

Research in Production and Operations Management University of Isfahan E-ISSN: 2981-0329 Vol. 14, Issue 4, No. 35, Winter 2024, p 49-59

Submit Date: 2023-07-17 / Accept Date: 2024-02-26 http://dx.doi.org/10.22108/POM.2024.138146.1513

do

(Research Paper)

Toward an Effective Panel Data Change Point Analysis Method

Naser Rafiee

Management and Industrial Engineering Faculty, Malek-Ashtar University of Technology, Tehran, Iran, naserrafiee67@gmail.com

Karim Atashgar*

Management and Industrial Engineering Faculty, Malek-Ashtar University of Technology, Tehran, Iran, atashgar@iust.ac.ir

Mehrdad FazlAli

Management and Industrial Engineering Faculty, Malek-Ashtar University of Technology, Tehran, Iran, mfazlal 1969@yahoo.com

Abstract

Purpose: Sometimes the performance of a process is better analyzed by panel data rather than measurements on only a time series. When a change manifests itself in the cross-section(s) of panel data, detection of the real-time change corresponding to the panel data leads practitioners to identify the responsible factor(s) that affected the distribution of the panel. This invaluable time analysis is referred to as the change point. This paper proposes a new effective method to identify the change point of the panel data case.

Design/methodology/approach: Considering cross-sectional time series, the identification of the change point of panel data has been evaluated. Two real cases have been studied to analyze the performance of the proposed method.

Findings: The comparative numerical performance analysis of different simulated cases indicates that the proposed method is more effective compared to the methods proposed in the literature.

Practical implications: The investigation of two real case studies in this paper addresses that the proposed method allows practitioners to use the proposed applied approach for analyzing different real panel cases.

Keywords: Changepoint, Panel data, Numerical analysis, Out-of-control condition

* Corresponding author, Orcid: 0000-0002-7857-9476 2981-0329/ © University of Isfahan

This is an open access article under the CC-BY-NC-ND 4.0 License (https://creativecommons.org/licenses/by-nc-nd/4.0/)



1. Introduction

Advances in data analysis have allowed engineers and practitioners to model a large number of process measurements to reconstruct panel data to evaluate its stability over time. In other words, there is in many applications, multi-dimensional data that includes time series observations of a large number of cross-sectional units. This structure of data is referred to as the panel data where one considers both sectional data and time series. The panel data approach provides the opportunity to investigate the individual effects of the cross-sectional data over time, separately.

Changepoint refers to the time when a process shifts from an in-control condition to an unexpected condition. Identification of the change point of panel data leads one to an effective root cause analysis of the process. Literature addresses that Bai (2010) used the least square error (LSE) method to estimate the common change point in means of panel data. Bai (2010) also used the quasi-maximum likelihood (QML) method to estimate the change point in mean, variance, and both. Horváth and Hušková (2012) investigated the statistics of the change point estimator proposed by Bai (2010) and proposed a test based on the likelihood approach. Horváth & Hušková (2012) used their proposed method to identify the change point of the Gini coefficient for the panel data of 33 countries including European countries, Australia, the United States, South America, China and Taiwan. Li et al. (2015) developed the statistic of Horváth and Huşková (2012) and proposed a new statistic based on the cumulative sum (CUSUM) method to identify the common change point of the variance in panel data. Chen and Hu (2017) focused on using CUSUM to estimate the mean change point of panel data. The estimator proposed by Chen and Hu (2017) is less complex as it is more accurate compared to the LSE method proposed by Bai (2010). Pestova and Pesta (2015) proposed a test to identify a common change point in means of panel data using Bootstrap term. Their proposed statistic is based on the CUSUM method. Pestova and Pesta (2017) also proposed a common change point estimator using panel data based on the LSE method. Maciak et al. (2018) proposed a method using CUSUM statistics. The three recent studies used claim amounts paid by insurance companies to evaluate the capability of the proposed methods. De Wachter and Tzavalis (2012) proposed a test to detect the change point in the structure of a dynamic linear panel data model. De Wachter and Tzavalis (2012) approached the GMM framework. Zhu et al. (2013) proposed a method for the detection of the change point in the case that unbalanced panel data follows a dependence structure. In this approach, the copula method is used to describe the dependency structure of the panel data. Zhu et al. (2013) used their proposed method to identify the start and the end of the credit crisis in Chinese banking. Enomoto and Nagata (2016) developed the Mahalanobis – Taguchi (MT) method proposed by Taguchi (2002) using Bayes inference to identify the change point in panel data. They used an annual beverage consumption case in Japan to evaluate the performance of the proposed method. Cho (2016) proposed the Double CUSUM (DC) statistic to identify the change point in panel data. Cho (2016) used financial data sets from stock prices of S & P 100 index components as a real case study to evaluate the capability of the proposed method. Atashgar and Rafiee (2019) using the exponentially weighted moving average (EWMA) and double CUSUM statistic proposed Double CUSUM-EWMA (DCE) statistic to detection of the change point in the panel data means. They showed the superiority of their proposed method compared to the Cho (2016) method. Atashgar et al. (2022) proposed a method for detecting the change point in panel data approaching a hybrid statistic called Double CUSUM-Modified EWMA (DCME). Atashgar et al. (2022) showed numerically higher sensitivity to their proposed method compared to the methods proposed by Cho (2016) and Atashgar and Rafiee (2019). Recently also, Atashgar and Rafiee (2020) investigated the annual car production in the world using change point analysis in the panel data. They expressed that the change point analysis has the capability of evaluating strategic issues, effectively.

