defined such that  $\psi_i = 1$  if  $u_i = \delta_j$  for some  $j = 1, 2, \dots, m$ , and  $\psi_i = 0$  otherwise, for  $i = n+k-1, n+k-2, \dots, n$ .

Then, under the setting  $\mathbf{H} = \mathbf{H}_n$ , the state space presentation for the Bayesian EML model can be expressed using Eqs. (9) and (10).

Based on the above formulation and by applying  $\mathbf{H} = \mathbf{H}_n$  in both the Kalman filter and fixed-interval smoothing algorithms, estimates of the state vector can be obtained. Consequently, the estimates  $\hat{\delta}$  for the outliers, as well as the estimates of the remaining components, can be extracted from  $\mathbf{x}_{n|N}$  for  $n = 1, 2, \dots, N-k+1$ . Specifically,  $\hat{\delta}$  corresponds to the elements from the  $(k+3)$ -th to the  $(k+2+m)$ -th entries in the vector  $\mathbf{x}_{N-k+1|N}$ , which depend only on the smoothing result at  $n = N-k+1$ . The corresponding variances of the outlier estimates are given by the diagonal elements of the covariance matrix  $\mathbf{V}_{N-k+1|N}$  from the  $(k+3)$ -th to the  $(k+2+m)$ -th rows.

### 2.2.3. Outlier Detection and Estimation

The key steps for identifying the positions of outliers and estimating their values are as follows. First, ignoring the presence of outliers, the time series is decomposed into constrained and remaining components using the ML model approach. The initial outlier positions are then set as the time points corresponding to the top  $m$  values of the remaining component in terms of their squared magnitudes.

Next, the EML model provides Bayesian-type estimates of the outliers, which approximate maximum likelihood estimates. These estimates enable the computation of an approximate maximum log-likelihood, which in turn allows the number of outliers to be determined using the minimum AIC criterion.

Subsequently, the estimated values of the outliers are normalized by their standard deviations. By ranking the normalized squared estimates, a new set of outlier positions is obtained. This update distinguishes the revised positions from the initial setting and enables more accurate estimation based on the updated locations. The procedure for updating outlier positions is repeated iteratively until the estimation results converge.

Although the final step is well-structured, it remains somewhat ad hoc, and its performance in detecting outliers is unstable due to the lack of consideration for interrelationships among outliers. Revising this step is a central objective of the present study.

### 3. New Development to Reinforce Previous Findings

#### 3.1. The Aims

In the current ML model approach, the WTI is estimated using the maximum likelihood method. From the perspective of statistical analysis, the use of maximum likelihood itself poses no particular issues. However, as maximum likelihood estimation is inherently model-dependent, our aim is not merely to justify its application but to provide theoretical support for both the model and the estimation method through logical and systematic discussion. In particular, we seek to develop a theoretical understanding of the critical role that WTI plays in improving the model's overall performance.

Although WTI was originally introduced as a practical device, we endeavor to clarify, from a theoretical standpoint, how it facilitates the effective separation of constrained and remaining components and contributes to the stabilization of subsequent estimation procedures. This theoretical elaboration is essential for reinforcing the model's interpretability and robustness.

In addition, when comparing different estimation methods, likelihood values are not always a reliable basis due to fundamental differences in model structure. To address this issue, another key objective is to propose new evaluation metrics based on the variance-covariance structures of the decomposed components. These metrics are intended to enable a more holistic assessment of the decomposition results and thereby enhance the credibility of the ML model approach. Unlike existing methods that rely narrowly on likelihood, the proposed metrics capture a wider range of estimation characteristics and offer a theoretically sound and practically useful basis for model comparison and evaluation.

Furthermore, as noted in Section 2.2.3, while the existing EML approach for identifying outlier positions is well-designed, it remains somewhat ad hoc and can exhibit unstable performance due to its neglect of the interrelationships among outliers. To address these limitations, we propose a method based on the maximization of AIC reduction, which identifies outliers by evaluating their individual contributions to the decrease in the AIC. This approach selects outlier positions that most significantly enhance model fit. By introducing this theoretically grounded and computationally efficient method, the outlier detection process becomes more stable, systematic, and reliable.

Collectively, these enhancements-clarifying the theoretical role of the ML approach, proposing additional evaluation metrics, and refining the outlier detection process-represent significant steps toward completing and strengthening the ML and EML model approaches as a coherent and reliable framework for time series analysis.

#### 3.2. Reinforcing the ML Model Approach

In this section, we aim to reinforce the ML model approach by offering theoretical insights into the evaluation of results and by proposing additional evaluation metrics.

##### 3.2.1. Variance-Preserving Adjustment of the Decomposed Components

Within the context of the ML model approach, the constrained-remaining decomposition yields  $\hat{s}_t$  and  $\hat{f}_t$  as the decomposed constrained and remaining components, respectively, such that the original time series  $y_t$  can be represented as:

$$y_t = \hat{s}_t + \hat{f}_t \quad (t = 1, 2, \dots, N). \quad (13)$$

Obviously, the decomposition in Eq. (13) is average-invariant, ensuring that the equality  $\bar{y} = \bar{s} + \bar{f}$  holds, where

$$\bar{y} = \frac{1}{N} \sum_{t=1}^N y_t$$

denotes the average of the time series  $y_t$ , and  $\bar{s}$  and  $\bar{f}$  denote the averages of the constrained and remaining components, respectively. Note that in the ML model approach, the remaining component is adjusted so that  $\bar{f}$  becomes zero. As a result,  $\bar{s}$  is equal to the average  $\bar{y}$ . This implies that the average level of the constrained component matches that of the original time series.

In addition to the average-invariance of the decomposition in Eq. (13), variance-invariance is also desirable; however, it cannot be consistently maintained by the ML model approach. Let  $\text{Var}(y_t)$  denote the sample variance of  $y_t$ , while  $\text{Var}(s_t)$  and  $\text{Var}(f_t)$  denote the sample variances of  $\hat{s}_t$  and  $\hat{f}_t$ , respectively, and let  $\text{Cov}(s_t, f_t)$  denote their sample covariance. Then, the following relationship holds:

$$\text{Var}(y_t) = \text{Var}(s_t) + \text{Var}(f_t) + 2\text{Cov}(s_t, f_t).$$

If  $\text{Cov}(s_t, f_t) = 0$ , indicating no correlation between the decomposed components, then we have

$$\text{Var}(y_t) = \text{Var}(s_t) + \text{Var}(f_t). \quad (14)$$

Thus, the decomposition becomes variance-invariant, meaning that the total variance of the original time series equals the sum of the variances of the decomposed components without any covariance term.

In the ML model approach, the decomposed components are generally uncorrelated due to the model specification. However, this property may vary depending on the parameter settings and the characteristics of the data. When a correlation exists between the components, an adjustment can be made to restore variance-invariance by introducing the following regression model:

$$\hat{f}_t = \gamma(\hat{s}_t - \bar{y}) + \varepsilon_t \quad (t = 1, 2, \dots, N), \quad (15)$$

where  $\hat{s}_t - \bar{y}$  is the centered constrained component,  $\gamma$  is the regression coefficient, and  $\varepsilon_t$  is the residual term.

Based on the least squares estimate  $\hat{\gamma}$  of  $\gamma$ , the remaining component can be adjusted as follows:

$$\tilde{f}_t = \hat{s}_t - \hat{\gamma}(\hat{s}_t - \bar{y}), \quad (16)$$

where  $\tilde{f}_t$  represents the adjusted remaining component, which can be used as the final decomposed result.

By the orthogonality property of least squares estimation,  $\tilde{f}_t$  is uncorrelated with  $\hat{s}_t$ , thereby ensuring that variance-invariance is preserved when  $\tilde{f}_t$  is used as the final decomposed result in place of  $\hat{f}_t$ . This adjustment, which renders the decomposed components uncorrelated, is referred to as a *variance-preserving adjustment*.

