Change point detection for tensors with heterogeneous slices

Lixing Zhu[1]

Joint work with: Jiaqi Huang[1], Junhui Wang[2] & Xuehu Zhu[3]

1 Beijing Normal University, Beijing
2 Chinese University of Hong Kong, Hong Kong
3 Xi'an Jiaotong University, Xi'an

December 10, 2023

Table of Contents

Introduction

- When the data are collected over time, the generation mechanism may change over time, leading to structural changes at some time points.
- In this research, we focus on offline tensor data.
- many applications present data structure: a structural mode of tensor may be categorical and the corresponding slices may present heterogeneity.

Introduction

- **Example.** Trade data of order-three tensor from Organization for Economic Co-operation and Development `https://data.oecd.org/trade/trade-in-goods-and-services.htm`.

- Mode 1). Import, export, and net trade (3-dimensional mode); Mode 2). Goods and services (2-dimensional mode) ; Mode 3). Four countries: Australia, South Africa, The United Kingdom, and The Unite States (4-dimensional structural mode). Time points: 176 quarters in total from the first quarter of 1979 to the fourth quarter of 2022.

- Our primary objective was to detect significant changes during this time spanning.

Introduction

- Heterogeneity may happen among different countries. This mode may represent specific structural information between countries.
- To detect changes, it should be better to define distance between tensors over time slice-wisely such that information of different slices would not be mixed up.
- Other examples include
  (a) the sensor mode of earthquake forecasting ([Xie et al., 2019], [Chen et al., 2022]),
  (b) the sub-region mode in the Brain cancer incidence about New Mexico in USA ([Kulldorff, 2012]),
  (c) the spatial grid mode of the 50-year high-resolution monthly gridded precipitation over Spain ([Herrera et al., 2012]).

The targets

- Propose a mode-based signal screening Frobenius distance for the moving sums of the difference between two adjacent tensors to mainly solve following three problems:

  (a) As the magnitudes of the changes are unknown, particularly in heterogeneous cases, how to appropriately determine the threshold is an issue.

  (b) It is a concern for how to estimate the sparsity level to adapt both dense and sparse model structure. See [Wang et al., 2021] who introduced a notion of sparsity level.

  (c) It is of interest to handle tensors with heterogeneous slices.

Table of Contents

Preliminary

Let $\mathcal{X}_i \in \mathbb{R}^{p_1 \times \cdots \times p_\kappa}, i = 1, 2, \cdots, n$ be independent order-κ tensors:

$$\mathcal{X}_i = E(\mathcal{X}_i) + \mathcal{E}_i, i = 1, 2, \cdots, n, \tag{1}$$

where $\mathcal{E}_i$'s are independent mean zero error tensors.

- Here, the order κ is the number of dimensions, and each dimension is called a mode. An order-κ tensor is a κ-dimensional array with κ modes.
- The element $(i_1, i_2, \cdots i_\kappa)$ of tensor $\mathcal{X}_i$ is denoted by $\mathcal{X}_{i,i_1 i_2 \cdots i_\kappa}$.
- Without loss of generality, assume the κ-th mode is the structural mode. Any l-th slice according to each l-th element of this mode is an order-$(\kappa - 1)$ tensor. Write it as $\mathcal{X}_{i,l}$.

Preliminary

- Assume that K change points $0 = z_0 < z_1 < \cdots z_k < \cdots < z_K < z_{K+1} = n$ to divide the original sequence to $K+1$ segments so that

$$E(\mathcal{X}_i) = \mathcal{M}^{(k)}, i = z_{k-1}+1, \cdots, z_k.$$

- Write $\alpha_n^\star = \min_{1\le k\le K+1} |z_k - z_{k-1}|$ as the minimum distance between any two change points.

- The number K and the locations z_k's of the change points are unknown.

Preliminary

- Recall that the κ-th mode is the structural mode.