This study attempts to propose a new method with high sensitivity compared to the methods proposed in the literature for detecting the common change point location in panel data.

The remainder of this paper is structured as follows; the next section introduces the concept of the change point issue and the importance of its identification. The third section is allocated to describe the research methodology and the proposed method for detecting the location of the common change point in the panel data. The fourth section provides the capability of the proposed method by investigating two real case studies. The fifth section is allocated to the findings of this research. Section 6 compares the capability of the proposed method numerically considering several different cases. Finally, the last section is dedicated to conclusions.

2. The concept and the importance of the change point

Assume y_{it} is panel data including independent observations, where $1 \le i \le N$, $1 \le t \le T$. In this case, N and T denote the number of cross-sectional units and the length of the time series for each cross-sectional unit, respectively. When a panel data of a process works under common causes, the values of the panel vary over time under a known normal distribution. Equation 1 indicates the case in the panel of the process is produced without affecting any special cause. This case is referred to as the in-control condition of the process panel data.

$$y_{it} = \mu_i + \epsilon_{it}; \quad i = 1, 2, ..., N \quad t = 1, 2, ..., T$$
 (1)

In Equation 1, μ_i denotes the mean of the cross-section *i* under the in-control condition, ε_{it} addresses the process error in the cross-section *i* and time *t* so that $E(\varepsilon_{it}) = 0$. The panel error random variable is independent and identically distributed in each cross-section of the panel. Now assume that the panel data affected by a sustained special cause(s) shifts to an unnatural status. In this case, the values of the panel do not follow the known distribution described for Equation 1. Let the expected value of a cross-section(s) of the panel data of the process after an unknown time τ (affected by an unnatural factor) shift to a new value and then the values

of the panel are produced with a new mean parameter. This unnatural condition of the process panel data can be described mathematically as the following equation 2:

$$y_{it} = \mu_i + \delta_i | \{T > \tau\} + \epsilon_{it}, \qquad 1 \le i \le N, 1 \le \tau \le T \tag{2}$$

In Equation 2, τ indicates, the time of possible change, and $\delta_i \neq 0$ addresses the change value in mean μ_i after an unknown time of $\tau \in [1, T)$. Time T indicates the time when practitioners are allowed to identify an unnatural condition of the process. In other words, T refers to the time of the current vector observation and practitioners are led to conclude that a special cause(s) has taken place in the process panel data and led the process to a departure from the in-control condition to an out-of-control condition. Figure 1 shows the description of one cross-section of the panel data. Time τ is referred to as the change point. This means that before a method triggers the new condition, a change has taken place in the process panel data and practitioners couldn't detect it instantly. In other words, the change point identification process will activate after the time that practitioners conclude an out-of-control condition for the process. As shown in Figure 1, between the change point τ and time $\tau + 1$ a special cause affected the mean of the cross-section and remained in the process panel data until time T when practitioners are allowed to identify the new condition of the process. In this case, a root-cause analysis should be started by practitioners to identify and eliminate the special cause(s) that manifested itself in the panel data of the process. A good estimate of the change point leads engineers to start with a good searching time point of the special cause and eliminate effectively the cause that shifted the panel data in an out-of-control condition.



