# Structural Change Point Detection
# Using A Large Random Matrix and Sparse Modeling

Katsuya Ito*
Graduate School of Economics
The University of Tokyo
Bunkyo, Tokyo, Japan
katsuya1ito@gmail.com

Akira Kinoshita
Preferred Networks, Inc.
Chiyoda, Tokyo, Japan
kino@preferred.jp

Masashi Yoshikawa
Preferred Networks, Inc.
Chiyoda, Tokyo, Japan
yoshikawa@preferred.jp

## ABSTRACT

This paper proposes a new structural change point detection method for time series using our new matrix decomposition method. We propose YMN decomposition which decomposes one time-series data matrix into a large random matrix and a sparse matrix. YMN decomposition can obtain the information about the latent structure from the sparse matrix. Thus our structural change point detection method using YMN decomposition can detect the change in higher-order moments of a mixing matrix as well as typical structural changes such as the change in mean, variance, and autocorrelation of time series. We also partly theorize our methods using existing theories of random matrix and statistics. Our experiment using artificial data demonstrates the effectiveness of our change point detection techniques and our experiment using real data demonstrates that our methods can detect structural changes in economics and finance.

## 1 INTRODUCTION

Unveiling the model of time-series data is necessary to explain the causal relationships and to forecast the future. In economic and financial modeling, the models of time-series data are not apparent in many cases; moreover, their models are time varying [2, 28, 32, 34]. Under such obscurity and instability of the models, explaining the causal relationships among data and forecasting the future are difficult but central problems in these fields [14, 16]. One straightforward method to overcome the instability is using *change-point detection* methods. Change-point detection methods divide piece-wise stable time-series data into several stable time-series data [38]. Moreover, detecting the change point of the *structure* is an essential task in time-series modeling in economics and finance [2]. Research with a similar purpose can be found in many fields such as biology [8], neuroscience [40], and computer network analysis [1].

Structural change point detection is difficult because we do not know the structure of the time-series before/after the change point, and how the change occurs. However, factor modeling is typically used to unveil the partial structures in economics [6] and finance [33]. Economic or financial data such as GDPs and stock prices are observed as high-dimensional time series data. Assume that we observe a vector $y_t$ at time $t$. A factor model assumes that a few typical "factors" generate the observed data. It formally introduces two latent variables $A$ and $z_t$ and assumes that the linear equation holds:

$$y_t = Az_t.$$

Table 1: Examples of observed data $y$, latent data $z$ and mixing matrix $A$ in several research fields

| Field | $y_t$ (observed) | $z_t$ (latent) | $A$ (structure) |
|---|---|---|---|
| Robotics | Sensor data | Motion of parts | Structure of data |
| Neuro-science | EEG/fMRI | Waves from parts | Structure of brain |
| Network | Traffic data | Packets | Network Structure |
| Economics | Macro data | Factors | Economic Structure |
| Finance | Stock prices | Factors | Structure of Market |

$z_t$ is a low-dimensional time-variant vector of factors, while $A$ is a time-invariant coefficient matrix. $A$ is called "factor loadings" in the context of econometrics and "mixing matrix" in the context of computer science. $A$ implies the latent "structure" behind the observed data. However, the structure may be changed abruptly in reality because of major economic or political events such as Brexit or the Global Financial Crisis. Most of the work performed by economists and quantitative financial analysts, such as the prediction and explanation of time-series, strongly rely on their models [19, 37]. It is, therefore, crucial for them to detect changes in their structures. As summarized in Table 1, the structures (i.e. the relationships between observed and latent data) are important, and similar types of problems occur in many fields.

The change exhibits several patterns. The simplest one is the change in the mean and variance of $A$'s entities; it can be detected by merely calculating the mean and variance of $y$. The change in the numbers of factors and the change in the autocorrelation of $y$ are also the essential but straightforward ones [9, 12, 20]. The third- and fourth-order moments are also significant in the literature of finance and econometrics, which attach special importance to them [15]. However, they are difficult to be detected because in high-dimensional cases, they approach the normal distribution by the central limit theorem[1] and the change cannot be observed in $y$. To the best of our knowledge, nonparametric methods that detect changes of such skewness and kurtosis do not exist, whereas many parametric methods do exist. [4, 30, 31].

Many previous studies showed the effectiveness of linear models in detecting the structural changes of high-dimensional data. Well-known methods such as principal component analysis (PCA), non-negative matrix factorization (NMF), and independent component analysis (ICA) can be interpreted as methods to

---

*This work was performed when the author was an intern at Preferred Networks, Inc.

[1]This interpretation of the central limit theorem is often used in ICA literature [23]. In this literature, skew and kurtosis are interpreted as *non-Gaussianity* and by using many samples, the non-Gaussianity of data will decrease by the central limit theorem.

decompose one data matrix into two matrices that satisfy specific properties. Moreover, a variant of these methods that specializes in change point detection is developed such as OMWRPCA [41], ILRMA [24], and TVICA [13].

We herein propose a new non-parametric change point detection method of the structures of multidimensional time series. Our method decomposes the observed data matrix into a large random matrix and a sparse coefficient matrix. First, we use randomly generated large matrices as latent time series. However, when a large random time series row vectors is generated, most of its row vectors are not related to the latent time series, and some are close to the latent time series. Therefore, we select good row vectors by applying a sparse matrix to this. Finally, we can detect the structural changes by calculating the difference of the sparse matrix. Some previous works reported similar frameworks such as latent feature Lasso [42] or extreme learning machine [22]; however, studies using both the random matrix and the sparse matrix have not been reported.

Moreover, we defined a new[2] type of change point, i.e., "the third or fourth moment of factor loading changes". Our experiment showed that our method can detect these change points. Our framework does not exert power in other tasks such as prediction and factor analysis. Meanwhile, in the context of non-parametric change point detection, it has a high expressive power.

The main contributions of this paper are as follows.

- We propose YMN decomposition, which decompose one time-series data matrix into a product of a random matrix and a sparse matrix.
- We propose a structural change point detection method using YMN decomposition.
- We demonstrate the effectiveness of our structural change point detection techniques using experimental data and real data.

The remainder of the paper is organized as follows. In the next section, we present the related work. In Section 3, we clearly define the problems to solve. In Section 4, we propose the method. The method is evaluated in Section 5. We discuss the experimental results, issues, and future work in Section 6. Finally, we conclude the paper in Section 7.

## 2 RELATED WORK

This paper proposes a new structural change point detection method using nonparametric matrix decomposition. In this section, we review the previous research that is related to ours in terms of (1) method and (2) purpose. We do not review the works that are not directly related to ours. Therefore, we refer to the survey by Truong et al. [38] for a general review of change point detection methods.

### 2.1 Non-parametric change point detection

As mentioned in the introduction, change point detection can be divided into two types: (1) parametric and (2) nonparametric change point detection.