Incidentally, as shown in Eqs. (15) and (16), the variance-preserving adjustment effectively eliminates the portion of the remaining component that is correlated with the constrained component, while maintaining the smoothness of the latter. Although an alternative method of achieving variance preservation – namely, adjusting the constrained component based on the remaining component – is theoretically possible, it is not recommended because it may compromise the smoothness of the trend.

Through the variance-preserving adjustment, the ML model approach can maintain the invariance properties of the constrained-remaining decomposition, namely, both average-invariance and variance-invariance. In statistical data, information is primarily conveyed through variance; thus, ensuring variance-invariance guarantees the preservation of information content before and after decomposition.

### 3.2.2. Structural Examination of Variances for the Decomposed Components

As discussed in the previous section, the variance-preserving adjustment ensures the maintenance of the variance-invariance property in the constrained-remaining decomposition, as shown in Eq. (14).

As a result, for a given time series  $y_t$ , the total variance of the two decomposed components remains constant, as expressed by the following equation:

$$\text{Var}(s_t) + \text{Var}(f_t) = \text{Var}(y_t) = \text{const.} \quad (17)$$

To further investigate the relationship between the decomposed components, define the product of their variances as

$$PV = \text{Var}(s_t)\text{Var}(f_t).$$

It can be confirmed that PV reaches its maximum when the variances of the decomposed components are equal, that is, when  $\text{Var}(s_t) = \text{Var}(f_t)$ . In this case, by also referring to equation (17), it becomes clear that  $\text{Var}(s_t) = \text{Var}(f_t) = 0.5\text{Var}(y_t)$ , and the maximum value of PV is given by  $0.25(\text{Var}(y))^2$ .

On the other hand, when the WTI,  $k$ , is small, the constrained component  $\hat{s}_t$  absorbs a large part of the fluctuations of the time series, exhibiting large variability. As a result, the remaining component  $\hat{f}_t$  contains relatively small fluctuations, leading to

$$\text{Var}(s_t) > \text{Var}(f_t).$$

As  $k$  increases,  $\text{Var}(s_t)$  decreases and  $\text{Var}(\hat{f}_t)$  increases, so that  $\text{Var}(s_t) - \text{Var}(f_t)$  gradually becomes smaller. Thus, the product of variances, PV, behaves as a monotonically non-decreasing function with respect to  $k$ .

When  $k$  continues to increase,  $\text{Var}(s_t)$  and  $\text{Var}(f_t)$  eventually become equal, so that PV reaches its maximum value. Beyond this point, as  $k$  increases further, PV becomes a monotonically non-increasing function of  $k$ .

When attempting to extract information related to business cycle fluctuations from time series data, such information is often captured by the variations of the remaining component. Therefore, it is natural to expect that the variance of the remaining component will be large in such contexts. Based on this expectation, one might consider choosing a larger value of the parameter  $k$  – particularly when starting from a small  $k$  – in order to achieve a higher value of PV and thereby better capture cyclical variations.

However, such an approach cannot be adopted indiscriminately. The structure of the model imposes constraints that must be respected. Simply increasing  $k$  in an attempt to maximize PV without regard to these structural constraints may lead to decompositions that are inconsistent with the fundamental assumptions of the model.

In summary, the absolute value of PV itself cannot be directly used as a criterion for evaluating the quality of a decomposition result. Nevertheless, PV serves as an important indicator of the relationship between the characteristics of the time series data and the parameter  $k$ . Therefore, for a given dataset, rather than focusing on the absolute magnitude of PV, it is more meaningful to observe how PV varies with  $k$ .

### 3.2.3. Assessing the Stability of Structural Changes in Decomposed Components

While the previously introduced indicator PV captures the relative changes in the variances of the two components as the parameter  $k$  varies, it does not reflect the internal stability or continuity of each component across successive values of  $k$ . In other words, PV illustrates how the variances of the constrained and remaining components,  $\text{Var}(s_t)$  and  $\text{Var}(f_t)$ , change in relation to each other, but it does not indicate whether the structural changes within each decomposed component occur gradually or abruptly as  $k$  increases.

To address this limitation, we consider an alternative approach that evaluates the similarity of each component's decomposed results with respect to a unit change in the successive value of  $k$ . This similarity can be quantified using the covariance between the decomposed results at the two settings. Specifically, for the constrained component, let  $\hat{s}_t(k-1)$  and  $\hat{s}_t(k)$  denote the decomposed results obtained at  $k-1$  and  $k$ , respectively. Instead of using the variance  $\text{Var}(s_t)$ , we introduce the covariance

$\text{Cov}(s_t(k-1), s_t(k))$  between  $\hat{s}_t(k-1)$  and  $\hat{s}_t(k)$  as an alternative indicator. If the decomposed results change only slightly with respect to  $k$ , the covariance  $\text{Cov}(s_t(k-1), s_t(k))$  will be close to  $\text{Var}(s_t)$ , suggesting that the component  $\hat{s}_t$  remains stable as  $k$  increases.

A similar argument applies to the remaining component. Instead of using the variance  $\text{Var}(f_t)$ , we apply the covariance  $\text{Cov}(f_t(k-1), f_t(k))$  between  $\hat{f}_t(k-1)$  and  $\hat{f}_t(k)$ , which denote the decomposed remaining components obtained at  $k-1$  and  $k$ , respectively.

Building upon this idea, we propose an extended measure analogous to PV that incorporates both the relative changes in variance between the two components and the degree of continuity within each component across successive values of  $k$ . Specifically, by replacing  $\text{Var}(s_t)$  and  $\text{Var}(f_t)$  with  $\text{Cov}(s_t(k-1), s_t(k))$  and  $\text{Cov}(f_t(k-1), f_t(k))$ , respectively, we define an extended indicator that simultaneously captures both the inter-component variance relationship and the intra-component consistency with respect to  $k$ . This combined measure provides a more comprehensive understanding of the decomposition behavior, particularly in applications where both smooth transitions and balanced variance allocation are critical considerations.

Moreover, for the constrained component, let  $\hat{s}_t^*$  denote a decomposed result obtained using a different method, and let  $\hat{s}_t(k)$  denote the result obtained using the ML model approach with a fixed value of  $k$ . We use the covariance  $\text{Cov}(s_t^*, s_t(k))$  between  $\hat{s}_t^*$  and  $\hat{s}_t(k)$  as a measure of similarity between these two results. A similar approach can be applied to the remaining component. Therefore, by integrating both aspects, we can compare the similarity between the decomposition results obtained by a different method and those from the ML model approach, with reference to the values of  $k$ .

The above concept serves as a useful basis for constructing indicators to evaluate decomposition results. From this perspective, a set of decomposed results obtained at a value of  $k$  that induces smaller fluctuations in the indicator can be regarded as reflecting a more stable underlying structure. Accordingly, it is preferable to adopt the decomposition corresponding to such a value of  $k$ , as it is likely to provide a more reliable representation of the time series structure. Furthermore, similar indicators can be developed to facilitate comparisons with alternative decomposition methods.

### 3.2.4. Evaluation Metrics for Assessing Decomposition Stability

In the constrained-remaining decomposition, evaluating the stability of the results is essential for understanding the underlying structure of a time series. One type of indicator for such evaluation is an extension of the product of variances. Since this indicator varies with the parameter  $k$ , it potentially reflects the structure of the decomposed results. However, it may exhibit unstable behavior depending on the characteristics of the time series and the chosen decomposition framework. This section discusses a more robust evaluation strategy based on the extended product of variances and explores its relationship to both likelihood-based estimation and alternative decomposition methods.