Fig. 1: change point for a cross-section of a panel data

3. Research methodology and the proposed method

Let y_t for $t=1,\ldots,T$ denotes a random variable of time series. Assume a change point occurs at an unknown time τ . In this case, y_t follows a common distribution function F_1 for $t=1,\ldots,\tau$, and it follows a common distribution function F_2 for $t=\tau+1,\ldots,T$, where $F_1\neq F_2$. Let U_t be defined for $t=1,\ldots,T$ as follows:

$$U_{t} = \sum_{j=1}^{t} \sum_{k=t+1}^{T} sgn(y_{k} - y_{j})$$
 (3)

where the function sgn(x) is defined as the follow:

$$sgn(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$
 (4)

Pettitt (1979) proposed a nonparametric test to identify the change point in the mean of time series. In this approach, to identify the presence of the change point over a given interval [1, T], the test statistic is defined as the following equation:

$$S_{\rm T} = 2 \times e^{\frac{-6k^2}{T^3 + T^2}} \tag{5}$$

where k is obtained as follows:

$$k = \max_{1 \le t < T} |U_t| \tag{6}$$

Based on Atashgar and Rafiee (2020) it is possible to use Monte Carlo simulation to access the test criterion. Once the test criterion is equal to π_T and $S_T < \pi_T$, the null hypothesis (i.e. no change has occurred) is rejected and the location of the change point can be estimated as Equation 7:

$$\hat{\tau} = \arg \max_{1 \le t < T} |U_t| \tag{7}$$

Based on the definition of Equation 2, the condition of Pettitt (1979) model is developed for the process panel data case. Assuming y_{it} is defined for i = 1,..., N and t = 1,..., T, as well as U_{it} is defined for i = 1,..., N and t = 1,..., T the following equation can be written:

$$U_{it} = \sum_{j=1}^{t} \sum_{k=t+1}^{T} sgn(y_{ik} - y_{ij})$$
 (8)

In this case, to detect the location of the common change point in the panel data Equation 9 is proposed.

$$PN = \max_{1 \le t < T} \sum_{i=1}^{n} |U_{it}| \cdot \mathbb{I}(S_{T}^{i} < \pi_{T}^{i})$$
(9)

where S_T^i is the test statistic of Equation (5) and π_T^i is the test criterion for time series observations of cross-section *i*. If $S_T^i < \pi_T^i$, then $\mathbb{I}(S_T^i < \pi_T^i) = 1$ and otherwise its value is zero. Therefore, assuming the presence of a change point in the panel data, the location of the common change point is identified as Equation 10.

$$\hat{\tau}^{PN} = \arg \max_{1 \le t < T} \sum_{i=1}^{n} |U_{it}| \cdot \mathbb{I}(S_T^i < \pi_T^i)$$
 (10)

4. Case studies

In this section to analyze the performance of the proposed method, two real cases are considered. The reports in this section show the capability of the proposed method to identify the change point of the process panel data.

4.1 Panel data change point analysis for a strategic monitoring

This section analyzes the change point identification for a real panel case to investigate the strategic terms of a holding company. The results of the real case of this section indicate that the proposed model of this paper is capable of performing an effective strategic analysis. The data in this section correspond to 4 important variables of 5 industries related to an Iranian holding company during the years 2010 to 2019. The terms of this case study are 1) total

sales (TS) in Iranian currency, 2) some supply chain companies (SCC), 3) the number of completed development projects (CDP) per year, and 4) time to turn an idea to a product (TIP) in a month. These four important variables are considered in the process of the monitoring strategic plan by the holding company.

The analysis using DCE and DCME methods addresses the presence of a common change point for the panel data. The analysis indicates the existence of change points for the variables except for CDP (see Tables 1, and 2). In other words, the analysis of the panel data change point indicates that the change point has occurred in TS, SCC and TIP variables. Tables 1 and 2 show the results of the analysis. The second and third columns of Tables 1 and 2 indicate the test statistic and the test criteria, respectively. Once the test statistic is larger than the test criteria, the fourth column of Tables 1 and 2 shows the estimated location of the common change point (τ) . Furthermore, the fifth column indicates the density of the industries affected by the estimated change point (m). Based on these two methods the 6^{th} year is identified for the change point of the panel data for TS, SCC and TIP variables, i.e. year 2015.