Because parametric change point detection places strong assumptions on the distribution of time series, it exhibits strong capability within the assumption of the model users, but any unexpected change point cannot be detected. Meanwhile, nonparametric change point detection methods place only a few assumptions on the time-series distribution of data. Therefore,

[2]This concept is not new. However, no research in econometrics has defined this type of change yet.

the model is free and can present a robust result; however, their results often cannot be interpreted, and they can be less accurate than parametric methods in some situations, e.g., the observed time-series data completely follows one stochastic model.

Nonparametric change point detection methods are primarily divided into two types. One type uses matrix decomposition and the other uses kernel methods. The kernel methods [21, 26] demonstrate a high ability when handling data with nonlinear models; however, they are similar to parametric methods in that the results strongly depend on the choice of the kernel functions. On the other hand, methods that fit each purpose and each data's property are developed in each field, such as Robust PCA [41], NMF [24], and ICA [3]. However, because a unique method to decompose one matrix into two matrices does not exist, it is necessary to place some assumptions on both matrices. These assumptions significantly restrict the degree of freedom of the model. PCA assumes no covariance on latent data,ICA assumes independence on latent data. Hence, the method of matrix decomposition used in each field differs significantly.

## 2.2 Structural change point detection

Next, we will explain the structural change point detection methods. In the parametric method, it is possible to assume a graph structure or a time-series structure such as auto-regressive (AR) and moving average [25, 40]. Meanwhile, in nonparametric methods, such a strong assumption cannot be made originally, but many empirical studies indicate the magnitude of the expressive power of the linear model. Among them, research to verify the magnitude of the changes in the linear model factor loadings is not attracting much attention; however, it is valuable in many fields including economics, finance, biology, neuroscience, and computer network analysis. However, economics and finance contain rich multidimensional time-series data from the beginning, and they unveil both their economic structure and structural changes. Changes in the number of factors are the most popular research topics [5], and the structural changes in AR models are also popular [39]. Recently, a method using wavelet transform and PCA has been developed, and this method can be applied to both structural changes above [7]. Meanwhile, in the finance literature, a method that uses ICA has been developed [13].

Finally, we clarify our position. First, our method is an online change point detection method using nonparametric matrix decomposition. Our method can be considered a matrix decomposition method; meanwhile, it is close to the kernel method in that it assumes that the latent time-series follows some model. However, in the context of online change point detection, time-series are generated within a small window width; therefore, the randomly generated time-series variation is large, and our method can be considered as model-free. Meanwhile, our purpose is to detect the change point of the structure, and can be applied to all the structural changes defined above. It is also applicable to new types of changes such as changes in higher moments of the mixing matrix. Under such circumstances, our method can be regarded as a new nonparametric method to detect structural changes.

## 3 PROBLEM STATEMENT

In this section, we mathematically and clearly explain the models that we will investigate, define the structural change that we want to detect and explain the change according to the real-world applications.Table 2 summarizes the notations used herein.

**Table 2: Notation.**

| Notation | Definition |
|---|---|
| $D$ | dimension of observed data |
| $K$ | dimension of latent data |
| $t$ | index of time: $t \in \{0, 1, \ldots, T-1\}$ |
| $t^*$ | time of the change point |
| $\boldsymbol{y}_t$ | $D$-dimensional observed data at time $t$, |
| $y_{td}$ | $d$-th element of $\boldsymbol{y}_t$ |
| $\boldsymbol{y}$ | observed time series: $\boldsymbol{y} = \langle \boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{T-1} \rangle$, |
| $A_t$ | a constant mixture matrix at time $t$ |
| $z_t$ | i.i.d. latent time series, which is a real $K$-vector, at time $t$ |
| $N_t$ | large random matrix used in Algorithm 1 |
| $Q$ | dimension of large random matrix |

We begin from the generative model. We observe a vector $\boldsymbol{y}_t \in \mathbb{R}^D$ at every time $t$. We assume that the vector $\boldsymbol{y}_t$ is generated as a mixture of latent time series

$$\boldsymbol{y}_t = A_t z_t,$$

where $A_t \in \mathbb{R}^{D \times K}$ is a mixing matrix and $z_t \in \mathbb{R}^K$ is a latent time series. We can observe only $\boldsymbol{y}_t$ and that both $A_t$ and $z_t$ are latent.

As is often the case in an econometric setting $z_t$ is assumed to follow one stable deterministic or stochastic model. $A_t$ is assumed to be piece-wise constant with respect to time and shows the *structure* behind data. Therefore, $A_t$ may change abruptly because major events occurred and the relation between the latent and the observed changed. In this study, we denote $t^*$ as the time of **a structural change point** if the following holds:

$$A_t = \begin{cases} A^{(1)} & (t < t^*), \\ A^{(2)} & (t \geq t^*), \end{cases}$$

where $A^{(1)}, A^{(2)} \in \mathbb{R}^{D \times K}$ are constants and $A^{(1)} \neq A^{(2)}$. Meanwhile, latent random signal $z_t$ is assumed to follow a Gaussian distribution at any time, and it is independent and identically distributed (i.i.d.) with respect to time:

$$z_t \sim N(0, I_K).$$

where $I_K$ is an identity matrix in $\mathbb{R}^{K \times K}$. Finally, the problem is to obtain the time of the structural change point $t^*$, and the estimation of $A_t$ is of no interest.

As for the type of changes that we will detect, we will follow the literature of econometric and finance [5, 7]. The structural changes defined in this literature can be divided into three types, that is

(1) change in the $n$-th moment of $A_t$
(2) change in the dimension of $A_t$ and $z_t$
(3) change in the autocorrelation of $\boldsymbol{y}_t$

Finally, we will explain the real-world applications and interpretation of these three changes.

Change (1) is the simplest change. For example, the 1st-order moment change in the $A_t$ means the abrupt increase and decrease of the values. The 2nd-order moment change in the $A_t$ means some of the latent factors are increasingly or decreasingly correlated to the observed data. The 4th-order moment change in the $A_t$ means the increase or decrease in the number of non-zero entries of $A_t$.

Change (2) is also under intense investigation in econometrics because a new connection between the latent and observed time-series are recognized and the explanation capability of the old model decreased if the number of factors increased. Although this is contrary to the setting, "The dimension of latent data $K$ is a constant," we will address this type of change point similarly.

Change (3) appears not related to our models, but by substituting $\langle z_t, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \cdots \rangle$ for $z_t$, we can also address the time-series models. This change is also important in many literature because the time-series becomes less predictable from the observed time-series if the autocorrelation of the $\boldsymbol{y}$ decreased and vice versa.

