entropy

MDPI

Article

Segmentation of High Dimensional Time-Series Data Using Mixture of Sparse Principal Component Regression Model with Information Complexity

Yaojin Sun and Hamparsum Bozdogan *

Department of Business Analytics and Statistics, University of Tennessee, Knoxville, TN 37996, USA; ysun52@vols.utk.edu

* Correspondence: bozdogan@utk.edu

Received: 29 August 2020; Accepted: 13 October 2020; Published: 17 October 2020

Abstract: This paper presents a new and novel hybrid modeling method for the segmentation of high dimensional time-series data using the mixture of the sparse principal components regression (*MIX-SPCR*) model with information complexity (ICOMP) criterion as the fitness function. Our approach encompasses dimension reduction in high dimensional time-series data and, at the same time, determines the number of component clusters (i.e., number of segments across time-series data) and selects the best subset of predictors. A large-scale Monte Carlo simulation is performed to show the capability of the *MIX-SPCR* model to identify the correct structure of the time-series data successfully. *MIX-SPCR* model is also applied to a high dimensional Standard & Poor's 500 (S&P 500) index data to uncover the time-series's hidden structure and identify the structure change points. The approach presented in this paper determines both the relationships among the predictor variables and how various predictor variables contribute to the explanatory power of the response variable through the sparsity settings cluster wise.

Keywords: high dimensional time-series; segmentation; mixture regression; sparse PCA; entropy-based robust EM; information complexity criteria

1. Introduction

This paper presents a new and novel method for the segmentation and dimension reduction in high dimensional time-series data. We develop hybrid modeling between *mixture-model cluster analysis* and *sparse principal components regression* (MIX-SPCR) model as an expert unsupervised classification methodology with *information complexity* (ICOMP) criterion as the fitness function. This new approach performs dimension reduction in high dimensional time-series data and, at the same time, determines the number of component clusters.

The research of time-series segmentation and change point positioning has been a hot topic of research for a long time. Different research groups have provided solutions with various approaches in this area, including, but not limited to, Bayesian methods Barber et al. [1], fuzzy systems Abonyi and Feil [2], and complex system modeling Spagnolo and Valenti [3], Valenti et al. [4], S Lima [5], Ding et al. [6]. We group these approaches into two branches, one based on complex systems modeling and the other on the statistical model through parameter estimation and inference. Among the complex systems-based modeling approaches, it is worth noting a series of papers that use the stochastic volatility model by Spagnolo and Valenti [3]. For example, these authors used a nonlinear Hestone model to analyze 1071 stocks on the New York Stock Exchange (1987–1998). After accounting for the stochastic nature of volatility, the model is well

suited to extracting the escape time distribution from financial time-series data. The authors also identified the NES (Noise Enhanced Stability) effect to measure market dynamics' stabilizing effect. The approach we propose in this paper belongs to another branch of using a statistical model on time scales. Along with the empirical analysis, we show a broader view of how different companies/sectors behaved across different periods. In particular, we use a mixture-model based statistical methodology to segment the time-series and determine change points.

The mixture-model cluster analysis of regression models is not new. These models are also known as *"cluster-wise regression"*, *"latent models"*, and *"latent structure models of choice"*. These models have been well-studied among statisticians, machine learning researchers, and econometricians in the last several decades to construct time-series segmentation models and identify change points. They have many useful theoretical and applied properties. Mixture-model cluster analysis of regression models is a natural extension of the standard multivariate Gaussian mixture-model cluster analysis. These models are beneficial to study heterogeneous data sets that involve not just one response variable but can have several responses or target-dependent variables simultaneously with a given set of independent variables. Recently, they have been proven to be a precious class of models in various disciplines in *behavioral and economic research*, *ecology*, *financial engineering*, *process control*, *and monitoring*, *market research*, *transportation systems*. Additionally, we also witness the mixture model's usage in the *analysis of scanner panel*, *survey, and other choice data to study consumer choice behavior and dynamics* Dillon et al. [7].

In reviewing the literature, we note that Quandt and Ramsey [8] and Kiefer [9] studied data sets by applying a mixture of two regression models using moment generating function techniques to estimate the unknown model parameters. Later, De Veaux [10] developed an EM algorithm to fit a mixture of two regression models. DeSarbo and Cron [11] used similar estimating equations and extended the earlier work done on a mixture of two regression models to a mixture of K-component regression models. For an excellent review article on this problem, we refer the reviewers to Wedel and DeSarbo [12].

In terms of these models' applications in the segmentation of time-series, they can be seen in the early work of Sclove [13], where the author applied the mixture model to the segmentation of US gross national product, a high dimensional time-series data. Specifically, Sclove [13] used the statistical model selection criteria to choose the number of classes.

With the currently existing challenges in mind in the segmentation of time-series data, in this paper, our objective and goal are to develop a new methodology which can:

- Identify and select variables that are sparse in the *MIX-SPCR* model.
- Treat each time segment continuously in the process with some specified probability density function (pdf).
- Determine the number of time-series segments and the number of sparse variables and estimate the structural change points simultaneously.
- Develop a robust and efficient algorithm for estimating model parameters.

We aim to achieve these objectives by developing the information complexity (ICOMP) criteria as our fitness function throughout the paper for the segmentation of high-dimensional time-series data.

Our approach involves a two-stage procedure. We first make a variable selection by using SPCA with the benefit of sparsity. We then fit the sparse principal component regression (SPCR) model by transforming the original high dimensional data into several main principal components and estimating relationships between the sparse component loadings and the response variable. In this way, the mixture model not only handles the curse of dimensionality but also maintains the model's excessive explanatory power. In this manner, we choose the best subset of predictors and determine the number of time-series segments in the *MIX-SPCR* model simultaneously using ICOMP.

The rest of the paper is organized as follows. In Section 2, we present the model and methods. In particular, we first briefly explain sparse principal component analysis (SPCA) due to Zou et al. [14] in Section 2.1. In Section 2.2, we modify SPCA and develop mixtures of the sparse principal component regression (*MIX-SPCR*) model for the segmentation of time-series data. In Section 3, we present a regularized entropy-based Expectation and Maximization (EM) clustering algorithm. As is well known, the EM algorithm performs through maximizing the likelihood of the mixture models. However, to make the conventional EM algorithm robust (not sensitive to initial values) and converge to global optimum, we use the robust version of the EM algorithm for the *MIX-SPCR* model based on the work of Yang et al. [15]. These authors addressed the robustness issue by adding an entropy term of mixture proportions to the conventional EM algorithm's objective function. While our EM algorithm is in the same spirit of the Yang et al. [15] approach, there are significant differences between our approach and theirs. Yang's robust EM algorithm merely deals with the usual clustering problem without involving any response (or dependent) variable or time factor in the data. We extend it to the case of the *MIX-SPCR* model in the context of time-series data. In Section 4, we discuss various information criteria, specifically the information complexity based criteria (ICOMP). We derive the ICOMP for the *MIX-SPCR* model based on Bozdogan's previous research ([16–20]). In Section 5, we present our Monte Carlo simulation study. Section 5.2 involves an experiment on the detection of structural points, and Section 5.3 presents a large scale Monte Carlo simulation verifying the advantage of the *MIX-SPCR* with statistical information criteria. We provide a real data analysis in Section 6 using the daily adjusted closing S&P 500 index and stock prices from the Yahoo Finance database that spans the period from January 1999 to December 2019. Finally, our conclusion and discussion are presented in Section 7.

2. Model and Methods

In this section, we briefly present the *sparse principal component analysis* (*SPCA*), *sparse principal component regression* (*SPCR*) as a background. Then, by hybridizing these two methods within the mixture model, we propose the *mixture-model cluster analysis of sparse principal component regression* (abbreviated as *MIX-SPCR* model hereafter), for segmentation of high dimensional time-series datasets. Compared with a simple linear combination of all explanatory variables (i.e., the dense PCA model), the new approach interprets better because it maintains a sparsity specification.

Referring to Figure 1, we first show the overall structure of the model in this paper. The overall processing flow is that we clean and standardize the data after obtaining the time-series data. Subsequently, we specify the number of time-series segments and how many Sparse Principal Components (SPCs) each segment contains. Using the Robust EM algorithm (Section 3), we estimate the model parameters, especially the boundaries (also known as *change points*) of each time segment. The information criterion values are calculated using the method of Section 4. By testing different numbers of time segments/SPCs, we obtain multiple criterion values. According to the calculated information criterion values, we choose the most appropriate model with the estimated parameters.

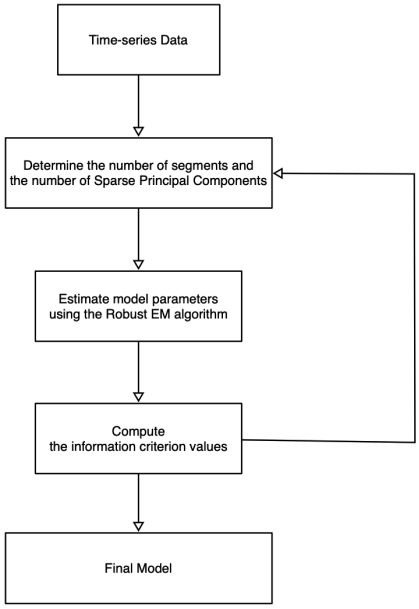

Figure 1. The flowchart of the MIX-SPCR method.

2.1. Sparse Principal Component Analysis (SPCA)

Given the input data matrix, $\mathbf{X}$ with n number of observations and p variables, we decompose $\mathbf{X}$ using the singular value decomposition (SVD). We write the decomposition procedure as $\mathbf{X} = \mathbf{UDV}^T$, where $\mathbf{D}$ is a diagonal matrix of singular values and orthogonal columns $\mathbf{U}$ and $\mathbf{V}$ as the left and right singular vectors. When we perform SVD of a data matrix $\mathbf{X}$ that has been centered, by subtracting each column's mean, the process is the well-known *principal component analysis* (*PCA*). As discussed by Zou et al. [14], PCA has several advantages as compared with other dimensionality reduction techniques. For example, the PCA can sequentially identify the source of variability by considering the linear combination of all the variables. Because of the orthonormal constraint during the computation, all the calculated *principal components* (*PCs*) have clear geometrical interpretation corresponding to the original data space as a dimension reduction technique. Because PCA can deal with *"the curse of dimensionality"* of high-dimensional data sets, it has been widely used in real-world scenarios, including biomedical and financial applications.