Table 1: Analysis of the common change point using the DCE method

Variable	S _{N,T} ^{DCE}	$\pi_{ ext{N,T}}$	$\hat{ au}^{DCE}$	m
TS	10.24	5.19	6	5
SCC	74.01	4.34	6	5
CDP	7.05	10.76		no change
TIP	9.68	5.20	6	5

Table 2: Analysis of the common change point using the DCME method

Variable	S _{N,T} ^{DCME}	$\pi_{ ext{N,T}}$	$\hat{ au}^{ extsf{DCME}}$	m
TS	10.67	5.07	6	4
SCC	77.10	4.37	6	5
CDP	8.01	11.56		no change
TIP	13.02	4.99	6	2

The analysis of the change point for the real panel data using the proposed model of this paper addresses $\hat{\tau}^{PN} = 5$. It means that the common change point for the three variables was triggered in the year 2014. Table 3 shows the results of the change point analysis.

Table 3: Location of the common change point

Variable	Statistic value of PN	Common change point location
TS	119	2014
SCC	120	2014
TIP	122	2014

To perform a strategic root cause analysis, the change point status for each industry of the holding company is considered. In this consideration, the retrospective time series data corresponding to the industries are analyzed separately to identify the change point for each industry. In this step, the change point detection method in the time series proposed by Pettitt (1979) is used. Table 4 shows the analysis results for *TS*, *SCC* and *TIP* terms. The last

column shows the increase/ decrease of the mean value of the variables after the detected change point.

Variable	Industry	Test statistic	Test criteria	Changepoint location	Change status	
_	1	0.07	0.18	2014	increase	
	2	0.09	0.18	2015	increase	
TS	3	0.07	0.18	2014	increase	
	4	0.07	0.18	2014	increase	
	5	0.07	0.14	2014	increase	
	1	0.07	0.18	2014	increase	
	2	0.09	0.18	2013	decrease	
SCC	3	0.07	0.18	2014	increase	
	4	0.07	0.18	2014	increase	
	5	0.09	0.18	2014	decrease	
	1	0.07	0.18	2014	decrease	
	2	0.07	0.18	2014	decrease	
TIP	3	0.09	0.18	2015	decrease	
	4	0.09	0.18	2013	decrease	
	5	0.07	0.18	2014	decrease	

Table 4: Results of the change status for each industry

The change point analysis of the above case study leads one to the following conclusions:

- (1) As shown in Figure 2, the total sale (TS) term affecting the responsible factor(s) increases after the identified change point of the industries.
- (2) After the identified mean change point corresponding to SCC, the mean observations of all 5 companies follow a different trend. Figure 3 addresses graphically the values after the detected change point.
- (3) Figure 4 indicates that the mean value of TIP for all 5 industries decreases, after the change point.



Fig. 2: Time series trend of the TS variable related to different industries

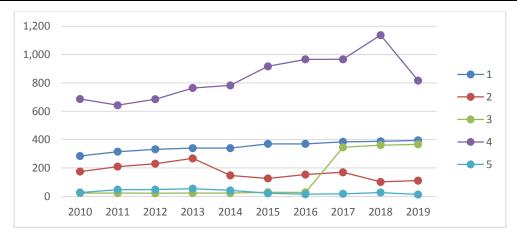


Fig. 3: Time series trend of the PCN variable related to different industries

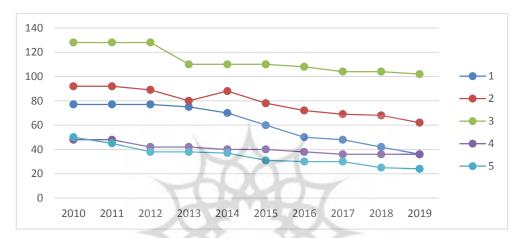


Fig. 4: Time series trend of the ITTP variable related to different industries

The investigation of the performance reports of the industries from 2006 to 2014 considering the change point analysis, indicates that executive policies of the strategic plan have been changed affecting some important factors such as turnover of management.

4.2 Automotive manufacturing change point identification

In this section, a real case corresponding to car manufacturing studied by Atashgar and Rafiee (2020) is considered. In this case, the number of cars manufactured by different companies around the world in 19 years (in the time interval of 2000 to 2018) is analyzed based on the change point concept. The countries considered in this evaluation are listed in Table 5. Based on the analysis of the proposed method of this paper, the location of the common change point for the automotive manufacturing case is estimated in 2008. Figure 5 indicates the result of the evaluation in obtaining the change point location in the case of panel data of the automotive industry. As shown in Figure 5 in the region of the maximum value the cure is relatively flat and it is not very peaked at the maximum value. This evaluation addresses the location of the change point and it is invaluable information for economic analysis. The automotive industry crisis in 2008-2010 was a part of the financial crisis of 2007-2008 and the great recession has been reported [11]. This crisis was reported for American industries and then it affected European, Canadian, and Asian industries.