## 4 METHOD

In this section, we propose our structural change point detection method. Our goal is to detect the change in the matrix $A_t$. Although both coefficients $A_t$ and factors $z_t$ are latent, the theory of random matrix and statistics [10, 11, 36] suggests that we can approximate $z_t$ by selecting some vectors in a large random matrix, and that $A_t$ can be determined by Lasso regression. We first decompose the matrix of observed vectors based on this suggestion, and subsequently evaluate the change in the estimated $A_t$.

### 4.1 Theory of Random Matrix and Lasso

In this section, we describe the methodology and the theory of our method. First, we will explain the problem, and how and why that problem can be solved.

As described in the previous section, our goal is to detect the change in $A_t$. Most of the existing methods such as PCA, NMF, and ICA solve this problem by estimating the $A_t$ and subsequently detect the change in the estimated $A_t$. However, we will directly calculate the change $d(A_t, A_{t-1})$ for an ideal distance of matrix $d$, and we are not interested in estimating $A_t$.

To calculate the change in structure $d(A_t, A_{t-1})$, our method decomposes the observed data $\boldsymbol{y}_t$ into a sparse matrix $M_t$ and a large random matrix $N_t$, and subsequently calculates the change in $M_t$. How our method operates can be explained by the following three basic ideas.

(1) If we generate many random vectors $\{N_t^{(i)}\}_{i=1}^Q$, the linear combination of $\{N_t^{(i)}\}_{i=1}^Q$ may be similar to the true latent variables $z_t$.

(2) If we perform a Lasso regression analysis, setting $N_t$ (random vectors) as explanatory variables and $\boldsymbol{y}_t$ (observed time-series) as explained variables, then random vectors in $N_t$ that are similar to $z_t$ may be automatically chosen and used to explain $\boldsymbol{y}_t$.

(3) If we perform such a Lasso regression analysis and obtain $M_t$ as a sparse coefficient matrix (i.e., $\boldsymbol{y}_t = M_t N_t$), the change in $M_t$ may be related with the change in $A_t$ (structural change = true coefficient matrix's change).

We explain our basic ideas briefly but they are mathematically formulated and confirmed by the existing random matrix theory and statistical theory in this section. For idea (1), we can use the random matrix theory. Suppose we generated $K$-dimensional random vectors denoted by $\{N_t^{(i)}\}_{i=1}^Q$ for $Q$ times. As above, we are considering the case where $Q$ is much larger than $K$. For example, in our experiments, we consider the case with $Q = 100 \max\{D, K\}$. If the matrix $\langle N_t^{(1)}, \cdots N_t^{(Q)} \rangle$ is full rank (i.e., rank = $K$), some of the $\{N_t^{(i)}\}_{i=1}^Q$ are linearly independent and we can obtain any vector in $\mathbb{R}^K$ (including $Z_t$) by the linear

combination of $\{N_t^{(i)}\}_{i=1}^Q$. The situation where the matrix is full rank can be interpreted in that any singular value of the matrix is non-zero. Many studies regarding the singular value of the random matrix have been conducted [35], and we used the result of Tao and Vu [36].

PROPOSITION 4.1 (THEOREM 1.3 [36]). *Let $\xi$ be a real random variable, $M_n(\xi)$ be the random $n \times n$ matrix whose entries are i.i.d. copies of $\xi$, and $\sigma_i(M)$ be the i-th largest singular value of a matrix $M$.*

*Suppose that $\mathrm{E}[\xi] = 0, \mathrm{E}[\xi^2] = 1$ and $\mathrm{E}[\xi^{C_0}] < \infty$ for some sufficiently large absolute constant $C_0$. Subsequently, for all $t > 0$, we have*

$$p(n\sigma_n(M_n(\xi))^2 \le t) = 1 - e^{-t/2 - \sqrt{t}} + O(n^{-c})$$

*where $c > 0$ is an absolute constant and implied constants in the $O(.)$ The notation depends on $\mathrm{E}[\xi^{C_0}]$ but are uniform on $t$*

This proposition can be used in any normalized square, random matrix (i.e., each entry are i.i.d. random variables whose mean and variance are 0 and 1, respectively). The stronger result is obtained in the Gaussian case (Theorem 1.1 in [11]). Moreover, these results can be extended to rectangular matrices (Theorem 6.5 in [11]).

Idea (2) is mathematically formulated as follows.

PROPOSITION 4.2 (MATHEMATICAL FORMULATION OF IDEA (2)). *Let*

- *$z_{t1}, \cdots, z_{tn} \in \mathbb{R}$ be i.i.d random variables.*
- *$a_1, \cdots a_n \in \mathbb{R}$ be constants.*
- *$y_t := a_1 z_{t1} + \cdots + a_n z_{tn}$ for all $t = 0, 1, \cdots, T - 1$*
- *$w_{t1}, \cdots, w_{tm} \in \mathbb{R}$ be i.i.d random variables*
- *where each pairs $z_{ti}, w_{tj}$ are independent.*

*and we estimate the coefficients by Lasso, such that*

$$y_t = b_1 z_{t1} + \cdots + b_n z_{tn} + b_{n+1} w_{t1} + \cdots + b_{n+m} w_{tm}$$

*holds. Then,*

- *$b_i \to a_i$ for $i = 1, \cdots, n$*
- *$b_i \to 0$ for $i = n + 1, \cdots n + m$*

*as $T \to \infty$.*

To prove this proposition, we use a seminal work by Candes and Tao [10, 11]. They proved this proposition within this assumption (restricted isometry constants)

*Definition 4.3 (Definition 1.1 [10] Restricted Isometry Constants).* Let $F$ be the matrix with the finite collection of vectors $(v_j)_{j \in J} \in \mathbb{R}^p$ as columns. For every integer $1 \le S \le |J|$, we define the $S$-restricted isometry constants $\delta_S$ to be the smallest quantity such that $F_T$ obeys

$$(1 - \delta_S)\|c\|^2 \le \|F_T c\|^2 \le (1 + \delta_S)\|c\|^2$$

for all subset $T \subset J$ of cardinality $|T| \le S$, and all real coefficients $(c_j)_{j \in T}$. Similarly, we define the $S, S'$-restricted orthogonality constants $\theta_{S, S'}$ for $S + S' \le |J|$ to be the smallest quantity such that

$$|\langle F_T c, F_T c' \rangle| \le \theta_{S, S'} \cdot \|c\|\|c'\|$$

holds for all disjoint sets $T, T' \subset J$ of cardinality $|T| \le S$ and $|T'| \le S'$.

The numbers $\delta_S$ and $\theta_{S, S'}$ measure how close the vectors are to behaving as an orthonormal system. Subsequently, Proposition 4.2 can be proven as follows.