Building on the previous discussion, we define the Index of Structural Similarity (ISS) by extending the product of variances (PV) as follows:

$$\text{ISS}(k) = \frac{2\sqrt{\text{Cov}(s_t(k-1), s_t(k))\text{Cov}(f_t(k-1), f_t(k))}}{\text{Var}(y_t)}, \quad (18)$$

where  $k$  is the parameter of interest, and all terms are defined in the previous sections. This formula incorporates dimensional adjustment and is based on the observation that the maximum of  $\text{Cov}(s_t(k-1), s_t(k))\text{Cov}(f_t(k-1), f_t(k))$  can approach  $(0.25, \text{Var}(y_t))^2$ , so the ISS may take values close to 1.

The  $\text{ISS}(k)$  measures the overall similarity of the decomposition results before and after an increase in  $k$ . A decrease in ISS suggests that the decomposition results change abruptly as  $k$  increases.

Furthermore, for a given value of  $k$ , the similarity between the components obtained by a different method — namely,  $s_t^*$  and  $f_t^*$  — and those from the ML model approach can be evaluated by replacing  $s_t(k-1)$  and  $f_t(k-1)$  with  $s_t^*$  and  $f_t^*$  in the ISS formula in Eq. (18). We refer to this kind of index of structural similarity as the mutual index of structural similarity (MISS).

However,  $\text{ISS}(k)$  may exhibit unstable behavior depending on the characteristics of the time series. Thus, based on  $\text{ISS}(k)$ , we define an index of instability (IOI) as

$$\text{IOI}(k) = \Delta \log \text{ISS}(k) = \log \text{ISS}(k) - \log \text{ISS}(k-1). \quad (19)$$

As shown in equation (19),  $\text{IOI}(k)$  is defined as the difference in the logarithm of  $\text{ISS}(k)$  with respect to  $k$ , and thus serves as a measure corresponding to the rate of increase of  $\text{ISS}(k)$  with respect to  $k$ . A small  $\text{IOI}(k)$  indicates that  $\text{ISS}(k)$  changes gradually as  $k$  increases. This suggests that the decomposition structure is relatively invariant with respect to  $k$ . Hence, the value of  $k$  corresponding to a small magnitude of  $\text{IOI}(k)$  may be selected as the most structurally stable decomposition point. Therefore,  $\text{IOI}(k)$  can be used, along with the likelihood, as a reference indicator for determining the value of  $k$  and evaluating the estimation method.

### 3.2.5. Bidirectional Processing and Recursive Decomposition Strategies

The original ML model approach is processed in the order of  $n = 1, 2, \dots, N - k + 1$ , and the component decomposition is conducted according to this order. This is referred to as forward processing. On the other hand, it is referred to as backward processing if the same model is configured with the component decomposition proceeding in reverse order, that is,  $n = N - k + 1, N - k, \dots, 1$ . Since the correspondence between decomposed results and observations differs between the forward and backward processings, the results of component decomposition may also differ.

By averaging the estimation results at the same time point obtained from these two processings, a more stable and accurate component decomposition can be achieved. In other words, integrating the results obtained through bidirectional estimation – namely, the forward and backward processings – can provide more reliable and robust outcomes compared to estimation based solely on forward processing.

It should be noted that the constrained-remaining components decomposition is not necessarily completed in a single step. If necessary, the ML model approach can be repeatedly applied to the resulting components to further decompose a given component into two or more subcomponents. However, the interpretation and utilization of such decomposition results should be carefully determined in accordance with the characteristics of the data under analysis and the objectives and context of the application.

Moreover, [1] proposed an orthogonalization method to transform the resulting components into mutually uncorrelated time series when multiple components are obtained.

### 3.3. Reinforce the EML Model Approach

In this section, we aim to refine the method for determining the locations of outliers to reinforce the EML model approach.

#### 3.3.1. Outlier Detection and Estimation

The original problem involves identifying and estimating  $m$  outliers within a sample dataset of size  $N$ , where  $m$  is much smaller than  $N$ . This task presents a combinatorial challenge and is therefore considered a significant endeavor. To address this issue, it is crucial to obtain some indication of the potential positions of the outliers.

For the sake of completeness, the full system enhanced in this study is presented below. Notably, the third step constitutes a novel contribution of this work.

### (1) Determining the potential Positions of Outliers

When we do constrained-remaining decomposition using the ML model approach with a large WTI ( $k$ ), most of the outliers become assimilated into the estimate of the remaining component. Thus, the estimates  $\tilde{f}_t$  for the remaining component can be used as references. That is, we can assume that

$$\tilde{f}_t \approx f_t + u_t \quad (t = 1, 2, \dots, N).$$

Let  $\hat{f}_t$  represent an observation from the random process  $\tilde{F}_t$ , and let  $f_t$  be the realized value for a random process  $F_t$ , where  $\tilde{F}_t = F_t + u_t$ , and  $u_t$  is assumed to contain a potential outlier. It is further assumed that  $E\{F_t\} = 0$  and  $E\{F_t^2\} = C > 0$ . The expected value  $E\{\tilde{F}_t^2\}$  can be expressed as follows:

$$E\{\tilde{F}_t^2\} = E\{F_t^2 + 2F_t u_t + u_t^2\}.$$

Thus, we have

$$E\{\tilde{F}_t^2\} = \begin{cases} C + \delta_j^2 & \text{if } u_t \text{ includes an outlier, say, } \delta_j \text{ for } 1 \leq j \leq m \\ C & \text{otherwise} \end{cases}.$$

This suggests that  $E\{\tilde{F}_t^2\}$  becomes particularly large when  $\tilde{F}_t$  may contain an outlier. Additionally, when  $u_t = 0$ ,  $\tilde{f}_t^2$  provides an unbiased estimation of  $\tilde{F}_t^2 = F_t^2$ .

In summary, a large  $\tilde{f}_t^2$  is likely to contain  $\delta_t^2$ . When considering the number of potential outliers as  $m$ , if  $\tilde{f}_{t_1}^2 \geq \tilde{f}_{t_2}^2 \geq \dots \geq \tilde{f}_{t_m}^2$ , it is reasonable to infer that  $\tilde{f}_{t_1}$  contains the largest outlier in amplitude,  $\tilde{f}_{t_2}$  contains the second-largest outlier, and so on. In other words, one can establish the order for searching potential outlier positions as  $t_1, t_2, \dots, t_m$ . Subsequently, the positions of the outliers can be determined based on this order.

### (2) Estimation of Outliers and Their Number

Assume that for a given integer  $M$  and a vector of potential outliers  $\delta = (\delta_1, \delta_2, \dots, \delta_M)^T$ , a subvector  $\delta^{(m)}$  consisting of  $m$  elements selected from  $\delta$  is incorporated into the model for  $m = 1, 2, \dots, M$ . When the positions of outliers are determined (though they may be tentative), applying the extended ML model approach yields a Bayesian-type estimate  $\hat{\delta}^{(m)}$  for the outliers. This estimate is a form of maximum a posteriori probability, closely approximating the maximum likelihood estimate. Therefore, the estimates  $\hat{\delta}^{(m)}$  can be substituted into  $\delta^{(m)}$  in  $Y_{1:N}^*(\delta^{(m)})$  as an approximation of the maximum likelihood  $LL(\hat{\sigma}^2, k | Y_{1:N}^*(\hat{\delta}^{(m)}))$ , which is defined in Eq. (11).

This allows for an approximate calculation of the maximum log-likelihood, and based on that, the AIC( $m$ ) defined by

$$AIC(m) = -2LL(\hat{\sigma}^2, k | Y_{1:N}^*(\hat{\delta}^{(j)})) + 2(m + 1), \quad (20)$$

which depends on the value of  $m$ . Thus, the estimation of the number  $m$  of outliers using the minimum AIC method (see [10]).