Even though PCA has excellent properties that are desirable in real-world applications and statistical analysis, the interpretation of PCs is often difficult since it includes all the variables as linear combinations of all the original variables in each of the PCs. In practice, the principal components always have a large number of non-zero coefficient values for corresponding variables. To resolve this drawback, researchers proposed various improvements focusing on PCA's sparsity while maintaining the minimal loss of information. Shen and Huang [21] designed an algorithm to iteratively extract top PCs using the so-called *penalized least sum of square* (*PLSS*) criterion. Zou et al. [14] utilized the lasso penalty (via Elastic Net) to maintain a sparse loading of the principal components, which is named *sparse principal component analysis* (*SPCA*).

In this paper, we use the sparse principal component analysis (SPCA) proposed by Zou et al. [14]. Given the data matrix $\mathbf{X}$, we minimize the objective function to obtain the SPCA results:

$$(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}) = \arg\min_{\mathbf{A},\mathbf{B}} \sum_{i=1}^{n} \left\| \mathbf{x}_i^T - \mathbf{A}\mathbf{B}^T\mathbf{x}_i^T \right\|^2 + \sum_{j=1}^{k} \lambda_{1,j} \left\| \mathbf{B}_{(j)} \right\|_1 + \lambda_2 \sum_{j=1}^{k} \left\| \mathbf{B}_{(j)} \right\|_2^2, \quad (1)$$

subject to

$$\mathbf{A}^T\mathbf{A} = I_k. \tag{2}$$

where I_k is the identity matrix. We maintain the hyperparameters $\lambda_{1,j}$ and λ_2 to be non-negative. The $\mathbf{A}$ and $\mathbf{B}$ matrices of size $(p \times k)$ are given by

$$\mathbf{B} = \begin{bmatrix} B_{1,1} & \cdots & B_{1,k} \\ \vdots & \ddots & \vdots \\ B_{p,1} & \cdots & B_{p,k} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{(1)} \mid \ldots \mid \mathbf{B}_{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_p \end{bmatrix}, \tag{3}$$

and

$$\mathbf{A} = \begin{bmatrix} A_{1,1} & \cdots & A_{1,k} \\ \vdots & \ddots & \vdots \\ A_{p,1} & \cdots & A_{p,k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{(1)} \mid \ldots \mid \mathbf{A}_{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_p \end{bmatrix}. \tag{4}$$

If we choose the first k principal components from the data matrix $\mathbf{X}$, then the estimate $\hat{\mathbf{B}}_{(j)}$ contains the sparse loading vectors, which are no longer orthogonal.

A bigger $\lambda_{1,j}$ means a greater penalty for having non-zero entries in $\hat{B}_{(j)}$. By using different $\lambda_{1,j}$, we control the number of zeros in the jth loading vector. If $\lambda_{1,j} = 0$ for $j = 1, 2, \ldots, k$, this problem reduces to usual PCA.

Zou et al. [14] proposed a generalized SPCA algorithm to solve the optimization problem in Equation (1). The algorithm applies the Elastic Net (EN) to estimate $\mathbf{B}_{(j)}$ iteratively and update matrix $\mathbf{A}$. However, this algorithm is not the only available approach for extracting principal components with sparse loadings. The SPCA could also be computed through dictionary learning by Mairal et al. [22]. By introducing the probability model of principal component analysis, SPCA is equivalent to the *sparse probabilistic principal component analysis* (*SPPCA*) if the prior is Laplacian distribution for each weight matrix element (Guan and Dy [23], Williams [24]). For further discussion on SPPCA, we refer readers to those related publications for more details.

Next, we introduce the *MIX-SPCR* model for the segmentation of time-series data.

2.2. Mixtures of SPCR Model for Time-Series Data

Suppose the continuous response variable is denoted as $\mathbf{y} = \{y_i | 1 \leq i \leq n\}$, where n represents the number of observations (time points). Similarly, we have the predictors denoted as $\mathbf{X} = \{\mathbf{x}_i | 1 \leq i \leq n\}$. Each observation $\mathbf{x}_i$ has p dimensions and is represented as $\mathbf{x}_i = [x_{1,i}, x_{2,i}, \cdots, x_{p,i}]^T$. Both the response variable and independent variables are collected sequentially labeled by time points $T = [t_1, t_2, \cdots, t_n]$.

The finite mixture model allows applying cluster analysis on conditionally dependent data into several classes. In the time-series data scenario, researchers cluster the data $((t_1, \mathbf{x}_1, y_1), (t_2, \mathbf{x}_2, y_2), \cdots, (t_n, \mathbf{x}_n, y_n))$ into several homogeneous groups where the number of groups G is unknown in general. Within each group, we apply the SPCA to extract top k principal components that each of them has a sparse loading of p variable coefficients. The extracted top k PCs are denoted as matrix $\mathbf{P}_{p \times k}$. We also use $\mathbf{P}_g$ to represent the principal component matrix obtained from the group indexed by $g = 1, 2, \ldots, G$.

The SPCR model assumes that each pair $(\mathbf{x}_i, y_i)$ is independently drawn from a cluster using both the SPCA and the regression model as follows.

$$y_i = \mathbf{x}_i^T \mathbf{P}_g \beta_g + \epsilon_{i,g}, i = 1, 2, \cdots, n, \tag{5}$$

where $\beta_g = \left[\beta_{g,1}, \beta_{g,2}, \cdots, \beta_{g,k}\right]^T$.

For each group g, the random error is assumed to be Gaussian distributed. That is, $\epsilon_{i,g} \sim \mathcal{N}(0, \sigma_g^2)$. If the response variable is multivariate, then the random error is usually also assumed to be a multivariate Gaussian distribution. Thus the probability density function (pdf) of the SPCR model is

$$f(y_i|\mathbf{x}_i, \mathbf{P}_g, \beta_g) = \mathcal{N}\left(y_i|\mathbf{x}_i^T \mathbf{P}_g \beta_g, \sigma_g^2\right). \tag{6}$$

We emphasize here that the noise (i.e., the error term) included in the statistical model is drawn from a normal distribution independent for each time-series segment, with different values of σ_g^2 for each period. Since we use the EM algorithm to estimate the parameters of the model, the noise parameter σ_g^2 can be estimated accurately as well. Future studies will consider introducing different noise distributions, such as α-stable Lévy noise [25], and other non-Gaussian noise distributions to further extend the current model.

We also consider time factor t_i in the SPCR model of time-series data to be continuous. The pdf of the time factor is

$$f(t_i|v_g, \sigma_g^{2,\mathsf{time}}) = \mathcal{N}\left(t_i|v_g, \sigma_g^{2,\mathsf{time}}\right), \tag{7}$$

where v_g is the mean, and $\sigma_g^{2,\mathsf{time}}$ is the variance of the time segment g. Apart from the normal distribution, our approach can also be generalized to other distributions for the time factor, such as skewed distributions, Student's t-distribution, ARCH, GARCH time-series models, and so on.

As a result, if we use the *MIX-SPCR* model to perform segmentation of time-series data, the likelihood function of the whole data $((t_1, \mathbf{x}_1, y_1), (t_2, \mathbf{x}_2, y_2), \cdots, (t_n, \mathbf{x}_n, y_n))$ with G number of clusters (or segments) is given by

$$L = \prod_{i=1}^{n} \prod_{g=1}^{G} \left[\pi_g f(y_i|\mathbf{x}_i, \mathbf{P}_g, \beta_g) f(t_i|v_g, \sigma_g^{2,\mathsf{time}})\right]^{z_{g,i}}, \tag{8}$$

where the π_g is the mixing proportion with the constraint that $\pi_g \geq 0$ and $\sum_{g=1}^{G} \pi_g = 1$. We follow the definition of missing values by Yang et al. [15] and let $\mathbf{Z} = \{Z_1, Z_2, \cdots, Z_n\}$. If $Z_i = g$, then $z_{g,i} = 1$, otherwise, $z_{g,i} = 0$. Then the log-likelihood function of the *MIX-SPCR* model models is

$$\begin{aligned}
\mathcal{L}_{\mathsf{mix}} &= \log(L) \\
&= \sum_{i=1}^{n} \sum_{g=1}^{G} z_{g,i} \log\left[\pi_g f(y_i|\mathbf{x}_i, \mathbf{P}_g, \beta_g) f(t_i|v_g, \sigma_g^{2,\mathsf{time}})\right] \qquad (9) \\
&= \sum_{i=1}^{n} \sum_{g=1}^{G} z_{g,i} \left[\log\, \pi_g + \log\, f(y_i|\mathbf{x}_i, \mathbf{P}_g, \beta_g) + \log\, f(t_i|v_g, \sigma_g^{2,\mathsf{time}})\right] \\
&= \underbrace{\sum_{i=1}^{n} \sum_{g=1}^{G} z_{g,i} \log\, \pi_g}_{\mathcal{L}_\pi} + \underbrace{\sum_{i=1}^{n} \sum_{g=1}^{G} z_{g,i} \log\, f(y_i|\mathbf{x}_i, \mathbf{P}_g, \beta_g)}_{\mathcal{L}_{\mathsf{SPCR}}} + \underbrace{\sum_{i=1}^{n} \sum_{g=1}^{G} z_{g,i} \log\, f(t_i|v_g, \sigma_g^{2,\mathsf{time}})}_{\mathcal{L}_{\mathsf{time}}}. \qquad (10)
\end{aligned}$$

We denote $\mathbf{z} = \left[z_{g,i}\right]$ where $g = 1, 2, \cdots, G$ and $i = 1, 2, \cdots, n$.

Given the number of segments, researchers usually apply the EM algorithm to determine the optimal segmentation by setting the objective function as $\mathcal{J}_{\mathsf{EM}} = \mathcal{L}_{\mathsf{mix}}$ (Gaffney and Smyth [26], Esling and Agon [27], Gaffney [28]).

3. Regularized Entropy-Based EM Clustering Algorithm

The EM algorithm is a method for iteratively optimizing the objective function. As discussed in Section 2.2, by setting the objective function as the log-likelihood function, we can use the EM algorithm to identify optimal segmentation of time series.

However, in practice, the EM algorithm is sensitive to model initialization conditions and cannot estimate the number of clusters appropriately. To deal with the initialization problem, in 2012, Yang et al. [15] proposed using an entropy penalty to stabilize the computation of each step. The improved method is called the *robust EM algorithm*. In this paper, we extend the robust EM algorithm to deal with time-series data for the *MIX-SPCR* model.

In Section 3.1, we discuss the entropy term of the robust EM algorithm. Then, we show the extension of the robust EM algorithm for the *MIX-SPCR* model in Sections 3.2 and 3.3.