PROPOSITION 4.4 (THEOREM 1.1 [11]). *Consider the linear model $y = X\beta + z$ where $X \in \mathbb{R}^{n \times p}, y \in \mathbb{R}^n, z \sim N(0, \sigma^2 I_n)$ Suppose $\beta \in \mathbb{R}^p$ is a vector of parameters that satisfies*

- *$\|\beta\|_{L^0} = |\{i \mid \beta_i \ne 0\}| = S$*
- *$\delta_{2S} + \theta_{S, 2S} < 1$*

*where $\delta_{2S}, \theta_{S, 2S}$ are the constants defined in Definition Def4.3. Here, we will estimate $\beta'$ by setting $\lambda_p = \sqrt{2 \log p}$*

$$\beta' = \operatorname{argmin}_{\tilde{\beta} \in \mathbb{R}^p} \|\tilde{\beta}\|_{L^1} \text{ subject to } \|X^*(y - X\tilde{\beta})\|_{L^\infty} \le \lambda_p \sigma$$

*Then $\beta'$ obeys*

$$\|\beta' - \beta\|_{L^2}^2 \le C_1^2 \cdot (2 \log p) \cdot S \cdot \sigma^2$$

*where $C_1 = 4/(1 - \delta_S - \theta_{S, 2S})$ with large probability.* [3]

## 4.2 YMN Matrix Decomposition and Structural Change Point Detection

Idea (1) can be interpreted as, "we can generate the true latent time series $z_t$;" idea (2) can be interpreted as "we can obtain the true latent time series $z_t$". Therefore, by combining ideas (1) and (2), we will obtain idea (3), "we can generate and obtain the true latent time-series and the true latent mixing matrix."

CONJECTURE 4.5 (MATHEMATICAL FORMULATION OF IDEA (3)). *Let*

- *$y_t = A_t z_t$*
- *$y_t \in \mathbb{R}^{D \times T}$ be an observed matrix,*
- *$z_t \in \mathbb{R}^{K \times T}$ be a latent matrix,*
- *$A_t \in \mathbb{R}^{D \times K}$ be a mixing matrix,*
- *$N_t \in \mathbb{R}^{Q \times T}$ be a large random matrix.*
- *$M_t \in \mathbb{R}^{D \times Q}$ be the sparsest matrix such that $\|y_t - M_t N_t\| \le (1 - \theta)\|y_t\|$ holds for a given hyperparameter $\theta$.*

*Then, distances $d_1$ and $d_2$ exist in $\mathbb{R}^{D \times K}$ and $\mathbb{R}^{D \times Q}$ such that*

$$d_1(A_{t_1}, A_{t_2}) \simeq d_2(M_{t_1}, M_{t_2})$$

*holds for all $0 < t_1 < t_2 < T$*

This conjecture suggests that we can approximate the change of mixing matrix $d_1(A_{t_1}, A_{t_2})$ by the change of the sparse matrix $d_2(M_{t_1}, M_{t_2})$ and that we don't have to directly identify $A$ in change point detection problems. In the remainder of this section, we show the matrix decomposition algorithm and the change point detection algorithm based on this conjecture. We do not know the properties of the distances $d_1 and d_2$. However, we used the function $d_2(M_{t_1}, M_{t_2}) := \|M_{t_1}\|_{L^2}/\|M_{t_2}\|_{L^2}$, which is not a distance but is robust under the permutations of the row vectors of $M$.

Algorithm 1[4] shows how our matrix decomposition using a large random matrix is conducted. Our ultimate goal is to decompose the observed data $y_t$ into a random matrix $N_t$ and sparse matrix $M_t$ such that the change in the $M_t$ is related to the change in $A_t$. To obtain a sparse matrix $M_t$, we first generate a large random matrix $N_t$; subsequently, our method minimizes the sparseness of the coefficient matrix within a constraint $\|Y - M_t N_t\| \le (1 - \theta)\|Y\|$ where $Y$ is the $D \times T$ matrix whose row vectors are $y_t$. Algorithm 2 shows our change point detection

---

[3] The term "with large probability" is used in the original paper [11]. This term means that the probability that the equation holds is above $1 - \frac{1}{\sqrt{\pi \log p}}$

[4] We name this algorithm "YMN" after the equation $Y = MN$. $M$ is the first initial of the mixing matrix and mechanism, and $N$ is the first initial of the normalized random matrix.

**Algorithm 1** Matrix decomposition using a large random matrix (YMN)

---

**Definition**
$Y$: $D \times T$ matrix that we want to decompose
$M$: $D \times K$ sparse matrix
$N$: $K \times T$ large random matrix
$N_i$: $1 \times T$ $i$-th row vector of $N$
$E[x], V[x]$: the mean and variance of vector $x$'s entities
$\theta$: threshold value of Lasso fitting score
$\alpha$: hyperparameter of Lasso, i.e., the coefficient of the $L_1$ term
$\delta$: step value of $\alpha$, used during hyperparameter optimization
Lasso$(X, T; \alpha)$: matrix $W$ that minimizes $||T - WX||^2 + \alpha ||W||^1$

**End Definition**
$N = \langle N_0, \cdots, N_{K-1} \rangle \sim N(0, I)$
**for** $i = 0, \cdots, K - 1$ **do**
  $N_i \Leftarrow (N_i - E[N_i]) / \sqrt{V[N_i]}$
**end for**
$M \Leftarrow$ Lasso$(Y, N; \alpha)$
**while** $||Y - MN|| \leq (1 - \theta)||Y||$ **do**
  $\alpha \Leftarrow \alpha + \delta$
  $M \Leftarrow$ Lasso$(Y, N; \alpha)$
**end while**

---

algorithm. First, we use the series of window frames $Y_t$ for each time and perform the matrix decomposition of $Y_t$. At this time, we have a series of matrix $M_t$. We calculate the change point score using the distance between $M_t$ and $M_{t+1}$.

---

**Algorithm 2** Change point detection algorithm using YMN decomposition

---

**Definition**
$w$: window size, $t$: current time
$y_t$: series vector, $s_t$: change point score
$\langle, \rangle$: concatenation of vectors into a matrix.
MatrixDecomp$(Y)$: Decompose $Y$ into two matrices by YMN decomposition.
**End Definition**

**for** $t = 0, \cdots, T - w - 1$ **do**
  $Y_t \Leftarrow \langle y_t, y_{t+1}, \ldots, y_{t+w-1} \rangle$
  $N_t M_t \Leftarrow$ MatrixDecomp$(Y_t)$
  $s_t \Leftarrow d_2(M_t, M_{t+1})$
**end for**

---

## 5 EXPERIMENT

To demonstrate that our methods can detect structural change points, we conducted two experiments. First, we conducted experiments in artificial settings where their structures and changes are clear. We demonstrate which type of structural change points can be detected and how these changes are detected clearly by these experiments.

Subsequently, we conducted experiments in real-world settings where the structures and changes in structures are not clear. We used economics [18] and financial [17] data that are popular in the literature of structural change point detection. We demonstrate that our method can be used in the real-world through these experiments.