- For any l with $l = 1, \cdots, p_\kappa$, the corresponding l-th slice $\mathcal{X}_{j,l} \in R^{p_1 \times p_2 \times \cdots \times p_{\kappa-1}}$, the set of changes $C_l = \{z_0, z_{k'}$ for $1 \leq k' \leq K_l\}$ is a subset of $0 = z_0 < z_1 < \cdots z_k < \cdots < z_K$ to divide the original sequence to $K_l + 1$ segments so that

$$E(\mathcal{X}_{i,l}) = \mathcal{M}^{(k')}, i = z_{k'-1} + 1, \cdots, z_{k'}.$$

- By the definition, the union of all the above subsets $\cup_l C_l$ equals $\{z_0, z_i, i = 1, \cdots K\}$. and $\alpha_n^\star = \min_{l=1,\cdots,p_\kappa} \min_{1 \leq k' \leq K_l+1} |z_{k'} - z_{k'-1}|$.

The distance

For the l-th slice, define the following mode-based signal-screening metric with a threshold $l_n(s) \to 0$ as $n \to \infty$:

Definition 1

$$||\mathcal{A}_l||_s^2 = \frac{\sum_{i_1=1}^{p_1} \sum_{i_2=2}^{p_2} \cdots \sum_{i_{\kappa-1}=1}^{p_{\kappa-1}} a_{i_1 i_2 \cdots i_{\kappa-1},l}^2 I(a_{i_1 i_2 \cdots i_{\kappa-1},l}^2 > l_n(s))}{\sum_{i_1=1}^{p_1} \sum_{i_2=2}^{p_2} \cdots \sum_{i_{\kappa-1}=1}^{p_{\kappa-1}} I(a_{i_1 i_2 \cdots i_{\kappa-1},l}^2 > l_n(s)) + 1/n}, \quad (2)$$

where $a_{i_1 i_2 \cdots i_{\kappa-1},l}$ denotes the $(i_1, i_2, \cdots, i_{\kappa-1}, l)$ element of the order-$(\kappa - 1)$ tensor $\mathcal{A}_l$
the value $1/n$ in the denominator is to avoid the undefined $0/0$ ratio.
When $l_n(s) = 0$, write $||\mathcal{A}_l||_s^2$ as $||\mathcal{A}_l||_{0s}^2$.

This is an average over significant elements of slice (handling sparsity level).

The moving sums(MOSUM)

Define the MOSUM of the tensor sequence:

$$\mathcal{D}_l(i) = \frac{1}{\alpha_n}\left(\sum_{j=i}^{i+\alpha_n-1} E(\mathcal{X}_{j,l}) - \sum_{j=i+\alpha_n}^{i+2\alpha_n-1} E(\mathcal{X}_{j,l})\right). \tag{3}$$

For each $1 \le k' \le K_l$, $||\mathcal{D}_l(i)||_{0s}^2$ has the following formula:

$$= \begin{cases} 0 & z_{k'-1} \le i < z_{k'} - 2\alpha_n, \\ \frac{\sum_{i_1=1}^{p_1}\cdots\sum_{i_{\kappa-1}=1}^{p_{\kappa-1}} \mathcal{D}^2_{i_1 i_2\cdots i_{\kappa-1},l}(i)}{\sum_{i_1=1}^{p_1}\cdots\sum_{i_{\kappa-1}=1}^{p_{\kappa-1}} I\left(\mathcal{D}^2_{i_1 i_2\cdots i_{\kappa-1},l}(i)>0\right)+\frac{1}{n}} & z_{k'} - 2\alpha_n \le i < z_{k'} - \alpha_n \\ ||\mathcal{M}^{(k+1)} - \mathcal{M}^{(k)}||_{0s}^2 & i = z_{k'} - \alpha_n \\ \frac{\sum_{i_1=1}^{p_1}\cdots\sum_{i_{\kappa-1}=1}^{p_{\kappa-1}} \mathcal{D}^2_{i_1 i_2\cdots i_{\kappa-1},l}(i)}{\sum_{i_1=1}^{p_1}\cdots\sum_{i_{\kappa-1}=1}^{p_{\kappa-1}} I\left(\mathcal{D}^2_{i_1 i_2\cdots i_{\kappa-1},l}(i)>0\right)+\frac{1}{n}} & z_{k'} - \alpha_n < i < z_{k'}. \end{cases}$$