3.1. The Entropy of EM Mixture Probability

As introduced in Equation (8), the π_g represents the mixture probability of each cluster or segment. In other words, the value of π_g is the probability that a data point belongs to group g. The clustering complexity is determined by the number of clusters and corresponding probability values, which could be obtained using entropy. Given $\{\pi_g | 1 \leq g \leq G\}$, the entropy of Z_i is

$$H(Z_i | \{\pi_g | 1 \leq g \leq G\}) = -\sum_{g=1}^{G} \pi_g log(\pi_g), \text{for } i = 1, 2, \cdots, n. \tag{11}$$

Then the entropy of $\mathbf{Z}$ is written as,

$$\begin{aligned} H(\mathbf{Z} | \{\pi_g | 1 \leq g \leq G\}) &= \sum_{i=1}^{n} H(Z_i | \{\pi_g | 1 \leq g \leq G\}) \\ &= -\sum_{i=1}^{n} \sum_{g=1}^{G} \pi_g log(\pi_g) \\ &= -n \sum_{g=1}^{G} \pi_g log(\pi_g). \end{aligned} \tag{12}$$

The objective function of the robust EM algorithm is

$$\mathcal{J}_{\mathsf{Robust\text{-}EM}} = \mathcal{L}_{\mathsf{mix}} - \lambda_{\mathsf{Robust\text{-}EM}} H(\mathbf{Z} | \{\pi_g | 1 \leq g \leq G\}), \tag{13}$$

where $\lambda_{\mathsf{Robust\text{-}EM}} \geq 0$. The log-likelihood term $\mathcal{L}_{\mathsf{mix}}$ is from Equation (9), which gives the goodness-of-fit.

Next, we present the steps of the EM algorithm for maximizing the objective function in Equation (13).

3.2. E-Step (Expectation)

From a Bayesian perspective, we let $\widehat{z}_{g,i}$ denote the posterior probability of the true cluster membership that a dataset triplet $(t_i, \mathbf{x}_i, y_i)$ is drawn from group g. Using the Bayes theorem, we have

$$\widehat{z}_{g,i} = \mathbb{E}(Z_i = g | y_i, \mathbf{x}_i, \mathbf{P}_g, \beta_g) \tag{14}$$

$$= \frac{\pi_g \mathcal{N}\left(y_i; \mathbf{x}_i \mathbf{P}_g \beta_g, \sigma_g^2\right) \mathcal{N}\left(t_i | v_g, \sigma_g^{2,\mathsf{time}}\right)}{\sum_{h=1}^{G} \pi_h \mathcal{N}\left(y_i; \mathbf{x}_i \mathbf{P}_h \beta_h, \sigma_h^2\right) \mathcal{N}\left(t_i | v_h, \sigma_h^{2,\mathsf{time}}\right)}. \tag{15}$$

3.3. M-Step (Maximization)

Using the robustified derivation of $\widehat{\pi}_g$, the estimated mixture proportion, we have

$$\widehat{\pi}_g^{\mathsf{new}} = \widehat{\pi}_g^{\mathsf{EM}} + \widehat{\lambda}_{\mathsf{Robust\text{-}EM}} \widehat{\pi}_g^{\mathsf{old}} \left(log(\widehat{\pi}_g^{\mathsf{old}}) - \sum_{h=1}^{G} \left(\widehat{\pi}_h^{\mathsf{old}} \log(\widehat{\pi}_h^{\mathsf{old}}) \right) \right), \tag{16}$$

where

$$\widehat{\pi}_g^{\mathsf{EM}} = \frac{\sum_{i=1}^{n} \widehat{z}_{g,i}}{n}. \tag{17}$$

We follow the recommendation of Yang et al. [15] for the value of $\widehat{\lambda}_{\mathsf{Robust\text{-}EM}}^{\mathsf{new}}$ as

$$\widehat{\lambda}_{\mathsf{Robust\text{-}EM}}^{\mathsf{new}} = min \left\{ \frac{\sum_{h=1}^{G} exp\left(-\eta n \left| \widehat{\pi}_g^{\mathsf{new}} - \widehat{\pi}_g^{\mathsf{old}} \right| \right)}{G}, \frac{1 - max\left\{ \sum_{i=1}^{n} \widehat{z}_{h,i}^{\mathsf{old}} / n | h = 1, 2, \cdots, G \right\}}{-max\left\{ \widehat{\pi}_h^{\mathsf{old}} | h = 1, 2, \cdots, G \right\} \sum_{h=1}^{G} \widehat{\pi}_h^{\mathsf{old}} \log \widehat{\pi}_h^{\mathsf{old}}} \right\}, \tag{18}$$

where

$$\eta = min\left\{1, 0.5^{\lfloor p/2 - 1 \rfloor}\right\}, \tag{19}$$

and p is the number of variables in the model.

We iterate E-step and M-step several times until convergence to obtain the parameter estimates. In particular, the β_g values get updated by maximizing the $\mathcal{J}_{\mathsf{Robust\text{-}EM}}$ from Equation (13). Since we fix the number of segments and principal components during each E-step and M-step, the updated values of β_g and σ_g can be calculated using $\mathcal{L}_{\mathsf{mix}}$ directly. The estimated values of β_g and σ_g are given as follows.

$$\widehat{\beta}_g^{\mathsf{new}} = \left[\sum_{i=1}^{n} \widehat{z}_{g,i}^{\mathsf{old}} (\mathbf{x}_i^T \mathbf{P}_g)^T (\mathbf{x}_i^T \mathbf{P}_g) \right]^{-1} \sum_{i=1}^{n} \widehat{z}_{g,i}^{\mathsf{old}} (\mathbf{x}_i^T \mathbf{P}_g)^T y_i$$

$$= \left[\sum_{i=1}^{n} \widehat{z}_{g,i}^{\mathsf{old}} \mathbf{P}_g^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{P}_g \right]^{-1} \sum_{i=1}^{n} \widehat{z}_{g,i}^{\mathsf{old}} \mathbf{P}_g^T \mathbf{x}_i y_i, \tag{20}$$

$$\widehat{\sigma}_g^{2,\mathsf{new}} = \sum_{i=1}^{n} \widehat{z}_{g,i}^{\mathsf{old}} \left\| y_i - \mathbf{x}_i^T \mathbf{P}_g \widehat{\beta}_g^{\mathsf{new}} \right\|_2^2 / \sum_{i=1}^{n} \widehat{z}_{g,i}^{\mathsf{old}}. \tag{21}$$

For the time factor, the estimated mean $\widehat{v}_g$ and variance $\widehat{\sigma}_g^{2,\mathsf{time}}$ are

$$\widehat{v}_g = \frac{\sum_{i=1}^{n} \widehat{z}_{g,i} t_i}{\sum_{i=1}^{n} \widehat{z}_{g,i}}, \tag{22}$$

$$\widehat{\sigma}_g^{2,\mathsf{time}} = \frac{\sum_{i=1}^{n} \widehat{z}_{g,i} \left(t_i - \widehat{v}_g\right)^2}{\sum_{i=1}^{n} \widehat{z}_{g,i}}. \tag{23}$$

As discussed above, our approach is flexible in considering other distributional models for the time-series factor, which we will pursue in separate research work.

4. Information Complexity Criteria

Recently, the statistical literature recognized the necessity of introducing model selection as one of the technical areas. In this area, the entropy and the Kullback–Leibler [29] information (or KL distance) play a crucial role and serve as an analytical basis to obtain the forms of model selection criteria. In this paper, we use information criteria to evaluate a portfolio of competing models and select the best-fitting model with minimum criterion values.

One of the first information criteria for model selection in the literature is due to the seminal work of Akaike [30]. Following the entropy maximization principle (EMP), Akaike developed the Akaike's Information Criterion (AIC) to estimate the expected KL distance or divergence. The form of AIC is

$$\mathsf{AIC} = -2 \log L(\hat{\theta}) + 2k, \tag{24}$$

where $L(\hat{\theta})$ is the maximized likelihood function, and k is the number of estimated free parameters in the model. The model with minimum AIC value is chosen as the best model to fit the data.

Motivated by Akaike's work, Bozdogan [16–20,31] developed a new information complexity (ICOMP) criteria based on Van Emden's [32] entropic complexity index in parametric estimation. Instead of penalizing the number of free parameters directly, ICOMP penalizes the covariance complexity of the model. There are several forms of ICOMP. In this section, we present the two general forms of ICOMP criteria based on the estimated inverse Fisher information matrix (IFIM). The first form is

$$\begin{aligned} \mathsf{ICOMP(IFIM)} &= -2 \log L(\hat{\theta}) + 2\mathsf{C}(\hat{\Sigma}_{model}) \\ &= -2 \log L(\hat{\theta}) + 2\mathsf{C}_1(\widehat{\mathcal{F}}^{-1}), \end{aligned} \tag{25}$$

where $L(\hat{\theta})$ is the maximized likelihood function, and $\mathsf{C}_1(\widehat{\mathcal{F}}^{-1})$ represents the entropic complexity of IFIM. We define $\mathsf{C}_1(\widehat{\mathcal{F}}^{-1})$ as

$$\mathsf{C}_1(\widehat{\mathcal{F}}^{-1}) = \frac{s}{2} \log \left(\frac{tr \widehat{\mathcal{F}}^{-1}}{s} \right) - \frac{1}{2} \log \left| \widehat{\mathcal{F}}^{-1} \right|, \tag{26}$$

and where $s = \mathrm{rank}(\widehat{\mathcal{F}}^{-1})$. We can also give the form of $\mathsf{C}_1(\widehat{\mathcal{F}}^{-1})$ in terms of eigenvalues,

$$\mathsf{C}_1(\widehat{\mathcal{F}}^{-1}) = \frac{s}{2} \log \left(\frac{\bar{\lambda}_a}{\bar{\lambda}_g} \right), \tag{27}$$

where $\bar{\lambda}_a$ is the arithmetic mean of the eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_s$, and $\bar{\lambda}_g$ is the geometric mean of the eigenvalues.

We note that ICOMP penalizes the lack of parsimony and the profusion of the model's complexity through IFIM. It offers a new perspective beyond counting and penalizing number of estimated parameters in the model. Instead, ICOMP takes into account interaction (i.e., correlation) among the estimated parameters through the model fitting process.