**Table 3: Settings of Experiment.**

| Notation | Setting | Definition |
|---|---|---|
| $D$ | 100 | dimensions of observed data |
| $K$ | 10 | dimensions of latent data |
| $T$ | 100 | max index of time $0 \leq t < T$ |
| $t^*$ | 50 | time of the change point |
| $A_t$ | see below | a mixture matrix at time $t$ |
| $z_t$ | $N(0, I_K)$ | i.i.d. latent time series |
| $y_t$ | $A_t z_t$ | observed data at time $t$, |
| $Q$ | 10000 | dimension of random matrix |
| $\theta$ | 0.99 | fitting score in Algorithm 1 |
| $w$ | 10 | time window in Algorithm 2 |
| $d_2(X, Y)$ | $||X||_{L^2} / ||Y||_{L^2}$ | $d_2$ in Conjecture 4.5 |

### 5.1 Artificial Data

To verify that our method can detect the structural changes described above, we conducted experiments with artificial data and compared the AUC scores to those of the existing methods (PCA, ICA, and NMF). All of the experiments were conducted in an "online" setting. In other words, we fetched the time-series data $y_t$ for $w$ times and decomposed the matrix $\langle y_t, y_{t+1}, \ldots, y_{t+w-1} \rangle$ into two matrices $M_t, N_t$ and detected the change in $M_t$. Change point detection algorithms using PCA, ICA, and NMF are executed in the same manner as Algorithm 2, that is, by defining MatrixDecomp in Algorithm 2 as PCA, ICA, and NMF. For example, if we perform a dimensional reduction from $Y_t$ to $Z_t$ such that $Z_t = M_t Y_t$ holds by PCA, then we calculate $d_2(M_t, M_{t+1})$. In addition, if we perform a matrix decomposition by ICA or NMF such that $Y_t = M_t N_t$ holds, then we calculate $d_2(M_t, M_{t+1})$. Note that if we perform $NMF$, we consider $Y_t + C$ instead of $Y_t$ where $C$ is a constant such that $Y_t + C > 0$ holds for all $t$. Table 3 summarizes the settings of these experiments.

In this experiment, we detected the following types of changes.
(1) Mean of Factor Loadings' Change
$$A_t = \begin{cases} A & (t < t^*), \\ A + W & (t \geq t^*), \end{cases}$$
where
$A_{i,j} \sim N(0, 1), W_{i,j} \sim N(\mu, 1), \mu \in \{0.5, 1.0, \cdots, 4.5, 5.0\}$
(2) Variance of Factor Loadings' Change
$$A_t = \begin{cases} A & (t < t^*), \\ A + W & (t \geq t^*), \end{cases}$$
where
$A_{i,j} \sim N(0, 1), W_{i,j} \sim N(0, \sigma), \sigma \in \{0.5, 1.0, \cdots, 4.5, 5.0\}$
(3) Skew and Kurtosis of Factor Loadings' Change
$$A_t = \begin{cases} A & (t < t^*), \\ A + W & (t \geq t^*), \end{cases}$$
where
$A_{i,j} \sim N(0, 1), W_{i,j} \sim t_{df}(0, 1), df \in \{1, 3, 5, \cdots, 19\}$
(4) Number of Factors' Change
$$A_t = \begin{cases} A^{(0)} \in \mathbb{R}^{D \times K} & (t < t^*), \\ A^{(1)} \in \mathbb{R}^{D \times (K+k)} & (t \geq t^*), \end{cases}$$
where $A_{i,j}^{(0)}, A_{i,j}^{(1)} \sim N(0, 1), k \in \{2, 4, 6, 8, 10\}$
(5) Coefficient change in AR(1)
$$y_{t+1} = \alpha_t y_t + \beta_t \mathbf{1} + \gamma_t \epsilon_t$$

**Table 4: Comparison of AUC scores. Experiments (1)-(5) shows our methods are effective and Experiments (6) shows we cannot detect that type of changes.**

| Methods | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| PCA | 0.877 ± 0.0502 | 0.961 ± 0.0239 | 0.671 ± 0.0566 | 0.729 ± 0.0470 | 0.707 ± 0.0524 | 0.628 ± 0.0548 |
| ICA | 0.881 ± 0.0474 | 0.949 ± 0.0335 | 0.701 ± 0.0508 | 0.800 ± 0.0418 | 0.701 ± 0.0425 | 0.611 ± 0.0560 |
| NMF | 0.629 ± 0.0705 | 0.804 ± 0.0642 | 0.583 ± 0.0500 | 0.721 ± 0.0541 | 0.614 ± 0.0502 | 0.659 ± 0.0507 |
| **Ours** | **1.000 ± 0.0000** | **1.000 ± 0.0000** | **0.952 ± 0.0204** | **0.998 ± 0.0010** | **0.999± 0.0001** | 0.552 ± 0.0637 |

$$\alpha_t = \begin{cases} \alpha_0 & (t < t^*), \\ \alpha_0 - \delta & (t \geq t^*), \end{cases}$$

where $\alpha_0 = 0.9$ and $\delta \in \{0.1, 0.2, \cdots, 0.8\}$ and $\beta_t = 1 - \alpha_t$ and $\gamma_t = \sqrt{1 - \alpha_t}$ and $\epsilon_t \sim N(0, 1)$. By this settings $y_t$ always satisfies normal conditions $E[y_t] \equiv 0$, $\text{Var}[y_t] \equiv 1$.

(6) Resample from same distribution

$$A_t = \begin{cases} A & (t < t^*), \\ B & (t \geq t^*), \end{cases}$$

where
$$A_{i,j} \sim N(0, 1), B_{i,j} \sim N(0, 1), \text{i.i.d.}$$

Table 4 shows the mean and standard deviation of the AUCs for each task.

The hyperparameters in all experiments in Table 4 are those of the easiest tasks, i.e., (1) $\mu = 5.0$, (2) $\sigma = 5.0$, (3) $df = 1$, (4) $k = 10$, (5) $\delta = 0.8$. The AUC score in each cell is the score to classify whether a given $y_t$ is the point before/after the change where $t \in \{t^*, t^* \pm 10, t^* \pm 20, t^* \pm 30\}$. For each task, we generated 100 samples at random and detected the change points.

Experiments (1) and (2) are easy tasks, and we can detect the change points with no mistakes. Experiments (3), (4), and (5) are relatively difficult tasks, but we can detect the change points with higher accuracy than the existing methods. Experiment (6) is the most challenging task, and all of the methods cannot detect the change points accurately because there is no change in moments nor distributions.

Figure 5.1 shows the $L^2$ norm of the $M_t$ and we found that the change in the original mixing matrix $A_t$ and the change in the sparse matrix $M_t$ are highly correlated.