### (3) Updating the Outlier Positions

Based on the potential setting of outlier locations  $\{t_1, t_2, \dots, t_m\}$  in sequential order, we estimate the parameters for each model corresponding to the number of outliers  $m = 0, 1, \dots, M$ , and compute the corresponding values of AIC( $m$ ) using Eq. (20). Specifically, the case of  $m = 0$  assumes that no outliers are present;  $m = 1$  assumes that a single outlier exists at time  $t_1$ ;  $m = 2$  assumes that outliers exist at times  $t_1$  and  $t_2$ ; and so forth. In this manner, we obtain the following sequence of AIC values:

$$\{AIC(m); m = 0, 1, \dots, M\} \quad (21)$$

Next, based on the AIC sequence in Eq. (21), we construct the sequence of AIC differences defined by  $\text{DAIC}(m) = \text{AIC}(m) - \text{AIC}(m - 1)$ , as follows:

$$\{\text{DAIC}(m); m = 1, 2, \dots, M\} \quad (22)$$

Within the sequence of AIC differences given in Eq. (22), any time point  $t_i$  associated with an index  $i$  such that  $\text{DAIC}(i) < 0$  can be considered to have special significance. That is, incorporating an additional outlier at time  $t_i$ , on top of the outliers already introduced at  $t_1, t_2, \dots, t_{i-1}$ , leads to a decrease in the AIC value, which implies an improvement in model fit. Therefore, such an outlier can be regarded as the  $i$ -th outlier that should be included in the model.

This reasoning allows us to update the outlier locations. Specifically, we examine the sequence of  $\text{DAIC}(m)$  in Eq. (22) in ascending order. For instance, if the AIC difference  $\text{DAIC}(m_1)$  corresponding to outlier number  $m_1$  is the smallest, we update the location of the  $m_1$ -th outlier to the corresponding time point  $t_1^*$ . Similarly, if  $\text{DAIC}(m_2)$  is the second smallest, we update the location of the  $m_2$ -th outlier to the corresponding time point  $t_2^*$ . In this way, we update the previous setting of outlier locations to the following:

$$\{t_1^*, t_2^*, \dots, t_M^*\} \quad (23)$$

We refer to the sequence in Eq. (23) as the updated outlier locations.

Once this update of outlier locations is performed, the updated set is used as a new setting, and the procedure of replacing outlier positions is repeated. In theory, this iterative process continues until the updated outlier locations coincide with the previous setting. However, from a practical perspective, it is generally sufficient if the elements contributing to negative DAIC can be identified in a stable manner before and after the update.

Ultimately, including outliers in the model that lead to negative DAIC differences results in a lower AIC value and, consequently, a better-fitting model. Thus, through this process, the number of outliers to be included in the model is determined automatically. Specifically, if the first  $\hat{m}$  outliers in the updated sequence of Eq. (23), up to  $t_{\hat{m}}^*$ , all result in negative values of  $\text{DAIC}(\hat{m})$ , then the outliers at the corresponding time points  $t_1^*, t_2^*, \dots, t_{\hat{m}}^*$  should be incorporated into the model. In doing so, we obtain the optimal model corresponding to the minimum AIC.

We shall refer to the aforementioned method of determining the locations and number of outliers as the *AIC reduction maximization method*. This method takes into account the dependencies among outliers and constitutes the next significant contribution of this study.

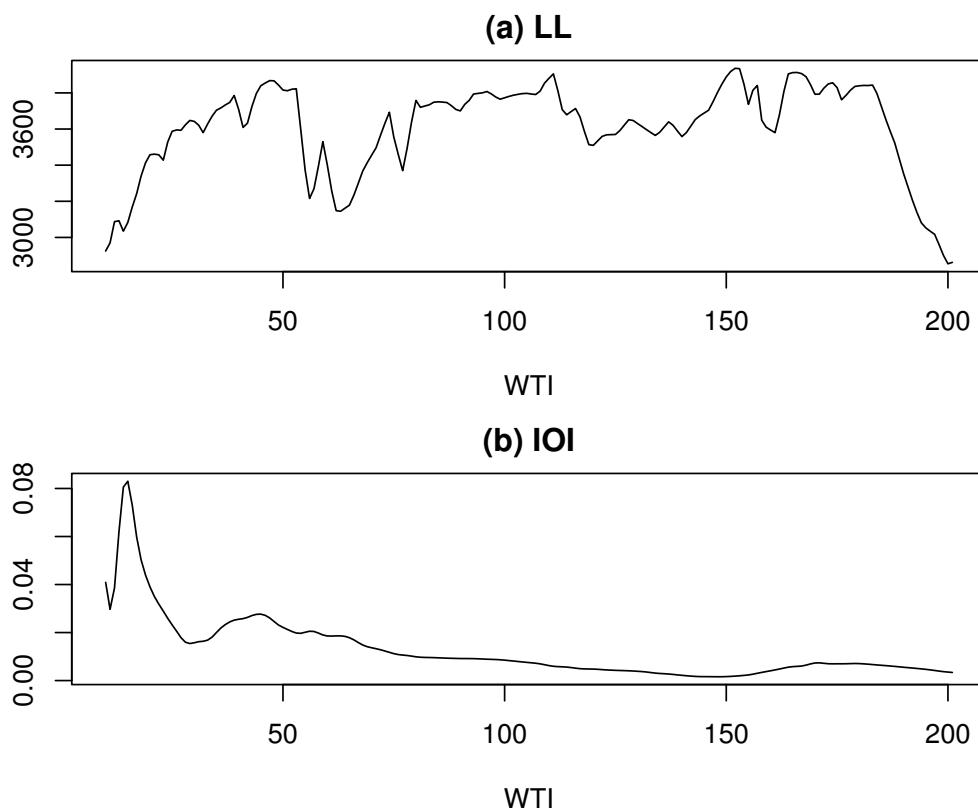
## 4. Empirical Examples

### 4.1. Empirical Analysis of Capital Investment in Japan

The first empirical example analyzes business expenditures for new plant and equipment (BE) in Japan. The goal of this example is to demonstrate the analytical procedure and highlight the effectiveness of the reinforcement in the ML model approach.

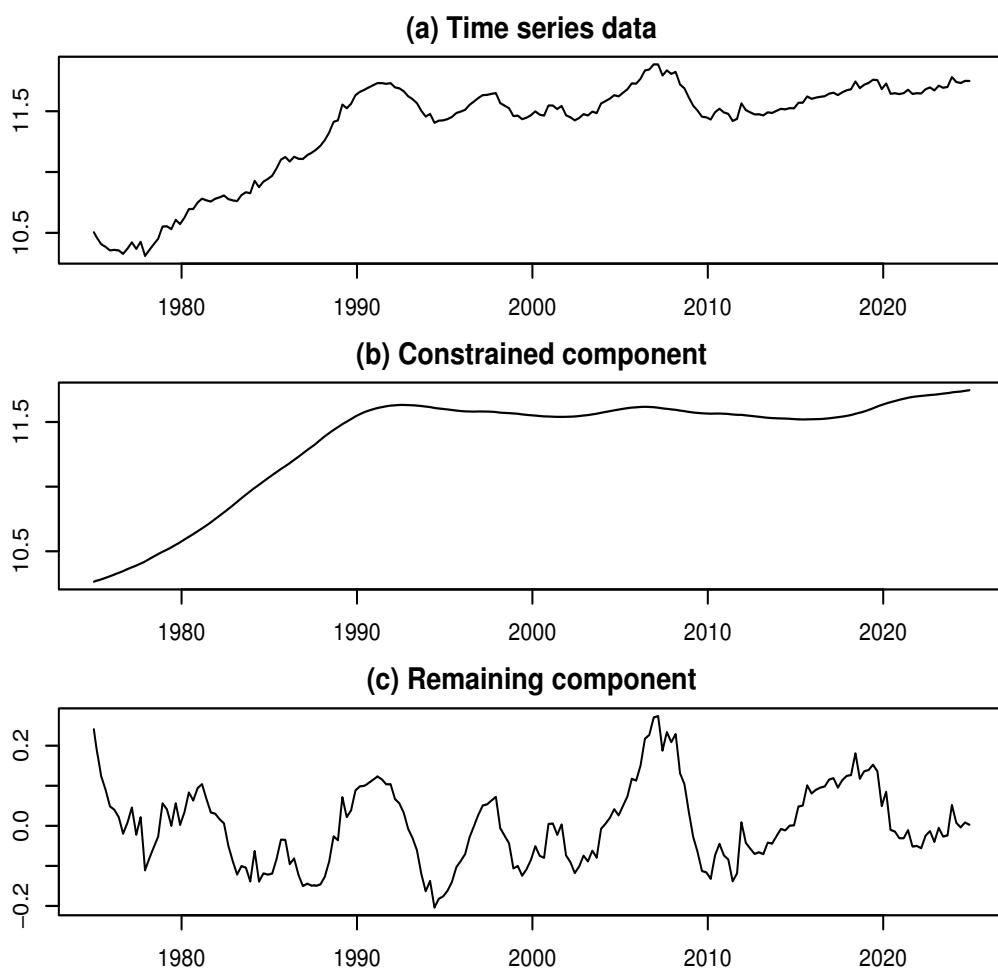
The time series of BE reflects levels of capital investment and is known for exhibiting distinct structural characteristics. The original BE data were obtained from the website of the Japanese Cabinet Office (see [11]). The data constitute a monthly time series spanning from January 1975 to December 2024, amounting to  $N = 600$  observations. Figure 2a displays a plot of the logarithmically transformed BE series (log-BE).