We define the second form of ICOMP as

$$\text{ICOMP(IFIM)}_{\mathsf{C_{1F}}} = -2\log L(\hat{\theta}) + 2\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1}), \tag{28}$$

where $\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1})$ is given by

$$\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1}) = \frac{s}{4}\frac{\frac{1}{s}tr\left(\left(\widehat{\mathcal{F}}^{-1}\right)^T\left(\widehat{\mathcal{F}}^{-1}\right)\right) - \left(\frac{tr\left(\widehat{\mathcal{F}}^{-1}\right)}{s}\right)^2}{\left(\frac{tr\left(\widehat{\mathcal{F}}^{-1}\right)}{s}\right)^2}. \tag{29}$$

In terms of the eigenvalues of IFIM, we write $\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1})$ as

$$\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1}) = \frac{1}{4\bar{\lambda}_a^2}\sum_{j=1}^{s}\left(\lambda_j - \bar{\lambda}_a\right)^2. \tag{30}$$

We want to highlight some features of $\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1})$ here. The term $\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1})$ is a second-order equivalent measure of complexity to the original term $\mathsf{C_1}(\widehat{\mathcal{F}}^{-1})$. Additionally, we note that $\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1})$ is scale-invariant and $\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1}) \geq 0$ with $\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1}) = 0$ only when all $\lambda_j = \bar{\lambda}_a$. Furthermore, $\mathsf{C_{1F}}(\widehat{\mathcal{F}}^{-1})$ measures the relative variation in the eigenvalues.

These two forms of ICOMP provide us an easy to use computational means in high dimensional modeling. Next, we derive the analytical forms of ICOMP in the *MIX-SPCR* model.

4.1. Derivation of Information Complexity in MIX-SPCR *Model for Time-Series Data*

We first consider the log-likelihood function of the *MIX-SPCR* model given in Equation (9),

$$\mathcal{L}_{\mathsf{mix}} = \mathcal{L}_{\pi} + \mathcal{L}_{\mathsf{SPCR}} + \mathcal{L}_{\mathsf{time}}. \tag{31}$$

After some work, the estimated inverse Fisher information matrix (IFIM) of the mixture probabilities is

$$\widehat{\mathcal{F}}_{\pi}^{-1} = \begin{bmatrix} \widehat{\pi}_1^{-1} & 0 & 0 & 0 \\ 0 & \widehat{\pi}_2^{-1} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \widehat{\pi}_G^{-1} \end{bmatrix}. \tag{32}$$

Similarly, for each segment g, the estimated IFIM, $\widehat{\mathcal{F}}_{g,\mathsf{SPCR}}^{-1}$, is

$$\widehat{\mathcal{F}}_{g,\mathsf{SPCR}}^{-1} = \begin{bmatrix} \hat{\sigma}_g^2\left[\sum_{i=1}^{n}\widehat{z}_{g,i}\left(\mathbf{x}_i^T\mathbf{P}_g\right)^T\left(\mathbf{x}_i^T\mathbf{P}_g\right)\right]^{-1} & \mathbf{0} \\ \mathbf{0}^T & 2\hat{\sigma}_g^4\left(\sum\widehat{z}_{g,i}\right)^{-1} \end{bmatrix}, \; g = 1, 2, \ldots, G. \tag{33}$$

Note that the IFIM should include both the SPCR models $\widehat{\mathcal{F}}^{-1}_{g,\mathsf{SPCR}}$ and the time factor $\widehat{\mathcal{F}}^{-1}_{g,\mathsf{time}}$ for each segment.

For each segment g, the time factor is under the univariate Gaussian distribution. As a result, the IFIM of the time factor is

$$\widehat{\mathcal{F}}^{-1}_{g,\mathsf{time}} = \begin{bmatrix} \widehat{\sigma}^{2,\mathsf{time}}_g/n & 0 \\ 0 & \frac{2}{n}\widehat{\sigma}^{4,\mathsf{time}}_g \end{bmatrix}. \tag{34}$$

By combining the two IFIMs for the SPCR model and the time factor, we have the inverse Fisher information

$$\widehat{\mathcal{F}}^{-1}_{g} = \begin{bmatrix} \widehat{\mathcal{F}}^{-1}_{g,\mathsf{SPCR}} & \mathbf{0} \\ \mathbf{0}^T & \widehat{\mathcal{F}}^{-1}_{g,\mathsf{time}} \end{bmatrix}. \tag{35}$$

Overall, the inverse of the estimated Fisher information matrix (IFIM) for the *MIX-SPCR* model becomes

$$\widehat{\mathcal{F}}^{-1} \cong \begin{bmatrix} \widehat{\mathcal{F}}^{-1}_{\pi} & 0 & 0 & \cdots & 0 \\ 0 & \widehat{\mathcal{F}}^{-1}_{1} & 0 & \cdots & 0 \\ 0 & 0 & \widehat{\mathcal{F}}^{-1}_{2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \widehat{\mathcal{F}}^{-1}_{G} \end{bmatrix}. \tag{36}$$

Using the above definition of ICOMP(IFIM) and the properties of block-diagonal matrices with their trace and determinant, we have

$$\mathsf{ICOMP(IFIM)} = -2\mathcal{L}_{\mathsf{mix}} + 2\mathsf{C}_1(\widehat{\mathcal{F}}^{-1}), \tag{37}$$

where

$$\mathsf{C}_1(\widehat{\mathcal{F}}^{-1}) = \frac{s}{2}\log\left[\frac{tr(\widehat{\mathcal{F}}^{-1}_{\pi}) + \sum_{g=1}^{G} tr(\widehat{\mathcal{F}}^{-1}_{g})}{s}\right] - \frac{1}{2}\left[\log\left|\widehat{\mathcal{F}}^{-1}_{\pi}\right| + \sum_{g=1}^{G}\log\left|\widehat{\mathcal{F}}^{-1}_{g}\right|\right], \tag{38}$$

and where $s = \mathrm{rank}(\widehat{\mathcal{F}}^{-1}) = r_\pi + \sum_{g=1}^{G} r_g = \dim(\widehat{\mathcal{F}}^{-1})$.

Similarly, we derive the second equivalent form of $\mathsf{ICOMP(IFIM)}_{\mathsf{C_{1F}}}$ as

$$\mathsf{ICOMP(IFIM)}_{\mathsf{C_{1F}}} = -2\mathcal{L}_{\mathsf{mix}} + 2\mathsf{C}_{\mathsf{1F}}(\widehat{\mathcal{F}}^{-1}). \tag{39}$$

Using the properties of the block-diagonal matrices, we have

$$tr\left(\left(\widehat{\mathcal{F}}^{-1}\right)^T\left(\widehat{\mathcal{F}}^{-1}\right)\right) = tr\left(\widehat{\mathcal{F}}^{-1}_{\pi}\right)^2 + \sum_{g=1}^{G} tr\left(\widehat{\mathcal{F}}^{-1}_{g}\right)^2. \tag{40}$$

Thus, an open computational form of $\mathsf{ICOMP(IFIM)}_{\mathsf{C_{1F}}}$ becomes

$$\mathsf{ICOMP(IFIM)}_{\mathsf{C_{1F}}} = -2\mathcal{L}_{\mathsf{mix}} + \frac{s}{2}\frac{\frac{1}{s}\left[tr\left(\widehat{\mathcal{F}}_{\pi}^{-1}\right)^2 + \sum_{g=1}^{G} tr\left(\widehat{\mathcal{F}}_{g}^{-1}\right)^2\right] - \left[\frac{tr\left(\widehat{\mathcal{F}}_{\pi}^{-1}\right) + \sum_{g=1}^{G} tr\left(\widehat{\mathcal{F}}_{g}^{-1}\right)}{s}\right]^2}{\left[\frac{tr\left(\widehat{\mathcal{F}}_{\pi}^{-1}\right) + \sum_{g=1}^{G} tr\left(\widehat{\mathcal{F}}_{g}^{-1}\right)}{s}\right]^2}. \tag{41}$$

We note that in computing both forms of ICOMP above, we do not need to build the full inverse of the estimated Fisher information matrix (IFIM) for the *MIX-SPCR* model given in Equation (36). All one requires is the computation of IFIM for each segment, which is appealing.

We also use AIC and CAIC (Bozdogan [33]) for comparison purposes given by

$$\mathsf{AIC} = -2\mathcal{L}_{\mathsf{mix}} + 2s^*, \text{and,} \tag{42}$$

$$\mathsf{CAIC} = -2\mathcal{L}_{\mathsf{mix}} + s^*\left(\log n + 1\right), \tag{43}$$

where $s^* = G(k+3)$ is the number of estimated parameters in the *MIX-SPCR* model and log denotes the natural logarithm of the sample size n.

Next, we show our numerical examples starting with a detailed Monte Carlo simulation study.

5. Monte Carlo Simulation Study

We perform numerical experiments in a unified computing environment: Ubuntu 18.04 operating system, Intel I7-8700, and 32 GB of RAM. We use the programming language Python and the scientific computing package NumPy [34] to build a computational platform. The size of the input data directly affects the running time of the program. At $n = 4000$ time-series observations, the execution time for each EM iteration is about 0.9 s. Parameter estimation can reach convergence within 40 steps of iterations, with a total machine run time of 37 s.

5.1. Simulation Protocol

In this section, we present the performance of the proposed *MIX-SPCR* model using synthetic data generated from a segmented regression model. Our simulation protocol has $p = 12$ variables and four actual latent variables. Two segmented regression models determine the dependent variable y, and each segment is continuous and has its own specified coefficients (β_1 and β_2). Our simulation set up is as follows:

$$\Lambda = \begin{bmatrix} 1.8 & 0 & 0 & 0 \\ 1.8 & 0 & 0 & 0 \\ 1.8 & 0 & 0 & 0 \\ 0 & 1.7 & 0 & 0 \\ 0 & 1.7 & 0 & 0 \\ 0 & 1.7 & 0 & 0 \\ 0 & 0 & 1.6 & 0 \\ 0 & 0 & 1.6 & 0 \\ 0 & 0 & 1.6 & 0 \\ 0 & 0 & 0 & 1.5 \\ 0 & 0 & 0 & 1.5 \\ 0 & 0 & 0 & 1.5 \end{bmatrix}, \tag{44}$$

$$\psi = \text{diag}\left(1.27, 0.61, 0.74, 0.88, 0.65, 0.81, 0.74, 1.3, 1.35, 0.74, 0.92, 1.32\right), \tag{45}$$

$$\Sigma = \Lambda\Lambda^T + \psi, \tag{46}$$

$$\mathbf{x}_t \sim \text{MVN}\left(\mathbf{0}, \Sigma\right), t = 1, 2, \cdots, 4000, \tag{47}$$

$$\beta_1 = (-10, 0.1, 0.1, 0.1, 2.1, 0, 0, 0.1, 0.1, 0, 0, 0), \tag{48}$$

$$\beta_2 = (0, 0, 0, 0, 0, 0.5, 0.3, 0.1, 2.1, 1, 2, 20), \tag{49}$$

$$y_{t,g=1} = \mathbf{x}_{1,t}\beta_1 + \varepsilon_{1,t}, t = 1, 2, \cdots, 2800, \tag{50}$$

$$y_{t,g=2} = \mathbf{x}_{2,t}\beta_2 + \varepsilon_{2,t}, t = 2801, 2802, \cdots, 4000. \tag{51}$$

We set the total number of time-series observations, $n = 4000$. The first segment has $n_1 = 2800$, and the second segment has $n_2 = 1200$ time-series observations. We randomly draw error term from a Gaussian distribution with zero mean and $\sigma^2 = 9$. Among all the variables, the first six observable variables explain the first segment, and the remaining six explanatory variables primarily determine the second segment. We set the mixing proportions $\pi_1 = 0.7$ and $\pi_2 = 0.3$ for two time-series segments, respectively.