## 5.2 Real-world Data

To verify that our method can be used in the real world, we conducted experiments with two famous datasets in economics [18] and finance [17]. First, we calculated $d_2(M_{t+1}, M_t)$ at each $t$. We then listed $A$ points whose $d_2(M_{t+1}, M_t)$ are in the top $A$. Finally, we clustered these $A$ points into $B$ points using k-means clustering. Note that the listing and clustering are performed by off-line settings, whereas our change-point detection method is an online method. Because no ground truth of change points exists in economic or financial data, we only consider true- and false-positive cases. In other words, we listed change points detected by our methods and checked whether a major event occurred in finance or economics.

### 5.2.1 US macroeconomic dataset.
We analyzed the US representative macroeconomic dataset of 101 time series, collected monthly between 1959 and 2006 ($T = 576$), for the change points. We listed 10 remarkable changes resulting from the aforementioned off-line listing method with $P = 30, Q = 10$ in Figure 5.2. Our change point detection methods clearly show the following major change points in the US economy.

- the Great Moderation period that started in 1983
- major economic recession in early 1970s
- major economic recession in early 1980s
- Internet Bubble between 1999 and 2001
- the oil crisis at 1973

These change points are consistent with the existing research on business cycles [29]. Moreover, the Internet Bubble and the oil crisis are not detected in previous research in econometrics [5, 7, 9, 27].

### 5.2.2 US 48 Industry Portfolios.
We analyzed the US 48 industry portfolio collected daily between 1926 and 2018 ($T = 24350$) for the change points [17]. We listed five most significant changes by the same clustering methods as Figure 5.2 in Figure 5.3. Our change point detection methods clearly show the following four major change points in the US stock market history: Wall Street Crash of 1929, Kennedy Slide of 1962, Black Monday of 1987, and financial crisis of 2007. Moreover, we observed that the magnitudes of these change by the scores.

## 6 DISCUSSION

We demonstrated that our method can detect all three types of structural changes defined in Section 3. This is because Conjecture 4.5 is true to some degree. That is, by Proposition 4.1, we can generate the same latent time-series by combining the random time-series; further, by Proposition 4.2, we can detect true latent time series by Lasso. Table 4 shows that this method is much more accurate than the existing matrix decomposition methods. This is because our method was initially created to detect the structural changes that we defined, but the existing methods cannot detect these changes.

Figure 5.1 shows the $L^2$ norm of the $M_t$ and we can see that the change of the original mixing matrix $A_t$ and the change of the sparse matrix $M_t$ are highly correlated. Hence this figure supports our Conjecture 4.5.

However, we could not detect the change (6). This is because the distance we used does not consider the permutation of the row. In other words, we consider only $\|M_t\|_{L^2}$ at each time $t$. Therefore, we cannot detect changes in swaps of elements of $A_t$. We expect that our methods can detect these changes by choosing some of the row vectors and by performing our decomposition repeatedly. From the facts above, it is clear that our method can be applied to the detection of structural changes in economics and finance. As mentioned in the Introduction, econometrics is an advanced research field in handling the change points; therefore, we think that this method can be applied to the detection of structural changes in many fields. For example, neuroscience and network analysis have similar objectives to ours because they cannot detect the exact model of the time-series, and structural changes are important to them.
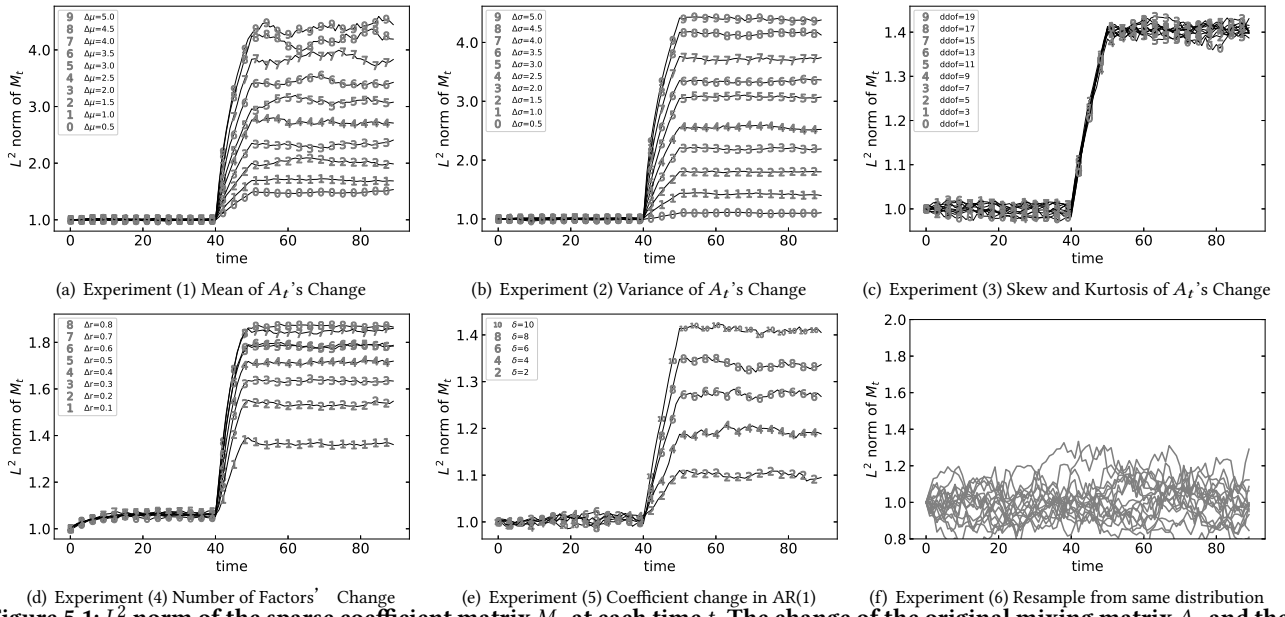
(a) Experiment (1) Mean of $A_t$'s Change

(b) Experiment (2) Variance of $A_t$'s Change

(c) Experiment (3) Skew and Kurtosis of $A_t$'s Change

(d) Experiment (4) Number of Factors' Change

(e) Experiment (5) Coefficient change in AR(1)

(f) Experiment (6) Resample from same distribution

**Figure 5.1:** $L^2$ **norm of the sparse coefficient matrix** $M_t$ **at each time** $t$. **The change of the original mixing matrix** $A_t$ **and the change of the sparse matrix** $M_t$ **are highly correlated.**