We applied the ML model approach to decompose the log-BE time series, performing the decomposition for each value of  $k$  ranging from 3 to 202. As shown in Figure 1a, the log-likelihood (LL) exhibits local maxima at three values of  $k$ : 47, 111, and 152, corresponding to LL values of 3867.97, 3905.51, and 3936.1, respectively. Among these, the highest LL is achieved at  $\hat{k} = 152$ , which is adopted as the estimate of  $k$ . Furthermore, Figure 1b indicates that the IOI reaches its minimum around  $\hat{k}$ . These findings suggest that the maximum likelihood estimate of  $k$  effectively captures a stable underlying structure.



**Figure 1.** Log-likelihood and IOI vs.  $k$  for the log-BE time series

Figure 2b,c illustrate the decomposed results for the log-BE time series, with Figure 2b depicting the constrained component and Figure 2c representing the remaining component.



**Figure 2.** Data and decomposed results for the log-BE time series

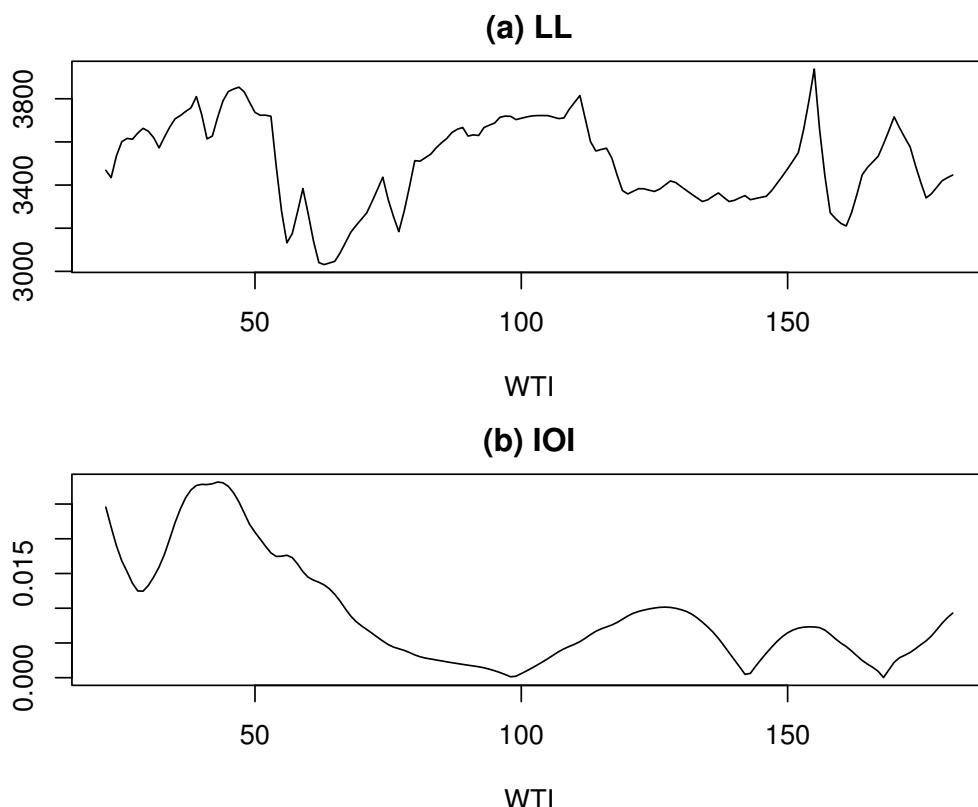
Incidentally, to validate the results of this study, we refer to the state-space modeling approach introduced in [8]. In that work, the R function `season` is employed to decompose a time series into three components: a trend component, an AR component, and observation noise. When this function was applied to the log-BE series, the order of the trend component was estimated to be 2, and that of the AR component was estimated to be 4. Based on these estimates, the time series was decomposed accordingly.

To align with the ML model framework, we treated the trend component as the constrained component and considered the combined AR component and observation noise as the remaining component. For each value of  $k$ , we computed the MISS values by comparing the decomposition results from the ML model approach with those obtained using the `season` function. The MISS reached its minimum around  $k = 11$ , suggesting that the conventional state-space-based decomposition closely corresponds to the ML model approach at this value of  $k$ . Notably, around  $k = 11$ , the IOI value is approximately 0.08, which is near its maximum. This implies that the conventional component decomposition using the state-space model lacks stability.

These findings suggest that the constrained component produces a highly smooth time series, effectively capturing the long-term trend in log-BE. Consequently, most short-term fluctuations are absorbed into the remaining component, which exhibits complex and irregular behavior. This observation provides a strong rationale for reapplying the ML model approach to further decompose the (first) remaining component.

Thus, we further decomposed the log-BE time series for each value of  $k$  ranging from 3 to 182. As shown in Figure 3a, similar to the first component decomposition, the LL exhibits local maxima

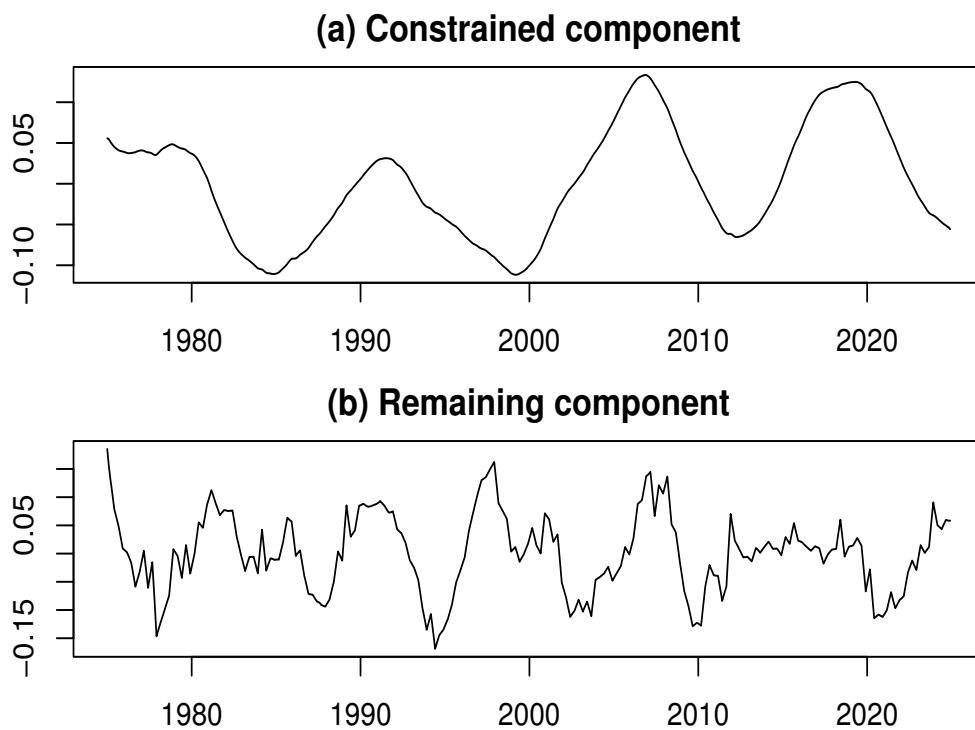
at three values of  $k$ : 47, 111, and 155, corresponding to LL values of 3853.70, 3814.70, and 3937.52, respectively. Among these, the highest LL is achieved at  $\hat{k} = 155$ .