Plot

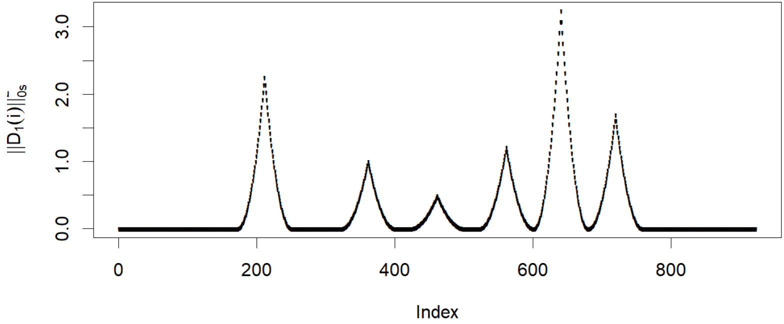

Figure 1: The dotted line is for $||\mathcal{D}_1(i)||^2_{0s}$.

Signal function: ridge ratio

- The periodic pattern of $||\mathcal{D}_l(i)||_{0s}^2$ over time is plotted in Figure 1 above. But we will not directly use it for detection as the highs of peaks would be different and unknown, determining the threshold calls for effort.

- We define a signal function:

$$T_l(i) = \frac{||\mathcal{D}_l(i)||_{0s}^2 + c_{nl}^{\star}(i)}{||\mathcal{D}_l(i+\alpha_n)||_{0s}^2 + c_{nl}^{\star}(i)}. \tag{4}$$

- The ridge function $c_{nl}^{\star}(i)$ to be selected later is to avoid the undefined $0/0$ of $T_l(i)$ such that $T_l(i)$ is close to 1 in the segments, as shown in Figure 1 above, where both $||\mathcal{D}_l(i)||_{0s}^2$ and $||\mathcal{D}_l(i+\alpha_n)||_{0s}^2$ involve no locations of change points.

Signal function: rational

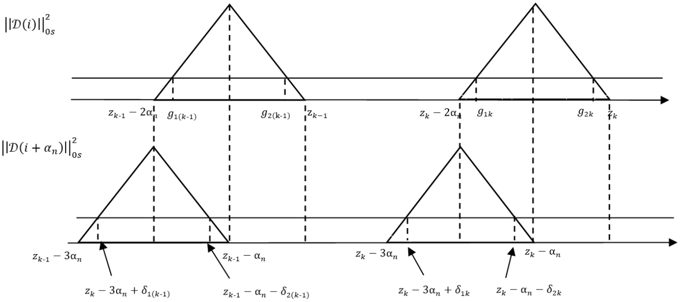

Figure 2: The plot is for the case where the distance between two true change points is not less than $3\alpha_n$.

Curve patterns

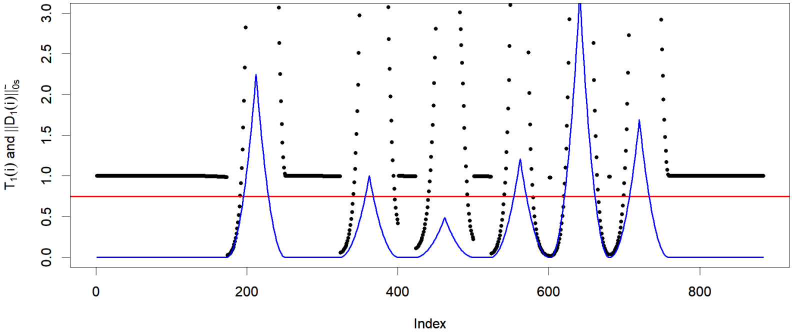

Figure 3: The blue dotted line represents $||\mathcal{D}_1(i)||_{0s}^2$ and the black dotted line corresponds to $T_1(i)$.