(a) with time window = 6 months
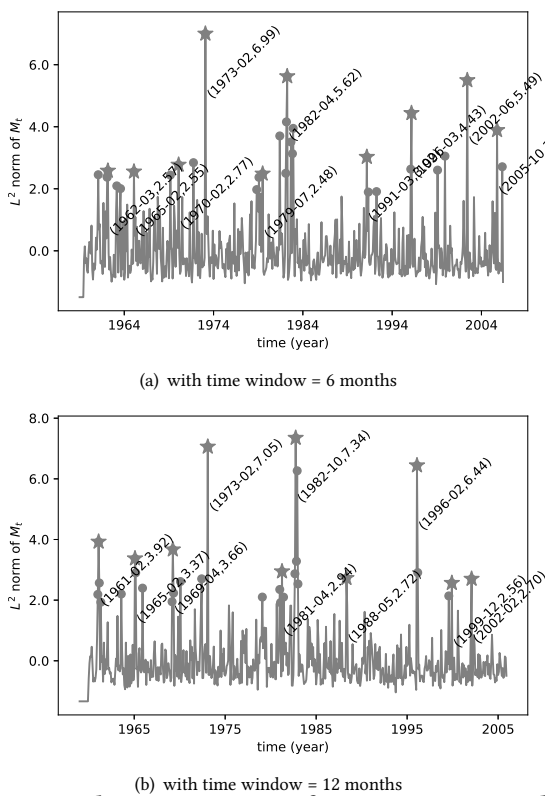


(b) with time window = 12 months

**Figure 5.2: Change point score of US macroeconomic data with time window =** 6**months and** 12**months. We pointed** $A = 30$ **largest change points by circles and labeled** $B = 10$ **selected change points by (year-month, score) and pointed them by stars for a better understanding.**
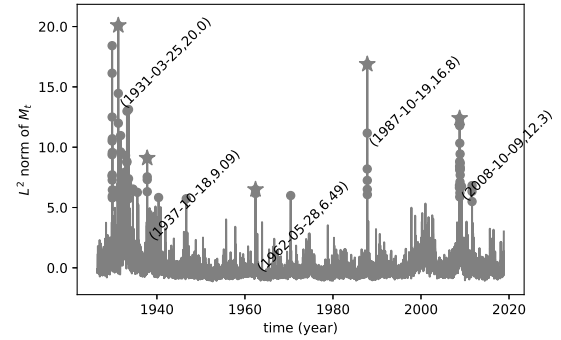


**Figure 5.3: Change point score of US stock data with time window =** 3 **days. We identified 100 most significant change points and labeled five selected change points by (year-month-day, score) and pointed them by stars for a better understanding.**

## 7 CONCLUSION

We herein proposed a new nonparametric change point detection method of the structures of multidimensional time series. We used a large random matrix to generate the latent time series, and Lasso to obtain good row vectors and to select the related row vectors. We also demonstrated that our method could detect not only typical structural changes used in econometrics and finance but also new types of structural changes such as the changes in the higher moment of the mixing matrix. With the random matrix theory and statistical theory of Lasso, we partly unveiled the mechanism and theory of our methods. However, a conjecture that fully supports our methods remains unproven. We demonstrated the effectiveness of our change point detection techniques by artificial data and real-world data in economics and finance. Similar structural changes may occur in many fields, but we have allocated comparison with the state-of-the-art methods and specialization in other fields as future work.

# ACKNOWLEDGMENT

# REFERENCES

[1] Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. 2016. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications* 60 (jan 2016), 19–31. https://doi.org/10.1016/j.jnca.2015.11.016

[2] Donald W. K. Andrews. 1993. Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica* 61, 4 (jul 1993), 821. https://doi.org/10.2307/2951764

[3] Yasunori Aoki, Ryouhei Ishii, Roberto D. Pascual-Marqui, Leonides Canuet, Shunichiro Ikeda, Masahiro Hata, Kaoru Imajo, Haruyasu Matsuzaki, Toshimitsu Musha, Takashi Asada, Masao Iwase, and Masatoshi Takeda. 2015. Detection of EEG-resting state independent networks by eLORETA-ICA method. *Frontiers in Human Neuroscience* 9 (feb 2015), 31. https://doi.org/10.3389/fnhum.2015.00031

[4] Reinaldo B. Arellano-Valle, Luis M. Castro, and Rosangela H. Loschi. 2013. Change Point Detection in The Skew-Normal Model Parameters. *Communications in Statistics - Theory and Methods* 42, 4 (feb 2013), 603–618. https://doi.org/10.1080/03610926.2011.611321

[5] Jushan Bai and Serena Ng. 2002. Determining the Number of Factors in Approximate Factor Models. *Econometrica* 70, 1 (jan 2002), 191–221. https://doi.org/10.1111/1468-0262.00273

[6] Jushan Bai and Serena Ng. 2008. Large Dimensional Factor Analysis. *Foundations and Trends® in Econometrics* 3, 2 (2008), 89–163. https://doi.org/10.1561/0800000002

[7] Matteo Barigozzi, Haeran Cho, and Piotr Fryzlewicz. 2018. Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics* 206, 1 (sep 2018), 187–225. https://doi.org/10.1016/j.jeconom.2018.05.003

[8] Vincent Brault, Sarah Ouadah, Laure Sansonnet, and Céline Lévy-Leduc. 2018. Nonparametric multiple change-point estimation for analyzing large Hi-C data matrices. *Journal of Multivariate Analysis* 165 (may 2018), 143–165. https://doi.org/10.1016/j.jmva.2017.12.005

[9] Jörg Breitung and Sandra Eickmeier. 2011. Testing for structural breaks in dynamic factor models. *Journal of Econometrics* 163, 1 (jul 2011), 71–84. https://doi.org/10.1016/j.jeconom.2010.11.008

[10] E.J. Candes and T. Tao. 2005. Decoding by Linear Programming. *IEEE Transactions on Information Theory* 51, 12 (dec 2005), 4203–4215. https://doi.org/10.1109/tit.2005.858979

[11] Emmanuel Candes and Terence Tao. 2007. The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics* 35, 6 (dec 2007), 2313–2351. https://doi.org/10.1214/009053606000001523

[12] Liang Chen, Juan J. Dolado, and Jesús Gonzalo. 2014. Detecting big structural breaks in large factor models. *Journal of Econometrics* 180, 1 (may 2014), 30–48. https://doi.org/10.1016/j.jeconom.2014.01.006

[13] Ray-Bing Chen, Ying Chen, and Wolfgang K. Härdle. 2014. TVICA—Time varying independent component analysis and its application to financial data. *Computational Statistics & Data Analysis* 74 (jun 2014), 95–109. https://doi.org/10.1016/j.csda.2014.01.002

[14] Peter F Christoffersen. 2001. Forecasting Non-Stationary Economic Time Series. *J. Amer. Statist. Assoc.* 96, 453 (mar 2001), 339–355. https://doi.org/10.1198/jasa.2001.s378

[15] R. Cont. 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1, 2 (feb 2001), 223–236. https://doi.org/10.1080/713665670