**Figure 3.** Log-likelihood and IOI vs.  $k$  for the second component decomposition

However, as shown in Figure 3b, the IOI values at these  $k$  points are elevated compared to their surroundings, suggesting that the most stable decomposition may not be achieved at these values. A closer examination of Figure 3b reveals a noticeable dip in the IOI at  $k = 97$ . Although the LL value at this point, 3719.58, is not exceptionally high, the LL values in its vicinity remain relatively stable at a high level. Therefore,  $k = 97$  is adopted as the estimate of  $k$  for the second component decomposition.

Figure 4 displays the results of the second decomposition applied to the remaining component obtained from the first decomposition. Figure 4a,b show the constrained and remaining components, respectively.



**Figure 4.** Results of the second component decomposition

The constrained component in the second decomposition, shown in Figure 4a, exhibits cyclical behavior with a period of over a decade, aligning with the Juglar cycle. This point deserves emphasis: in this case, the time series subject to decomposition is itself a remaining component containing cyclical fluctuations of varying periods. Consequently, the constrained component extracted through this process captures a long-period, highly smooth cyclical fluctuation.

Furthermore, the remaining component in the second decomposition, shown in Figure 4b, is characterized by short-term cyclical fluctuations. These findings reinforce the consistency of the results with business cycle theory and underscore the effectiveness of the proposed approach in identifying and analyzing economic cycles (see [1]).

The results presented above suggest the following: by incorporating additional metrics and estimation methods alongside the conventional maximum likelihood approach, more stable decomposition results can be achieved. Naturally, different decomposition outcomes may arise depending on the number of components extracted and the choice of  $k$  values. This underscores the fact that data analysis is inherently both data-driven and purpose-driven; thus, trial and error, guided by the analytical objectives of the data under investigation, is essential.

#### 4.2. Empirical Analysis of Industrial Production in Japan

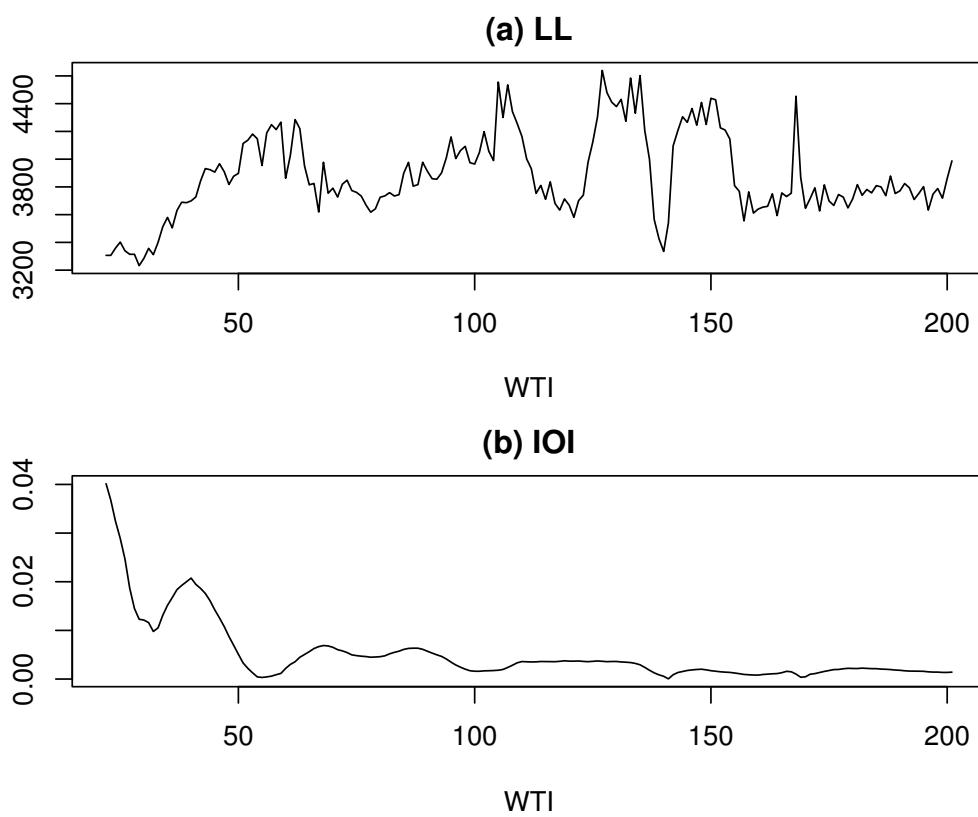
In this section, we present an empirical analysis of the seasonally adjusted index of industrial production (SAIIP) in Japan as a second example. This example demonstrates the effectiveness of the reinforced EML model approach in detecting and estimating outliers in time series data.

The SAIIP data were obtained from the same source as the BE data. As SAIIP serves as a principal indicator for analyzing economic trends, it is frequently used in empirical studies. Due to its sensitivity to sudden changes in economic conditions, outliers tend to appear frequently.

To facilitate comparison with the empirical example in [7], we use the data as a monthly time series from January 1975 to December 2022, comprising a total of  $N = 576$  months. A logarithmic transformation is applied to the SAIIP time series, which we refer to as log-SAIIP. The objective is to analyze the time series behavior of log-SAIIP.