5.2. Detection of Structural Change Point

In the first simulation study, we limit the actual number of segments equal to two, which means that the first segment expands from the starting point to a structural change point, and the second segment expands from the change point to the end. By design, each segment is continuous on the time scale, and different sets of independent variables explain the trending and volatility. We run the *MIX-SPCR* model to see if it can successfully determine the position of the change point using the information criteria. If a change point is correctly selected, we expect that the information criteria is minimized at this change point.

Figures 2 and 3 show our results from the *MIX-SPCR* model. Specifically, it shows the sample path of the information criteria at each time point. We note that all the information criteria values are minimized from $t = 2800$ to $t = 3000$, which covers the time-series's actual change point position. As the *MIX-SPCR* model selects different change points, the penalty term of AIC and CAIC remain the same because both the number of model parameters and the number of observations do not change. In this simulation scenario, the fixed penalty term means that the AIC and CAIC reflect the changes only in the "lack of fit" term of various models without considering model complexity. This indicates that using AIC-type criteria just counting and penalizing the number of parameters may be necessary but not sufficient in model selection.

As a comparison, however, we note that the penalty term of information complexity-based criteria, C_1 and C_{1F}, are adjusted in selecting different change points. They are varying but not fixed.

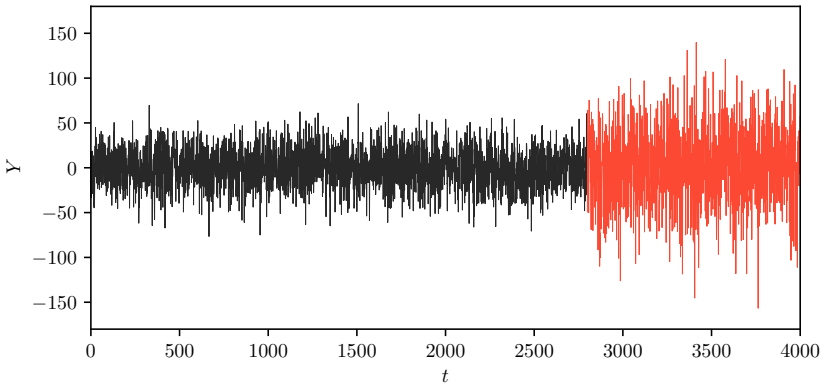

Figure 2. The plot of two-segment simulated time-series data. We show the plot of the simulated time-series data through the whole-time scale. Note that the first segment is from the starting point $t = 1$ to the change point $t = 2800$, and the second time segment expands from the change point $t = 2801$ to the end $t = 4000$.

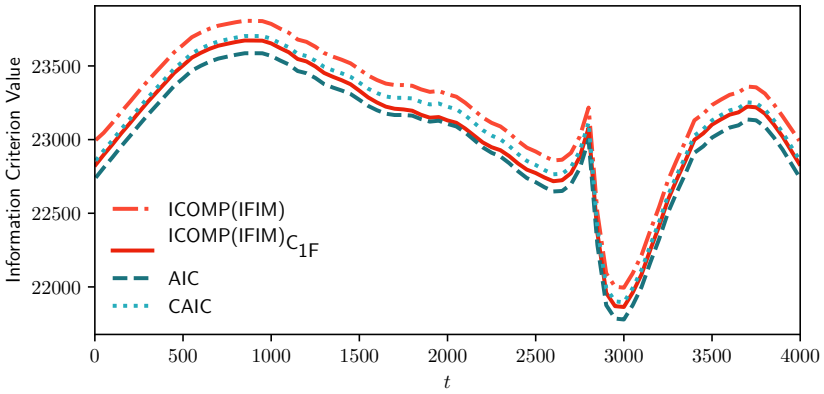

Figure 3. Sample path of information criteria for the simulated time-series data. The horizontal coordinate represents the position of the possible change points, and the vertical coordinate represents the corresponding information criterion (IC) values. The lower the IC value, the more likely the selected position of the change point is the real position. The real change point is $t = 2800$.

5.3. A Large-Scale Monte Carlo Simulation

Next, we perform a large-scale Monte Carlo simulation to illustrate the *MIX-SPCR* model's performance in choosing the correct number of segments and the number of latent variables. A priori, in this simulation, we pretend that we do not know the actual structure of the data and use the information criteria to recover the actual construction of the *MIX-SPCR* model. To achieve this, we follow the above simulation protocol using a different number of time points by varying $n = 1000, 2000, 4000$. As before, there are twelve explanatory variables drawn from four latent variable models generated from a multivariate Gaussian distribution given in Equation (47). The simulated data again consist of two time-series segments with mixing proportions $\pi_1 = 0.7$ and $\pi_2 = 0.3$, respectively. For each data generating process, we replicate the simulation one hundred times and record both information complexity-based criteria (ICOMP(IFIM) & $\text{ICOMP(IFIM)}_{C_{1F}}$) and classic AIC-type criteria (AIC & CAIC).

In Table 1, we present how many times the *MIX-SPCR* model selects different models in the one hundred simulations. In this way, we can assess different information criteria by measuring the hit rates.

Looking at Table 1, we see that when the sample size $n = 1000$ (small), AIC selects the correct model ($G = 2$, $k = 4$) 69 times, CAIC selects 80 times, ICOMP(IFIM) selects 48 times, and ICOMP(IFIM)$_{C_{1F}}$ selects 76 times, respectively, in 100 replications of the Monte Carlo simulation. When the sample size is small, ICOMP(IFIM) tends to choose a sparser regression model sensitive to the sample size. However, as the sample size increases, when $n = 2000$ and $n = 4000$, ICOMP(IFIM) consistently outperforms other information criteria in terms of hit rates. The percentage of the correctly identified model is above 90%, as reported above.

Table 1. Frequency of the choice of the true model with information criteria in 100 replications of the experiment for each sample size (n) of time-series observations. The true model is $G = 2$ and $k = 4$.

		$n = 1000$		$n = 2000$		$n = 4000$	
		G = 2	**G = 3**	**G = 2**	**G = 3**	**G = 2**	**G = 3**
AIC	k = 2	0	0	0	0	0	0
	k = 3	0	6	0	3	0	1
	k = 4	69	0	77	0	75	0
	k = 5	24	1	20	0	24	0
CAIC	k = 2	1	0	0	0	0	0
	k = 3	1	3	0	1	0	1
	k = 4	80	0	96	0	93	0
	k = 5	14	1	3	0	6	0
ICOMP(IFIM)	k = 2	31	2	1	0	0	0
	k = 3	2	5	0	2	0	1
	k = 4	48	0	96	0	96	0
	k = 5	11	1	1	0	3	0
ICOMP(IFIM)$_{C_{1F}}$	k = 2	2	1	0	0	0	0
	k = 3	0	7	0	3	0	1
	k = 4	76	0	93	0	93	0
	k = 5	13	1	4	0	6	0

Our results show that the *MIX-SPCR* model works well in all settings to estimate the number of time-series segments and the number of latent variables.

Figure 4 illustrates how the *MIX-SPCR* model performs if the number of segments and the number of sparse principal components are unknown beforehand.

The choice of the number of segments (G) has a significant impact on the results. For all the simulation scenarios, the correct choice of the number of segments ($G = 2$) has information criterion values less than the incorrect choice ($G = 3$). This pattern emerges consistently among all the sample sizes, both the classical ones and information-complexity based criteria.

In summary, the large-scale Monte Carlo simulation analysis highlights the performance of the *MIX-SPCR* model. As the sample size increases, the *MIX-SPCR* model improves its performance. As shown in Figure 3, the *MIX-SPCR* model can efficiently determine the structural change point and estimate the mixture proportions when the number of segments is unknown beforehand. Another key finding is that, by using the appropriate information criteria, the *MIX-SPCR* model can correctly identify the number of segments and the number of latent variables from the data. In other words, our approach can extract the main factors not only from the intercorrelated variables but also classify the data into several clearly defined segments on the time scale.

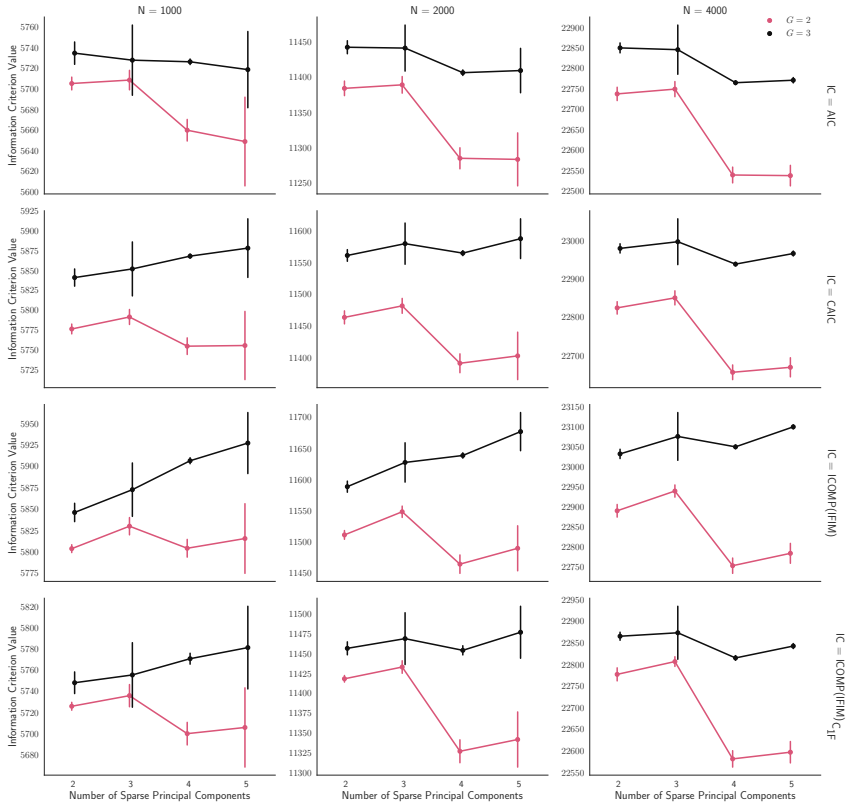

Figure 4. Plot of average and 1SD (standard deviation) of information criterion values over different sample sizes in all simulations with three Sparse Principal Components (SPCs) and $G = 2$ segments. The red line indicates the estimated *MIX-SPCR* model based on two groups ($G = 2$). Correspondingly, the black line indicates the estimated *MIX-SPCR* model for three groups ($G = 3$). Horizontal coordinates represent different numbers of SPCs.