The properties of the curve

- 1). When the spacing length $\alpha_n^\star > 3\alpha_n$, $T_l(i)$ monotonically decreases on the left-hand side of $z_{k'} - 2\alpha_n + 1$, drops to 0 at $z_{k'} - 2\alpha_n + 1$, and then increases up to infinity.

- 2). When $2\alpha_n \leq \alpha_n^\star < 3\alpha_n$, $T_l(i)$ monotonically decreases to a local minimum first and then suddenly jumps up to one in the interval with length $f + 1, 0 \leq f < \alpha_n$, and immediately drops down to another local minimum before increasing up to infinity.

- 3). One period corresponds to one change point. The local minimizer $z_{k'} - 2\alpha_n + 1$ in the period corresponds to the change point $z_{k'}$. As the local minima are zero, the threshold for detecting change points can be easily determined.

The adaptive-to-change ridge function

- The ridge function for any positive constant s_1 has the formula

$$c^{\star}_{nl}(i) = \frac{s_1 \epsilon_n (\log n)^{0.55}}{I(i \in \mathcal{S}^{\star}_l) + \frac{1}{n}} \tag{5}$$

where

$$S^{\star}_l = \left\{ i : \sum_{i_1=1}^{p_1} \sum_{i_2=2}^{p_2} \cdots \sum_{i_{\kappa-1}=1}^{p_{\kappa-1}} I\left(\mathcal{D}^2_{nl,i_1 i_2 \cdots i_{\kappa-1}}(i) > 0\right) > 0 \right\}.$$

- ϵ_n is the convergence upper bound of $|\mathcal{D}_{nl,i_1,\cdots,i_{\kappa-1}}(i) - \mathcal{D}_{l,i_1,\cdots,i_{\kappa-1}}(i)|$ over all $\{i_1, \cdots, i_{\kappa-1}, l\}$ and i.

- In the segment with no changes, $I(i \in \mathcal{S}^{\star}_l) = 0$ and $c^{\star}_{nl}(i)$ can be very large, $T_l(i)$ is close to 1; in the segment with changes, $c^{\star}_{nl}(i)$ converges to zero at a slower rate than ϵ_n, the curve pattern is shown in Figure 3 below.

Signal function

- The final signal function is

$$T^v(i) = \min_{1 \le l \le p_\kappa} T_l(i).$$

- $T^v(i)$ have different patterns in the following cases.

Plot of Case 1

Case 1: all T_l's have changes at the same locations.

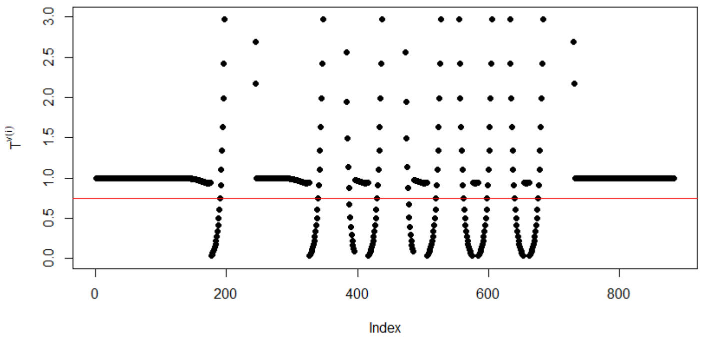

Figure 4: The plot presents the curve when all T_l's have changes at the same locations.

Plot of Case 2

Case 2: There are at least two T_l's having changes at different locations.