[16] Graham Elliott and Allan Timmermann. 2008. Economic Forecasting. *Journal of Economic Literature* 46, 1 (feb 2008), 3–56. https://doi.org/10.1257/jel.46.1.3

[17] Kenneth R. French. 2018. Detail for 48 Industry Portfolios. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_48_ind_port.html

[18] Giampiero Gallo. 2010. The methodology and practice of econometrics by J. L. Castle; N. Shephard. *Journal of Economics* 101 (01 2010), 99–101. https://doi.org/10.2307/41795703

[19] Allan Gibbard and Hal R. Varian and. 1978. Economic Models. *Journal of Philosophy* 75, 11 (1978), 664–677. https://doi.org/10.5840/jphil1978751111

[20] Xu Han and Atsushi Inoue. 2014. TESTS FOR PARAMETER INSTABILITY IN DYNAMIC FACTOR MODELS. *Econometric Theory* 31 (sep 2014), 1–36. https://doi.org/10.1017/s0266466614000486

[21] Zaïd Harchaoui, Francis Bach, and Éric Moulines. 2008. Kernel Change-point Analysis. In *Proceedings of the 21st International Conference on Neural Information Processing Systems (NIPS'08)*. Curran Associates Inc., USA, 609–616. http://dl.acm.org/citation.cfm?id=2981780.2981856

[22] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. 2004. Extreme learning machine: a new learning scheme of feedforward neural networks, In 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541). *IEEE International Conference on Neural Networks - Conference Proceedings* 2, 985 – 990 vol.2. https://doi.org/10.1109/ijcnn.2004.1380068

[23] A. Hyvärinen and E. Oja. 2000. Independent component analysis: algorithms and applications. *Neural Networks* 13, 4-5 (jun 2000), 411–430. https://doi.org/10.1016/s0893-6080(00)00026-5

[24] Daichi Kitamura, Shinichi Mogami, Yoshiki Mitsui, Norihiro Takamune, Hiroshi Saruwatari, Nobutaka Ono, Yu Takahashi, and Kazunobu Kondo. 2018. Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation. *EURASIP Journal on Advances in Signal Processing* 2018, 1 (may 2018), 28. https://doi.org/10.1186/s13634-018-0549-5

[25] Jeremias Knoblauch and Theodoros Damoulas. 2018. Spatio-temporal Bayesian On-line Changepoint Detection with Model Selection. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Jennifer Dy and Andreas Krause (Eds.), Vol. 80. PMLR, Stockholmsmassan, Stockholm Sweden, 2718–2727. http://proceedings.mlr.press/v80/knoblauch18a.html

[26] Shuang Li, Yao Xie, Hanjun Dai, and Le Song. 2015. M-statistic for Kernel Change-point Detection. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'15)*. MIT Press, Cambridge, MA, USA, 3366–3374. http://dl.acm.org/citation.cfm?id=2969442.2969615

[27] Shujie Ma and Liangjun Su. 2018. Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics* 207, 1 (nov 2018), 1–29. https://doi.org/10.1016/j.jeconom.2018.06.019

[28] R. David McLean and Jeffrey E. Pontiff. 2016. Does Academic Research Destroy Stock Return Predictability? *The Journal of Finance* 71, 1 (jan 2016), 5–32. https://doi.org/10.1111/jofi.12365

[29] Inc. National Bureau of Economic Research. 2012. US Business Cycle Expansions and Contractions. https://www.nber.org/cycles.html

[30] Grace Ngunkeng and Wei Ning. 2014. Information Approach for the Change-Point Detection in the Skew Normal Distribution and Its Applications. *Sequential Analysis* 33, 4 (oct 2014), 475–490. https://doi.org/10.1080/07474946.2014.961845

[31] Wei Ning and A. K. Gupta. 2009. Change Point Analysis for Generalized Lambda Distribution. *Communications in Statistics - Simulation and Computation* 38, 9 (oct 2009), 1789–1802. https://doi.org/10.1080/03610910903125314

[32] Bradley S. Paye and Allan Timmermann. 2006. Instability of return prediction models. *Journal of Empirical Finance* 13, 3 (jun 2006), 274–315. https://doi.org/10.1016/j.jempfin.2005.11.001

[33] Mario Pitsillis. 2005. 1 - Review of literature on multifactor asset pricing models. In *Linear Factor Models in Finance*, John Knight and Stephen Satchell (Eds.). Butterworth-Heinemann, Oxford, 1 – 11. https://doi.org/10.1016/B978-075066006-8.50002-1

[34] James Stock and Mark Watson. 1994. *Evidence on Structural Instability in Macroeconomic Time Series Relations*. Technical Report. National Bureau of Economic Research. https://doi.org/10.3386/t0164

[35] Terence Tao. 2008. Random matrices: A general approach for the least singular value problem. https://terrytao.wordpress.com/2008/05/22/random-matrices-a-general-approach-for-the-least-singular-value-problem/

[36] Terence Tao and Van Vu. 2010. Random Matrices: the Distribution of the Smallest Singular Values. *Geometric and Functional Analysis* 20, 1 (mar 2010), 260–297. https://doi.org/10.1007/s00039-010-0057-8

[37] Jennifer Thompson. 2017. Smart beta funds pass $ 1tn in assets. https://www.ft.com/content/bb0d1830-e56b-11e7-8b99-0191e45377ec 2017-12-27.

[38] Charles Truong, Laurent Oudre, and Nicolas Vayatis. 2018. Selective review of offline change point detection methods. arXiv:arXiv:1801.00718

[39] Stefan De Wachter and Elias Tzavalis. 2012. Detection of structural breaks in linear dynamic panel data models. *Computational Statistics & Data Analysis* 56, 11 (nov 2012), 3020–3034. https://doi.org/10.1016/j.csda.2012.02.025

[40] Beilun Wang, arshdeep Sekhon, and Yanjun Qi. 2018. Fast and Scalable Learning of Sparse Changes in High-Dimensional Gaussian Graphical Model Structure. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research)*, Amos Storkey and Fernando Perez-Cruz (Eds.), Vol. 84. PMLR, Playa Blanca, Lanzarote, Canary Islands, 1691–1700. http://proceedings.mlr.press/v84/wang18f.html

[41] Wei Xiao, Xiaolin Huang, Jorge Silva, Saba Emrani, and Arin Chaudhuri. 2017. Online Robust Principal Component Analysis with Change Point Detection. arXiv:arXiv:1702.05698

[42] Ian En-Hsu Yen, Wei-Cheng Lee, Sung-En Chang, Arun Sai Suggala, Shou-De Lin, and Pradeep Ravikumar. 2017. Latent Feature Lasso. In *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Doina Precup and Yee Whye Teh (Eds.), Vol. 70. PMLR, International Convention Centre, Sydney, Australia, 3949–3957. http://proceedings.mlr.press/v70/yen17a.html