6. Case Study: Segmentation of the S&P 500 Index

6.1. Description of Data

The financial market often generates a large amount of time-series data, and in most cases, the generated data is high-dimensional. In this paper, we use the S&P 500 index and its related hundreds of company stocks categorized into eleven sectors, which are high dimensional time-series data. The index value is the response variable mixed by plenty of companies' variations at each time point. These long time-series values often consist of different regimes and states. For example, the stock market experienced a boom period from 2017 to 2019, which is a dramatic change compared with the stock market during the 2008 financial crisis. If we analyze each sector or company, some industries perform more actively than others during a particular period.

In this section, we implement the *MIX-SPCR* model on the adjusted closing price of the S&P 500 (^GSPC) as a case study. We extract the daily adjusted closing prices from the Yahoo Finance database (https://finance.yahoo.com/) that spans the period from 1 January 1999 to 31 December 2019. By removing weekends and holidays, there are $n = 5292$ tradable days in total. The main focus of this section is to split the time-series into several self-contained segments. Besides, we expect the extracted sparse principal components to explain the variance and volatility in each segment.

6.2. Computational Results

To have a big picture of how the S&P 500 index values reflect the changes of 506 company stock prices, Figure 5 shows the plot of the normalized values of adjusted closing prices. We use the *MIX-SPCR* model with the information criteria to determine the number of segments and the number of sparse principal components. To achieve interpretable results, we limit our search space to a maximum of seven time-series and six sparse principal components. Table 2 shows the optimal combination of three self-contained segments and three sparse principal components for each of the segments by using the information complexity ICOMP(IFIM). The other three information criteria also choose this combination as the best-fitting model. Figure 6 illustrates the probability and time range of each segment. We can see that the first segment is from 1 January 1999, to 26 October 2007. The second time-series segment spans from 29 October 2007, to the end of 2016. The last segment extends from 30 December 2016 to 31 December 2019.

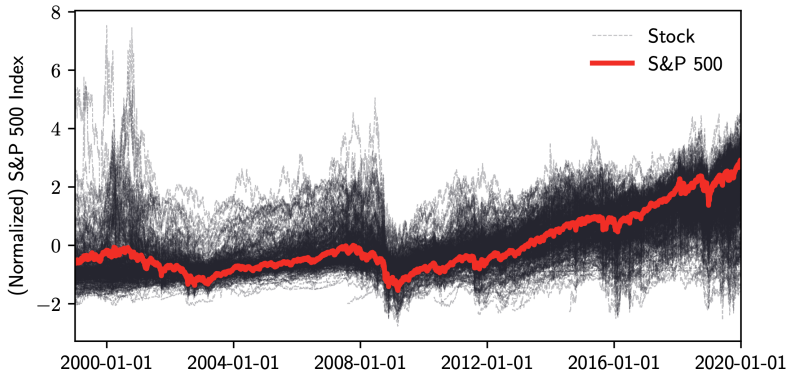

Figure 5. Normalized S&P 500 index and stock prices from January 1999 to December 2019.

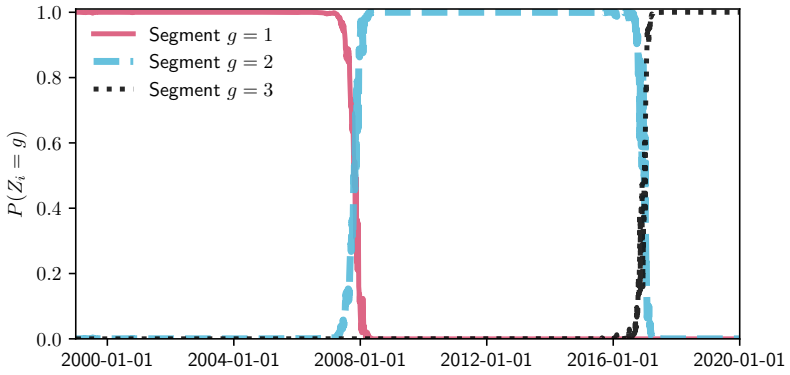

Figure 6. Segmented periods and probability. The plot's vertical coordinate indicates the probability that an individual time-series data point belongs to each segment.

Table 2. The ICOMP(IFIM) values of segmentation results for S&P 500 index data (Lower is better).

		Number of Sparse Principal Components				
		1	**2**	**3**	**4**	**5**
Number of Segments	1	30,097.04	30,092.45	30,106.50	30,121.64	30,145.13
	2	29,975.01	30,058.40	30,293.55	30,234.65	30,347.94
	3	30,010.70	30,062.19	29,241.52	30,453.74	30,526.20
	4	29,877.27	29,825.73	29,811.53	30,571.39	30,628.61
	5	29,904.35	29,973.47	30,011.18	30,311.52	30,554.82
	6	30,111.35	30,361.39	30,388.47	30,665.26	30,581.29
	7	30,031.39	30,564.65	30,597.14	30,823.76	31,057.54

We emphasize that many factors may explain the stock market variation, and this is not a research on how the socioeconomic events influence the S&P 500 index. However, it does raise our interest in the distribution of two structural change points from the segmentation results. The first change point is October 2007, which is the early stage of the 2008 financial crisis. The second structural change point is December 2016, the transitional period of the USA presidential election. Identification of these two change points shows that our proposed method can detect the underlying physical and structural change from the available time-series data.

Table 3 lists the estimated coefficients (β_g) from sparse principal component regression. Because all the collected stock prices and S&P 500 index values are standardized before implementing the *MIX-SPCR* model, we make dimension reduction, remove the constant term, and perform regression analysis using the SPCR model. The R^2 values are above 0.8 across all three different time segments.

Table 3. SPCR coefficients (β_g) of three different segments.

	Segment 1 ($R^2 = 0.82$) **01-01-1999 $\sim$ 26-10-2007**	**Segment 2 ($R^2 = 0.94$)** **27-10-2007 $\sim$ 29-12-2016**	**Segment 3 ($R^2 = 0.97$)** **30-12-2016 $\sim$ 31-12-2019**
SPC1	0.0964	0.1240	0.1512
SPC2	0.0729	−0.0439	0.0359
SPC3	0.0079	0.0191	−0.0051

6.3. Interpretation of Computational Results

One may ask a question, "Can the *MIX-SPCR* model identify the key variables from the hundreds of companies?" If the constructed model is dense, the selected companies would include all the sectors whereby the dense model is limiting the interpretation of the data. Our analysis identifies all the companies with non-zero coefficient values and maps them back to each of the sectors in Tables A1–A3. Each calculated sparse principal component vector consists of around fifty companies, much less than the original data dimension ($p = 506$). We observe that these selected companies are grouped into a few sectors within different time segments. For example, energy companies load in the first sparse principal component vector from 1999 to 2007 (segment 1) and diminish after that.

To have a detailed analysis of how different sectors perform across three segments, we do the stem plot to show the sparse principal component coefficients $\mathbf{P}_g$ of four sectors, namely financials, real estate, energy, and information technology (IT). Figures 7–8 indicate a similar behavior that happened in financial and real estate companies. Both sectors play an essential role in the first two time-series segments but have no contribution in the third segment, which is the period after December 2016. Notice that in Figure 9, energy companies act as an essential player before 2016. However, during the recession in 2008, energy company loadings are negated from the first SPC to the second SPC. Compared with other industries, the variation in energy company stock prices does not contribute to the S&P 500 index after 2016.

Another question is "What sector/industry is the main contributing factor after the 2016 United States presidential election?" A possible answer is, as shown in Figure 10, the SPC coefficients of information technology companies. From 1999 to the recession in 2008, IT companies work mainly on the second SPC and the third SPC, which do not contribute much to the main variation. After the recession, the variations of IT companies do not contribute compared with other sectors. However, after December 2016, companies from the IT industry play an essential role in the primary stock price volatility.

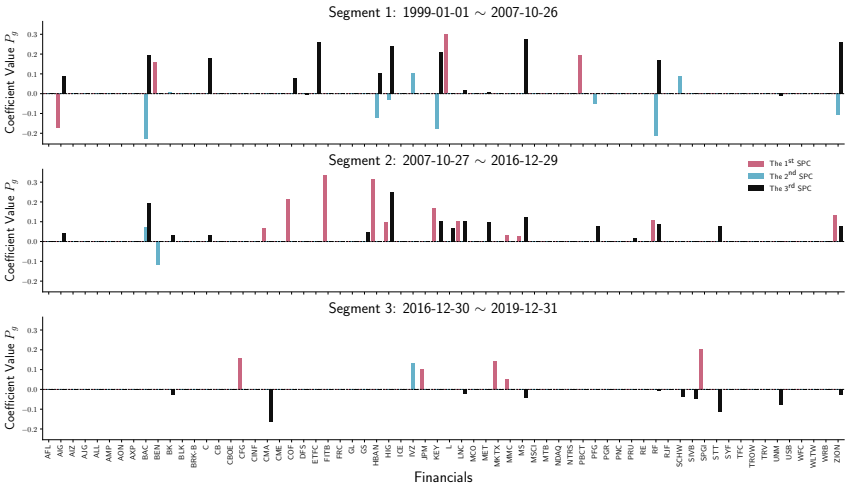

Figure 7. Stem plot of SPC coefficients $\mathbf{P}_g$ for financial companies within each time segment. From top to bottom, the three panels represent different segmented periods, respectively. The horizontal axis of each panel indicates the company in the industrial sector. The vertical axis shows the SPC coefficient values.