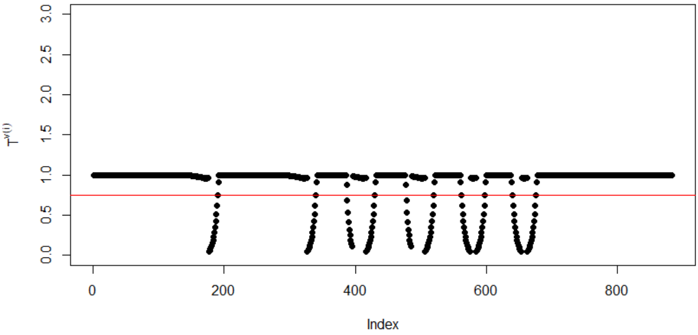

Figure 5: The solid line is for $||\mathcal{D}_1(i)||_{0s}^2$ and the dotted line for $T^v(i)$. There are at least two T_l's having changes at different locations.

The signal statistic

- The sample version of $\mathcal{D}_l(i)$ is

$$\mathcal{D}_{nl}(i) = \frac{1}{\alpha_n}\left(\sum_{j=i}^{i+\alpha_n-1}\mathcal{X}_{j,l} - \sum_{j=i+\alpha_n}^{i+2\alpha_n-1}\mathcal{X}_{j,l}\right).$$

- The sample version $T_{nl}(i)$ is

$$T_{nl}(i) = \frac{||\mathcal{D}_{nl}(i)||_s^2 + c_{nl}(i)}{||\mathcal{D}_{nl}(i+\alpha_n)||_s^2 + c_{nl}(i)}.$$

The signal statistic

- $c_{nl}(i)$ is an estimator of $c^{\star}_{nl}(i)$: for a $\nu > 1/2$ and any positive constant s_1,

$$c_{nl}(i) = \frac{s_1 \epsilon_n (\log n)^{0.55}}{I(i \in \mathcal{S}_l) + \frac{1}{n}}, \tag{6}$$

- The final signal statistic is defined as

$$T_n^v(i) = \min_{1 \le l \le p_\kappa} T_{nl}(i).$$

Inconsistency

- In the intervals where $||\mathcal{D}_l(i)||_{0s}^2$ and $||\mathcal{D}_l(i+\alpha_n)||_{0s}^2$ take small positive values, the inconsistency occurs. $T_n^v(\cdot)$ is inconsistent consequently. See Figure 5 below.
- The inconsistency causes the curve pattern of $T_n^v(\cdot)$ to be unlike that of $T^v(\cdot)$.
- We call these intervals the uncertain sets.

Inconsistency

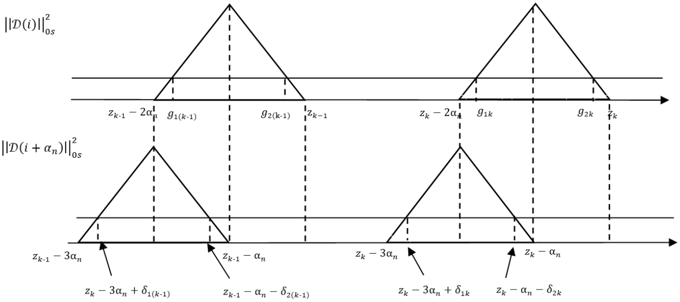

Figure 6: The plot is for the case where the distance between two true change points is not less than $3\alpha_n$.

Detection criterion

We suggest an algorithm to choose the intervals $(m_k^v(\tau), M_k^v(\tau))$ to rule out the local minimizers of the "false " changes in the uncertain sets, and detect a local minimizer converging to the corresponding minimizer of $T^v(\cdot)$ in a certain sense.

- **Choosing the disjoint intervals each containing one true change point.** We determine the disjoint intervals $(m_k^v(\tau), M_k^v(\tau))$ for $k = 1, \cdots, \hat{K}$
- For a pre-determined threshold $0 < \tau < 1$, define $M_k^v(\tau)$ satisfying

$$T_n^v(M_k^v(\tau)) < \tau, \quad T_n^v(M_k^v(\tau) + 1) \geq \tau \tag{7}$$

and let $m_k^v(\tau) = M_k^v(\tau) - \frac{2\sqrt{\tau}}{\sqrt{\tau}+1}\alpha_n$.