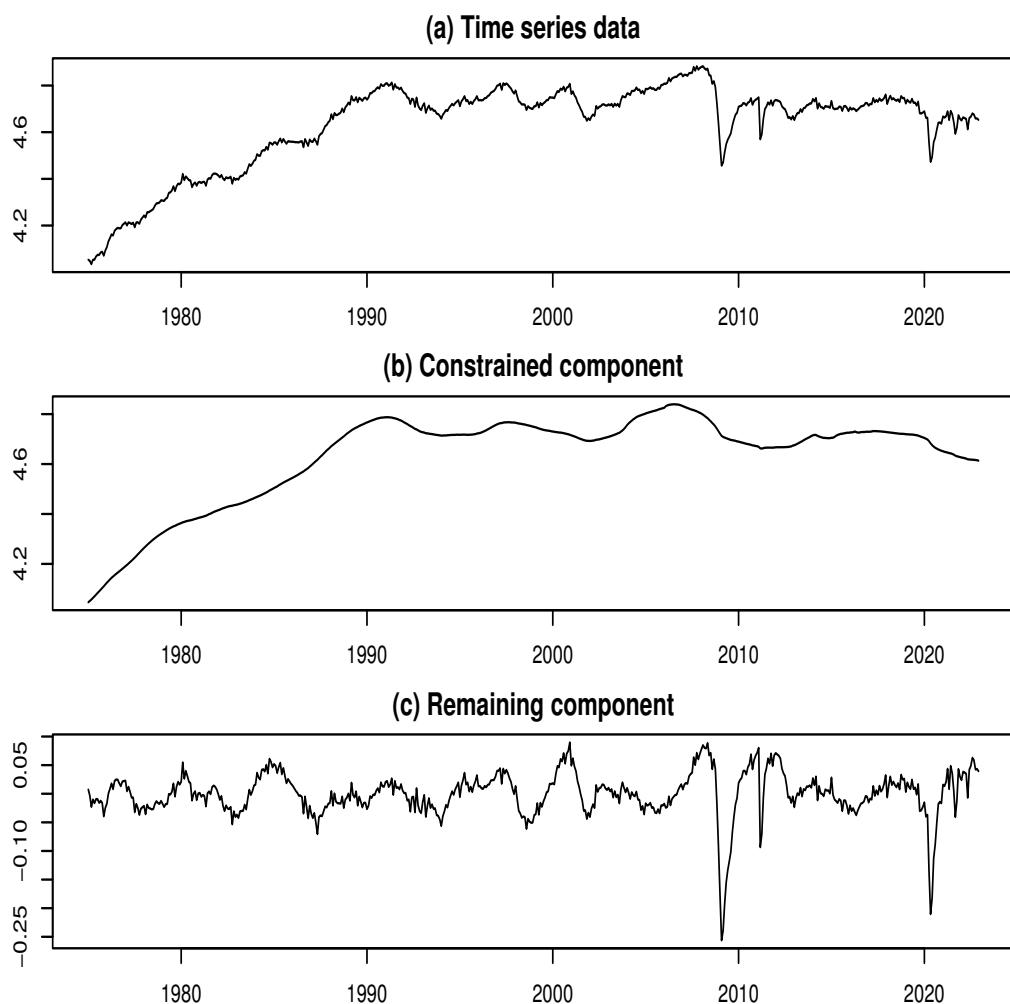
Figure 6a presents the time series plot of log-SAIIP. Two prominent declines are clearly observable in the series. The first, occurring around February 2009, is attributable to the aftermath of the 2007–2008 global financial crisis. The second, around May 2020, reflects the impact of the COVID-19 pandemic. These sharp drops are likely to be associated with the presence of multiple outliers.

We applied the ML model approach to decompose the log-SAIIP time series, performing the decomposition for each value of  $k$  ranging from 3 to 202. As shown in Figure 5a, the log-likelihood (LL) exhibits three prominent local maxima at  $k = 62$ ,  $k = 105$ , and  $k = 135$ , corresponding to LL values of 4285.37, 4555.44, and 4601.83, respectively. Furthermore, at  $k = 168$ , the LL shows a spike-like surge, reaching a peak value of 4453.67. A closer examination of Figure 5b reveals a noticeable dip in the IOI at  $k = 62$ , which corresponds to the first local maximum of the LL. Although the LL at this point is not the global maximum, the surrounding LL values remain relatively stable at a high level. Therefore, for the purpose of comparison with the findings of Kyo (2025b), and based on the above observation, we adopt  $k = 62$  as the estimated value of  $k$  for the second component decomposition.



**Figure 5.** Log-likelihood and IOI vs.  $k$  for the log-SAIIP time series

Figure 6b,c present the results of the constrained-remaining component decomposition using the ML model approach with  $\hat{k} = 62$ . In Figure 6b, the constrained component exhibits a smooth trend, while Figure 6c shows that the remaining component displays cyclical variations, indicating the presence of business cycles in Japan. Notably, a substantial portion of the sharp declines observed around February 2009 and May 2020 can be attributed to the remaining component. These abrupt drops in the time series are likely to reflect the influence of outliers caused by sudden economic shocks. Therefore, the elements associated with these variations should be identified and treated as outliers, as they may distort the analysis of business cycles.



**Figure 6.** Data and decomposed results for the log-SAIIP time series

The upper bound for the number of potential outliers was set to  $M = 30$ , and the potential locations of the outliers were determined based on large squared values in the decomposed remaining component time series.

Furthermore, based on the AIC reduction maximization method, the number of outliers was estimated as  $\hat{m} = 24$ , with the minimum AIC value being  $-8799.09$ . The estimated outlier positions are at time points 410, 411, 545, 412, 413, 409, 414, 546, 547, 415, 416, 436, 435, 548, 544, 408, 149, 398, 401, 399, 397, 394, 400, and 403. Arranged in chronological order, these positions mainly correspond to the periods from December 2008 to August 2009 and from April 2020 to July 2020.

Finally, we performed the estimation of outliers using  $\hat{m} = 24$ . Figure 7a depicts the time series containing outliers, Figure 7b displays the final outlier-adjusted time series, while Figure 7c,d show the estimates of the constrained and remaining components for the outlier-adjusted time series. It can be confirmed that, compared to the initial estimate of the remaining component in Figure 6c, the estimation results with the outlier-adjusted data in Figure 7d show a significant improvement in uniformity.