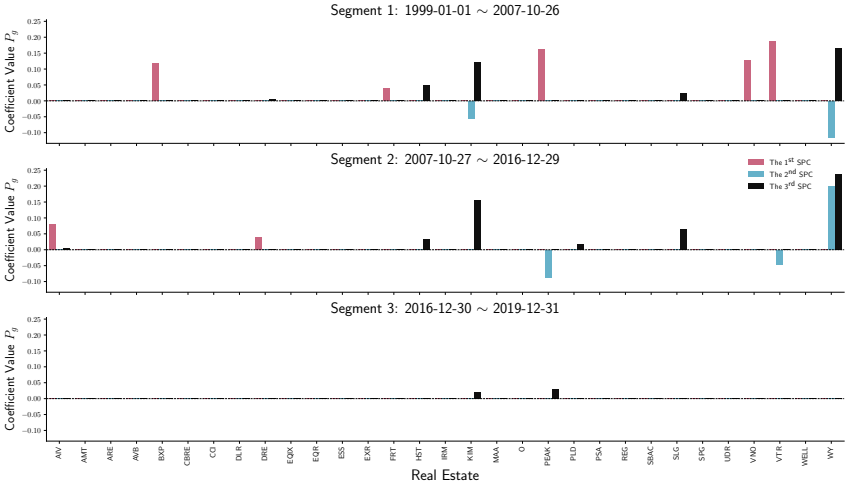

Figure 8. Stem plot of SPC coefficients $\mathbf{P}_g$ for real estate companies within each time segment. From top to bottom, the three panels represent different segmented periods, respectively. The horizontal axis of each panel indicates the company in the industrial sector. The vertical axis shows the SPC coefficient values.

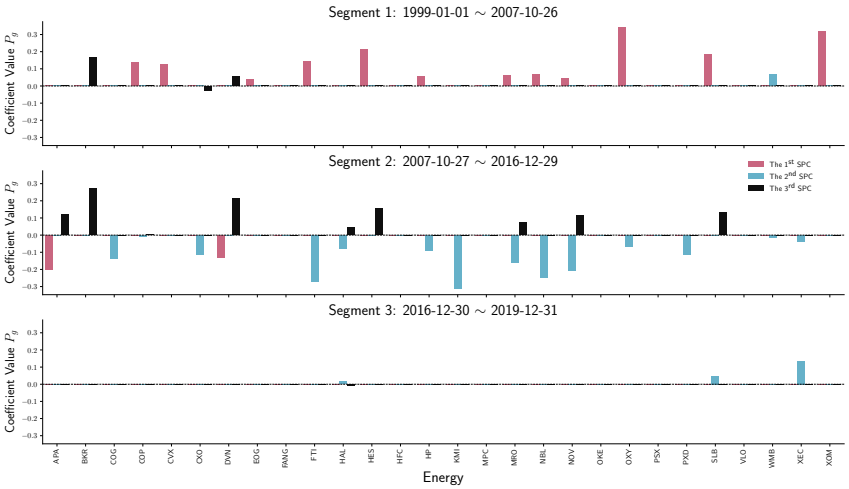

Figure 9. Stem plot of SPC coefficients $\mathbf{P}_g$ for energy companies within each time segment. From top to bottom, the three panels represent different segmented periods, respectively. The horizontal axis of each panel indicates the company in the industrial sector. The vertical axis shows the SPC coefficient values.

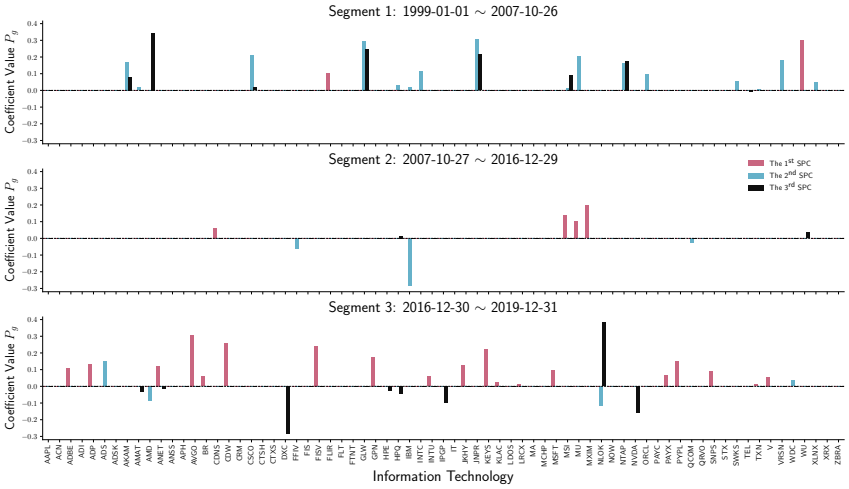

Figure 10. Stem plot of SPC coefficients $\mathbf{P}_g$ for information technology companies within each time segment. From top to bottom, the three panels represent different segmented periods, respectively. The horizontal axis of each panel indicates the company in the industrial sector. The vertical axis shows the SPC coefficient values.

As discussed above, Figures 7–10 provide a clear picture of how different sectors perform (via coefficient $\mathbf{P}_g$) without considering the effects on the S&P 500 index. It might raise the interest in how the SPCR coefficient $\mathbf{P}_g\beta_g$ changes before/after certain socioeconomic events. We follow the research implemented by Aït-Sahalia and Xiu [35] about how the Federal Reserve addressing heightened liquidity from March 10 to 14 March 2008, affects the stock market. The data analyzed by Aït-Sahalia and Xiu [35] are the S&P 100 index values using the traditional PCA, and the authors grouped stocks into financial and non-financial categories. Instead of PCA, we apply the SPCR model on the S&P 500 index and analyze how eleven sectors react before/after Federal Reserve operations. Figure 11 shows that financials, consumer discretionary, real estate, and industrials experienced more significant perturbations than other sectors in terms of SPCR coefficients $\mathbf{P}_g\beta_g$. This conclusion is consistent with the results from Aït-Sahalia and Xiu [35] that the average loadings of first and second principal components of financial companies

are distinct from non-financial companies. However, considering that we have 506 companies in the raw data and make a sparse loading of companies for comparison, the excessive explanatory power is still maintained in this high-dimensional case using the SPCR model, which is more interpretable.

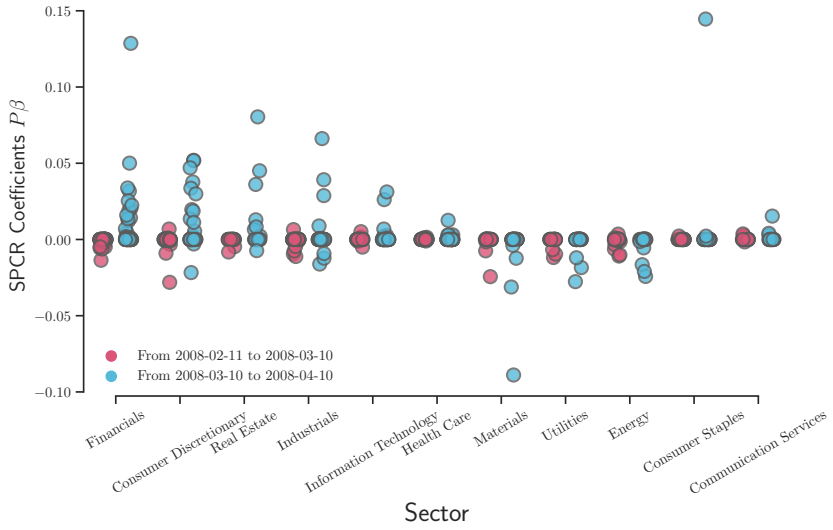

Figure 11. Overlay plot of the SPCR coefficients before/after 2008 financial crisis.

7. Conclusions and Discussions

In this paper, we presented a new and novel method to segment high-dimensional time-series data into different clusters or segments using the mixture model of the sparse principal components model (*MIX-SPCR*). The *MIX-SPCR* model considers both the relationships among the predictor variables and how various predictor variables contribute the explanatory power to the response variable through the sparsity settings. Information criteria have been introduced and derived for the *MIX-SPCR* model. These criteria are applied to study their performance under different sample sizes and to select the best-fitting model.

Our large-scale Monte Carlo simulation exercise showed that the *MIX-SPCR* model could successfully identify the real structure of the time-series data using the information criteria as the fitness function. In particular, based on our results, the information complexity-based criteria—i.e., ICOMP(IFIM) and $\text{ICOMP(IFIM)}_{\text{C1F}}$—outperformed the conventional standard information criteria, such as the AIC-type criteria as the data dimension and the sample size increase.

Later, we empirically applied the *MIX-SPCR* model to uncover the S&P 500 index data (from 1999 to 2019) and identify two change points of this data set.

We observe that the first change point physically coincides with the early stages of the 2008 financial crisis. The second change point is immediately after the 2016 United States presidential election. This structural change point coincides with the election of President Trump and his transition.

Our findings showed how the S&P 500 index and company stock prices react within each time-series segment. The *MIX-SPCR* model presents excessive explanatory power by identifying how different sectors fluctuated before/after the Federal Reserve's addressing heightened liquidity from 10 March to 14 March 2008.

Although this is not a traditional event study paper, it is the first paper to use the sparse principal component regression model with mixture models in the time-series analysis. The proposed new and novel *MIX-SPCR* model enlightens us to explore more interpretable results on how macroeconomic

factors/events influence the stock prices on the time scale. Later, in a separate paper, we will incorporate the event study in the *MIX-SPCR* model as our future research initiative.

This paper's time segmentation model builds on time-series data, constructs likelihood functions, and performs parameter estimation by introducing error information unique to each period. Researchers have recently realized that environmental background noise can positively affect the model building and analysis under certain circumstances ([36–42]). For example, in Azpeitia and Wagner [40], the authors highlighted that the introduction of noise is necessary to obtain information about the system. In our next study, we would like to explore this positive effect of environmental noise even further and use it to build better statistical models for analyzing high-dimensional time-series data.

Author Contributions: Conceptualization, H.B. and Y.S.; methodology, H.B. and Y.S.; software, Y.S.; validation, H.B. and Y.S.; formal analysis, H.B. and Y.S.; investigation, H.B. and Y.S.; resources, H.B. and Y.S.; data curation, Y.S.; writing–original draft preparation, H.B. and Y.S.; writing–review and editing, H.B. and Y.S.; visualization, H.B. and Y.S.; supervision, H.B.; project administration, H.B. and Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The first author expresses his gratitude to Bozdogan in bringing this challenging problem to his attention as part of his doctoral thesis chapter and spending valuable time with him that resulted in this joint work. We also express our thanks to Ejaz Ahmed for inviting us to make a contribution to the Special Issue of Entropy. We extend our thanks and gratitude to anonymous reviewers. Their constructive comments further improved the paper.

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

Abbreviations

The following abbreviations are used in this manuscript:

MIX-SPCR	Mixture of the sparse principal component regression model
CV	Cross validation
TICC	Toeplitz inverse covariance-based clustering
GGS	Greedy Gaussian Segmentation
PCA	Principal Component Analysis
PC	Principal Component
SPCA	Sparse Principal Component Analysis
SPCR	Sparse Principal Component Regression
SPC	Sparse Principal Component
SPPCA	Sparse Probabilistic Principal Component Analysis
EM	Expectation–Maximization (Algorithm)
IC	Information Criterion
ICOMP	Information Complexity

Appendix A. Tables

Table A1. Sparse Principal Component (SPC) of Segment 1 (1 January 1999 ∼ 26 October 2007).