- In our numerical studies, we recommend $\tau = 0.8$.

Detection criterion

- **Ruling out spurious intervals.** Due to the inconsistency of $T_n^v(\cdot)$ occurs in the uncertain sets, spurious intervals might exist: that is, there might also exist some M_g's corresponding to false change points.
- Rule out those M_g's when $M_{g+1}(\tau) - M_g(\tau) \leq 3\alpha_n/2$, where $g \in \{g : T_n^v(M_g) < \tau \text{ and } T_n^v(M_g + 1) \geq \tau\}$.

- **Identifying the locations.** Identify the local minimizer $r_k := \max\{\arg\min_{i \in (m_k^v(\tau), M_k^v(\tau))} T_n^v(i)\}$ for $1 \leq k \leq \hat{K}$;
- define the estimated location as $\hat{z}_k = r_k + 2\alpha_n - 1$.

Table of Contents

Conditions on the tensor size p and heterogeneity

- Assume that $\mathcal{X}_i - E(\mathcal{X}_i)$ are independent random error tensors.

1. Only under 4-th moment of all elements of the tensor, the maximal divergence rate of $p = \prod_{l=1}^{\kappa} p_l$ can be up to n^2.
2. Under stronger conditions including sub-Gauss condition (exponential tail), the elements η_{ij} of the standardized error vectors $\boldsymbol{\eta}_i$ for $i = 1, 2, \cdots, n; j = 1, 2, \cdots, p$ are independent and identically distributed, the maximal divergence rate of p can be close to $\exp(\alpha_n^{\frac{1}{7}} - logn)$.

Consistency of change point estimation

Theorem 2

- *Under some other regularity conditions including the above, when* $\alpha_n^\star \geq 2\alpha_n$ *and* $\frac{n}{K\alpha_n^\star} \to \infty$,

$$Pr(\hat{K} = K) \to 1,$$

and the estimators $\{\hat{z}_1, \hat{z}_2, \cdots, \hat{z}_K\}$ *with* $\hat{z}_j = 0$ *for* $\hat{K} < j \leq K$*:*

$$\Pr\left\{\max_{1\leq k\leq K}\left|\frac{\hat{z}_k - z_k}{\alpha_n}\right| < \epsilon\right\} \to 1$$

for every $\epsilon > 0$.

Table of Contents

Simulations

- Recall the tensor model (Equation (1)):
  $\mathcal{X}_i = \mathcal{M}_i + \mathcal{E}_i, i = 1, 2, \cdots, n,$
- To substantiate the credibility of the real data instance, we align the sample size and dimensionality with the those observed in the authentic data scenario.
- The sample size is $n = 200$.
- The number of change points is $K = 4$
- $p_1 = 3$, $p_2 = 2$, and $p_3 = 4, 20, 40$.
- The precise locations of the change points are identified as $40, 80, 110, 140$.
- The signal statistics $T_{nl}(\cdot)$ are over the elements of mode 3 (p_3).

Order-three tensor

- *Case 1.* (Variance Heterogeneity)
  - 5 segments: $I_1 = (1, 39)$, $I_2 = (40, 79)$, $I_3 = (80, 109)$, $I_4 = (110, 139)$, $I_5 = (140, n)$.
  - All elements within the odd slices of $\mathcal{E}_i$ are independently generated from $\mathcal{N}(0, 2)$.
    All elements within the odd slices of $\mathcal{M}_i$ a uniform value of 3 when $i \in I_2 \cup I_4$, and 2.8 when $i \in I_1 \cup I_3 \cup I_5$.
  - All elements within the even slices of $\mathcal{E}_i$ are independently drawn from $\mathcal{N}(0, 1)$.
    All elements within the even slices of $\mathcal{M}_i$ are set uniformly to 0.2 when $i \in I_2 \cup I_4$, and 0.4 when $i \in I_1 \cup I_3 \cup I_5$.