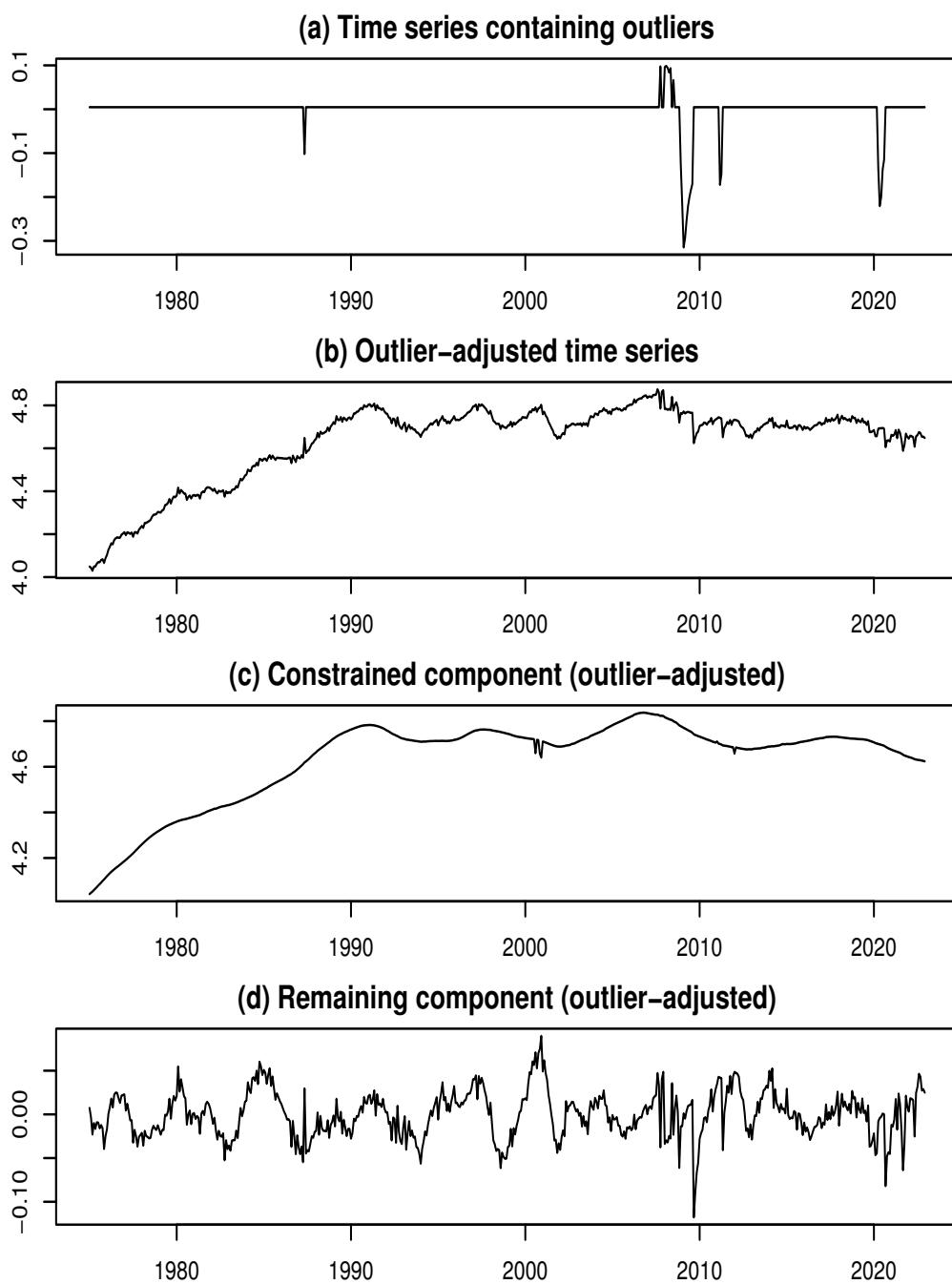


Figure 7. Final results of decomposition for the log-SAIIP time series

As a reference for comparison, we again present the main results obtained in [7]. In [7], we used an EML model approach, which is reviewed in Section 3.3, for detecting and estimating outliers. The number of outliers was estimated as  $\hat{m} = 22$  based on the method of determining outlier locations according to the order of the squared standardized outliers, resulting in a minimum AIC value of  $-8009.47$ . The outlier positions were estimated at the following time points: 410, 411, 412, 413, 409, 414, 415, 435, 416, 545, 436, 546, 408, 149, 547, 544, 398, 394, 397, 399, 401, and 548.

Thus, by updating the method for detecting outliers, the AIC decreased by 789.62, thereby improving the reliability of the estimation results. As a result, two additional outliers were detected. These newly added outliers are active at time points 400 and 403, corresponding to April and July of 2008, respectively. The marginal contributions to the AIC reduction for each are 13.43 and 17.53, respectively. The difference between the component decomposition results obtained in [7] and those

shown in Figure 7 is not easily visually discernible in the graph, so it will be omitted. However, as described below, a comparison based on the numerical values of the evaluation metrics is possible.

Therefore, we evaluate the decomposition results using the index of symmetry and uniformity (ISU), proposed by [12], as a benchmark. For a given time series, the ISU is defined as the logarithmic ratio of the standard deviation of the series' absolute values to the standard deviation of the original series. For the outlier-adjusted remaining component, a larger standard deviation of the absolute values indicates a smoother outlier-adjusted constrained component. This suggests that the influence of outliers on the decomposition result is weaker, leading to greater stability, which is desirable. Conversely, a smaller standard deviation of the absolute values implies greater symmetry around the mean and also a weaker influence of outliers on the remaining component – another desirable outcome. Therefore, the decomposition results can be evaluated using the minimum ISU criterion applied to the outlier-adjusted remaining component.

The following results were obtained based on the above findings. For the result derived from the method proposed by [7], the ISU was  $-0.2386$ , whereas for the result based on the present AIC reduction maximization method, the ISU decreased to  $-0.2403$ . This demonstrates the advantage of the AIC reduction maximization method in updating conditions, as evaluated by the minimum ISU principle.

Considering these outcomes, a key feature of the proposed method is its thorough implementation of the ML model approach, which makes it particularly promising for handling outliers in the decomposition of constrained and remaining components.

## 5. Summary and Discussion

We began by reviewing the moving linear (ML) model approach proposed by [1], along with its extension, the extended moving linear (EML) model introduced by [7], which together form the theoretical foundation of this study. Building on these frameworks, we introduced several methodological advancements, particularly focusing on improving outlier detection and estimation.

Although WTI was initially developed as a practical tool, we clarified its theoretical significance in separating constrained and remaining components, thereby contributing to the stability of estimation and enhancing both the interpretability and robustness of the model. To support model evaluation, we also proposed new metrics that offer a broader perspective than traditional likelihood-based methods. A core contribution is the index of instability (IOI), constructed from the index of structural similarity (ISS), which serves as a powerful supplementary criterion to the log-likelihood and reinforces the validity of maximum likelihood estimation results.

Other key methodological developments include bidirectional processing strategies that integrate forward and backward decompositions. Averaging the results from both directions leads to more stable and reliable component estimates. We also proposed a novel three-step approach for outlier detection: identifying potential outlier positions, estimating the number and values of outliers, and updating their positions systematically. This process employs the Akaike information criterion (AIC) to iteratively refine the model by selecting positions that reduce AIC and improve model fit.

A major contribution in this regard is the AIC reduction maximization method, which accounts for interdependencies among outliers and automatically determines their optimal number and locations. This approach significantly enhances the model's explanatory and predictive performance, strengthening its ability to handle outliers in time series data.

We demonstrated the effectiveness of the reinforced EML model through two empirical examples. The first used Business Cycle Indicator (BE) data from January 1975 to December 2022. After logarithmic transformation, notable outliers were detected, particularly around the global financial crises and the COVID-19 pandemic. The application of the ML model improved business cycle analysis by adjusting for these outliers, achieving a lower AIC and identifying additional structural shifts. This example also illustrated the utility of the IOI-based parameter estimation.

The second example analyzed the seasonally adjusted index of industrial production (SAIIP) over the same period. As with the first case, economic shocks resulted in discernible outliers. The AIC reduction maximization method estimated 24 such outliers, leading to significantly improved decomposition. Evaluation metrics, including the index of symmetry and uniformity (ISU), confirmed that the proposed method was more reliable than prior approaches. This case highlighted the effectiveness of our method for outlier detection and estimation based on AIC reduction.

Together, these empirical findings underscore the robustness of the reinforced EML model in addressing outliers, enabling more accurate and insightful analyses of business cycles and economic dynamics.

**Funding:** Not applicable.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** All the data used in this paper are publicly available. The authors possess these data and can provide them upon a reasonable request.

**Conflicts of Interest:** The authors declare no competing interests.

## References

1. Kyo, K.; Kitagawa, G. A moving linear model approach for extracting cyclical variation from time series data. *J. Bus. Cycle Res.* **2023**, *19*, 373-397.
2. Kyo, K.; Noda, H.; Fang, F. An integrated approach for decomposing time series data into trend, cycle and seasonal components. *Math. Comput. Model. of Dyn. Syst.* **2024**, *30*, 792-813.
3. Ren, H.; Xu, B.; Wang, Y.; Yi, C.; Huang, C.; Kou, X.; Xing, T.; Yang, M.; Tong, J.; Zhang, Q. Time-series anomaly detection service at Microsoft. <https://doi.org/10.1145/3292500.3330680>
4. Vishwakarma, G.K.; Paul, C.; Elsawah, A.M. An algorithm for outlier detection in a time series model using backpropagation neural network. *Journal of King Saud University - Science*, **2020** *32*, 3328-3336.
5. Jamshidi, E.J.; Yusup, Y.; Kayode, J.S.; Kamaruddin, M.A. Detecting outliers in a univariate time series dataset using unsupervised combined statistical methods: A case study on surface water temperature. *Ecological Informatics*, **2022** *69*, 101672.
6. Kyo, K. An approach for the identification and estimation of outliers in a time series with a nonstationary mean. In *Proceedings of The 2023 World Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE'23)*, IEEE Computer Society, 1477-1482.
7. Kyo, K. Enhancing business cycle analysis by integrating anomaly detection and components decomposition of time series data. *Statistical Methods & Applications*. 2025, <https://doi.org/10.1007/s10260-025-00780-6>.
8. Kitagawa, G. *Introduction to Time Series Modeling with Application in R*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2020.
9. Kitagawa, G.; Gersch, W. A smoothness priors state space modeling of time series with trend and seasonality. *Journal of the American Statistical Association*, **1984**, *79*, 378-389.
10. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control*, **1974**, *AC-19*, 716-723.
11. Japanese Cabinet Office (2025). Coincident index, <https://www.esri.cao.go.jp/en/stat/di/di-e.html>.
12. Kyo, K. Identifying and estimating outliers in time series with nonstationary mean through multi-objective optimization method. In *Big Data, Data Mining and Data Science: Algorithms, Infrastructures, Management and Security*, 2025, De Gruyter.

**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.