	SPC1		SPC2		SPC3	
	Count	**Percentage**	**Count**	**Percentage**	**Count**	**Percentage**
Health Care	4	6.56	6	9.84	1	1.64
Industrials	6	8.57	3	4.29	6	8.57
Utilities	5	17.86	3	10.71	3	10.71
Materials	2	7.14	1	3.57	1	3.57
Consumer Discretionary	5	7.81	6	9.38	6	9.38
Energy	13	46.43	1	3.57	3	10.71
Financials	5	7.58	10	15.15	15	22.73
Real Estate	5	16.13	2	6.45	5	16.13
Consumer Staples	2	6.06	0	0.00	1	3.03
Communication Services	1	3.85	2	7.69	1	3.85
Information Technology	2	2.82	16	22.54	8	11.27

Table A2. Sparse Principal Component (SPC) of Segment 2 (27 October 2007 ∼ 29 Decmeber 2016).

	SPC1		SPC2		SPC3	
	Count	**Percentage**	**Count**	**Percentage**	**Count**	**Percentage**
Health Care	7	11.48	2	3.28	1	1.64
Industrials	5	7.14	6	8.57	4	5.71
Utilities	0	0.00	0	0.00	5	17.86
Materials	6	21.43	2	7.14	3	10.71
Consumer Discretionary	7	10.94	14	21.88	3	4.69
Energy	2	7.14	14	50.00	9	32.14
Financials	12	18.18	3	4.55	16	24.24
Real Estate	2	6.45	3	9.68	6	19.35
Consumer Staples	0	0.00	0	0.00	0	0.00
Communication Services	5	19.23	3	11.54	1	3.85
Information Technology	4	5.63	3	4.23	2	2.82

Table A3. Sparse Principal Component (SPC) of Segment 3 (30 Decmeber 2016 ∼ 31 Decmeber 2019).

	SPC1		SPC2		SPC3	
	Count	**Percentage**	**Count**	**Percentage**	**Count**	**Percentage**
Health Care	10	16.39	14	22.95	2	3.28
Industrials	9	12.86	4	5.71	4	5.71
Utilities	1	3.57	0	0.00	0	0.00
Materials	1	3.57	3	10.71	6	21.43
Consumer Discretionary	3	4.69	10	15.63	8	12.50
Energy	0	0.00	3	10.71	1	3.57
Financials	5	7.58	1	1.52	10	15.15
Real Estate	0	0.00	0	0.00	2	6.45
Consumer Staples	0	0.00	6	18.18	3	9.09
Communication Services	2	7.69	5	19.23	5	19.23
Information Technology	19	26.76	4	5.63	9	12.68

References

1. Barber, D.; Cemgil, A.T.; Chiappa, S. *Bayesian Time Series Models*; Cambridge University Press: Cambridge, UK, 2011.
2. Abonyi, J.; Feil, B. *Cluster Analysis for Data Mining and System Identification*; Springer Science & Business Media: New York, NY, USA, 2007.
3. Spagnolo, B.; Valenti, D. Volatility effects on the escape time in financial market models. *Int. J. Bifurc. Chaos* **2008**, *18*, 2775–2786. [CrossRef]
4. Valenti, D.; Fazio, G.; Spagnolo, B. Stabilizing effect of volatility in financial markets. *Phys. Rev. E* **2018**, *97*, 062307. [CrossRef] [PubMed]
5. S Lima, L. Nonlinear Stochastic Equation within an Itô Prescription for Modelling of Financial Market. *Entropy* **2019**, *21*, 530. [CrossRef]
6. Ding, W.; Wang, B.; Xing, Y.; Li, J.C. Correlation noise and delay time enhanced stability of electricity futures market. *Mod. Phys. Lett. B* **2019**, *33*, 1950375. [CrossRef]
7. Dillon, W.R.; Böckenholt, U.; De Borrero, M.S.; Bozdogan, H.; De Sarbo, W.; Gupta, S.; Kamakura, W.; Kumar, A.; Ramaswamy, B.; Zenor, M. Issues in the estimation and application of latent structure models of choice. *Mark. Lett.* **1994**, *5*, 323–334. [CrossRef]
8. Quandt, R.E.; Ramsey, J. Estimating Mixtures of Normal Distributions and Switching Regressions. *J. Am. Stat. Assoc.* **1978**, *73*, 730–738.
9. Kiefer, N.M. Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica* **1978**, *46*, 427–434.
10. De Veaux, R.D. Parameter Estimation for a Mixture of Linear Regressions. Ph.D. Thesis, Department of Statistics, Stanford University, Stanford, CA, USA, 1986. Tech. Rept. No. 247.
11. DeSarbo, W.S.; Cron, W.L. A maximum likelihood methodology for clusterwise linear regression. *J. Classif.* **1988**, *5*, 249–282. [CrossRef]
12. Wedel, M.; DeSarbo, W.S. A Review of Recent Developments in Latent Class Regression Models; In *Advanced Methods of Marketing Research*; Bagozzi, R., Ed.; Blackwell Pub.: Hoboken, NJ, USA, 1994; pp. 352–388.
13. Sclove, S.L. Time-series segmentation: A model and a method. *Inf. Sci.* **1983**, *29*, 7–25. [CrossRef]
14. Zou, H.; Hastie, T.; Tibshirani, R. Sparse principal component analysis. *J. Comput. Graph. Stat.* **2006**, *15*, 265–286. [CrossRef]
15. Yang, M.S.; Lai, C.Y.; Lin, C.Y. A robust EM clustering algorithm for Gaussian mixture models. *Pattern Recognit.* **2012**, *45*, 3950–3961. [CrossRef]
16. Bozdogan, H. On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Commun. Stat. Theory Methods* **1990**, *19*, 221–278. [CrossRef]
17. Bozdogan, H. Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-Fisher information matrix. In *Information and Classification*; Springer: New York, NY, USA, 1993; pp. 40–54.
18. Bozdogan, H. Choosing the number of clusters, subset selection of variables, and outlier detection in the standard mixture-model cluster analysis. In *New approaches in Classification and Data Analysis*; Springer: New York, NY, USA, 1994; pp. 169–177.
19. Bozdogan, H. Mixture-model cluster analysis using model selection criteria and a new informational measure of complexity. In *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach*; Springer: New York, NY, USA, 1994; pp. 69–113.
20. Bozdogan, H. A new class of information complexity (ICOMP) criteria with an application to customer profiling and segmentation. *İstanbul Üniversitesi İşletme Fakültesi Derg.* **2010**, *39*, 370–398.
21. Shen, H.; Huang, J.Z. Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivar. Anal.* **2008**, *99*, 1015–1034. [CrossRef]
22. Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G. Online dictionary learning for sparse coding. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 689–696.

23. Guan, Y.; Dy, J. Sparse probabilistic principal component analysis. In Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics, Clearwater, FL, USA, 16–19 April 2009; pp. 185–192.
24. Williams, P.M. Bayesian regularization and pruning using a Laplace prior. *Neural Comput.* **1995**, *7*, 117–143. [CrossRef]
25. Guarcello, C.; Valenti, D.; Spagnolo, B.; Pierro, V.; Filatrella, G. Josephson-based threshold detector for Lévy-distributed current fluctuations. *Phys. Rev. Appl.* **2019**, *11*, 044078. [CrossRef]
26. Gaffney, S.; Smyth, P. Trajectory clustering with mixtures of regression models. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 15–18 August 1999; pp. 63–72.
27. Esling, P.; Agon, C. Time-series data mining. *ACM Comput. Surv. (CSUR)* **2012**, *45*, 1–34. [CrossRef]
28. Gaffney, S. Probabilistic Curve-Aligned Clustering and Prediction with Regression Mixture Models. Ph.D. Thesis, University of California, Irvine, CA, USA, 2004.
29. Kullback, A.; Leibler, R. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
30. Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. In *Second International Symposium on Information Theory*; Petrox, B., Csaki, F., Eds.; Academiai Kiado: Budapest, Hungary, 1973; pp. 267–281.
31. Bozdogan, H. Akaike's Information Criterion and Recent Developments in Information Complexity. *J. Math. Psychol.* **2000**, *44*, 62–91. [CrossRef] [PubMed]
32. Van Emden, H.M. An analysis of complexity. In *Mathematical Centre Tracts*; Mathematisch Centrum: Amsterdam, The Netherlands, 1971.
33. Bozdogan, H. Model Selection and Akaike's Information Criteria (AIC): The General Theory and its Analytical Extensions. *Psychometrica* **1987**, *52*, 317–332. [CrossRef]
34. van der Walt, S.; Colbert, S.C.; Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30. [CrossRef]
35. Aït-Sahalia, Y.; Xiu, D. Principal component analysis of high-frequency data. *J. Am. Stat. Assoc.* **2019**, *114*, 287–303. [CrossRef]
36. Spagnolo, B.; Valenti, D.; Guarcello, C.; Carollo, A.; Adorno, D.P.; Spezia, S.; Pizzolato, N.; Di Paola, B. Noise-induced effects in nonlinear relaxation of condensed matter systems. *Chaos Solitons Fractals* **2015**, *81*, 412–424. [CrossRef]
37. Valenti, D.; Magazzù, L.; Caldara, P.; Spagnolo, B. Stabilization of quantum metastable states by dissipation. *Phys. Rev. B* **2015**, *91*, 235412. [CrossRef]
38. Spagnolo, B.; Guarcello, C.; Magazzù, L.; Carollo, A.; Persano Adorno, D.; Valenti, D. Nonlinear relaxation phenomena in metastable condensed matter systems. *Entropy* **2017**, *19*, 20. [CrossRef]
39. Serdukova, L.; Zheng, Y.; Duan, J.; Kurths, J. Stochastic basins of attraction for metastable states. *Chaos Interdiscip. J. Nonlinear Sci.* **2016**, *26*, 073117. [CrossRef] [PubMed]
40. Azpeitia, E.; Wagner, A. The positive role of noise for information acquisition in biological signaling pathways. *bioRxiv* **2019**, *2019*, 762989.
41. Addesso, P.; Filatrella, G.; Pierro, V. Characterization of escape times of Josephson junctions for signal detection. *Phys. Rev. E* **2012**, *85*, 016708. [CrossRef]
42. Li, J.h.; Łuczka, J. Thermal-inertial ratchet effects: Negative mobility, resonant activation, noise-enhanced stability, and noise-weakened stability. *Phys. Rev. E* **2010**, *82*, 041104. [CrossRef]