Order-three tensor

Table 1: Order-three tensor with $K = 4$, $p_1 = 3$, $p_2 = 2$.

			$\hat{K} - K$						
Scenario	Mean	MSE	≤ -3	-2	-1	0	1	2	≥ 3
Variance Heterogeneity									
$p_3 = 4$	4.245	3.080	0	6	40	74	59	21	0
$p_3 = 20$	4.215	3.186	0	8	44	70	56	19	3
$p_3 = 40$	4.290	2.924	1	11	32	69	58	29	0

Order-three tensor

- *Case 2.* (Distribution Heterogeneity)
  - The setting is identical to that in Case 1 except that all elements within the odd slices of $\mathcal{E}_i$ are independently sampled from the uniform distribution $U(-2, 2)$ (against $N(0, 2)$).
- Summary: *The results are reported in Table 1 below and our method shows robust performances in low and high dimensions.)*

Order-three tensor

Table 2: Order-three tensor with distribution heterogeneity (uniform distribution and normal distribution). $K = 4$, $p_1 = 3$, and $p_2 = 2$.

			$\hat{K} - K$						
Scenario	Mean	MSE	≤ -3	-2	-1	0	1	2	≥ 3
$p_3 = 4$	4.135	3.478	1	12	48	60	56	23	0
$p_3 = 20$	4.080	3.686	2	9	44	75	56	14	0
$p_3 = 40$	4.080	3.686	2	9	44	75	56	14	0

Table of Contents

Trade data (order-three tensor) I

- We conducted an analysis of import, export, and net trade (3-dimensional mode) data for goods and services (2-dimensional mode) from the first quarter of 1979 to the fourth quarter of 2022 (176 quarters in total) in four countries: Australia, South Africa, The United Kingdom, and The Unite States (4-dimensional structural mode).
- These are order-three tensor data of dimensions $(3, 2, 4)$. The data set was obtained from `https://data.oecd.org/trade/trade-in-goods-and-services.htm`. Our primary objective was to detect significant changes during this time spanning. The four countries are regarded as a 4-dimensional structural mode.
- As a result of our analysis, we identified four change points: “1989-Q1”, “1998-Q1”, “2005-Q1”, and “2012-Q1”, where “Q1” represents the first quarter.

Trade data (order-three tensor) II

- Might the following events be close to and play crucial roles in these changes?
  - (1) (1989-Q1) In October 1987, a worldwide stock market crash, called "Black Monday," caused a big drop in global stock markets in just one day.
  - (2) (1998-Q1) In 1997, there was a big money problem in Asia. This made money in some countries there worth less, and their financial systems got messed up. It affected the whole world's economy.
  - (3) (2005-Q1) The Iraq War, commencing in 2003 and continuing in 2004, possibly affecting the import and export expenditures and net trade balances of the United States. The war's impact could have led to shifts in resource allocation and trade relations;
  - (4) (2012-Q1) The Libya War in 2011, wherein Libya's significant role as an oil-exporting nation and its oil production and exports influenced the global oil market. These events may significantly affect the international trade of these four countries.

Table of Contents

References I

Chen, Y., Wang, T., Samworth, R. J., et al. (2022).
High-dimensional, multiscale online changepoint detection.
Journal of the Royal Statistical Society Series B, **84**:234–266.

Herrera, S., Gutiérrez, J. M., Ancell, R., Pons, M., Frías, M., and Fernández, J. (2012).
Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over spain (spain02).
International Journal of Climatology, **32**:74–85.

Kulldorff, M. (2012).
Brain cancer incidence in new mexico.
http://www.satscan.org/datasets/nmbrain/index.html.

References II

Wang, D., Yu, Y., and Rinaldo, A. (2021).
Optimal change point detection and localization in sparse dynamic networks.
The Annals of Statistics, **49**:203–232.

Xie, L., Xie, Y., and Moustakides, G. V. (2019).
Asynchronous multi-sensor change-point detection for seismic tremors.
In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 787–791. IEEE.

Thanks!