No.	Country	No.	Country	No.	Country	No.	Country
1	Argentina	11	Germany	21	Portugal	31	Thailand
2	Austria	12	Hungary	22	Romania	32	Turkey
3	Belgium	13	India	23	Russia	33	UK
4	Brazil	14	Indonesia	24	Serbia	34	Ukraine
5	Canada	15	Iran	25	Slovakia	35	USA
6	China	16	Italy	26	Slovenia	36	Uzbekistan
7	Czech Rep.	17	Japan	27	South Africa	37	Others
8	Egypt	18	Malaysia	28	South Korea		
9	Finland	19	Mexico	29	Spain		
10	France	20	Poland	30	Taiwan		

Table 5: The considered countries of the evaluation

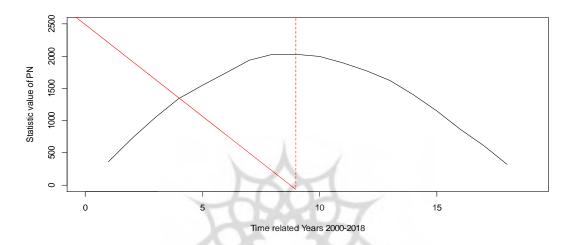


Fig. 5: the change point location in the case of the automotive industry

5. Findings

Identification of the change point in a panel data case is an important step for practitioners, in the case that the process works under an out-of-control condition. An identified change point allows one to analyze and identify the source(s) that affected the process at the time when the process has been shifted to an unnatural condition. In this paper, a new statistic (as defined in Equation 10) is proposed to lead practitioners to find the change point statistically.

ربال حامع علوم ال

6. Discussion

The above analysis indicated that the proposed method is capable of detecting the change point effectively. To analyze the performance of the proposed method using a comparative approach, a low-dimension (10×10) process is simulated. In this analysis, it is assumed that the panel data is affected by a step shift after the change point. In this section, three models of the literature including Cho (2016), Atashgar and Rafiee (2019) and Atashgar et al. (2022) are compared with the performance of the proposed model of this paper. The simulated data follow the ARMA (2, 2) model, as Equation 11:

$$\varepsilon_{it} = 0.2\varepsilon_{it-1} - 0.3\varepsilon_{it-2} + u_{it} + 0.2u_{it-1}$$
(11)

where
$$u_{it} = \sum_{k=0}^{99} \varrho_k v_{i-k,t}$$
 (12)

and $v_{it}\sim_{iid} N(0,\sigma_v^2)$, $\sigma_v=0.1\varrho^{-1}$, $\varrho_k=\varrho(k+1)^{-1}$ and $\varrho\in\{0.2,\ 0.5\}$. It is noted that ϱ adjusts the correlation degree of the cross-section.

Assume that all the cross-sections of the panel are affected by a step change at time τ . The change size for each cross-section follows the uniform function U (0.75, 1.25) with ratio $\delta \in \{0.1, 0.2, 0.3\}$. In this evaluation, according to Atashgar et al. (2022), $\lambda = 0.6$ and $k = \lambda / 2$ values are considered. In this analysis, the simulation is iterated 1000 times for each change value. To evaluate the location accuracy indicator of the change point, $|\hat{\tau} - \tau| < 1$ is considered, where $\hat{\tau}$ indicates the estimated change point τ . Table 6 shows the results of the analysis. Table 1 compares numerically the accuracy of the change point location detection for the four methods, i.e. DC, DCE, DCME, and PN (the proposed method of this paper). The analysis of Table 6 leads one to conclude the following results:

- (1) When the common change point manifests itself into the panel near the beginning and the end time points of the time interval [1, T], the location estimation accuracy of the change point by the proposed PN statistic of this paper is much better compared to the other three models.
- (2) When the real common change point occurs in the middle of the time interval [1, T], the location accuracy of the change point estimated by all four models improves as the size of the change is increased.
- (3) The reports indicate that the capability of improving the proposed model PN is superior compared to the other three models.

	$\varrho = 0.2$				$\varrho = 0.5$				
δ	τ	DC	DCE	DCME	PN	DC	DCE	DCME	PN
0.1	0.2T	0	0	0	215	0	0	0	213
	0.5T	429	401	429	371	417	389	409	346
	0.8T	0	0	0	217	0	0	0	192
0.2	0.2T	0	0	0	444	0 -	0	0	424
	0.5T	688	621	689	697	694	611	686	698
	0.8T	0	0	0	417	0	0	0	413
0.3	0.2T	0	0	0	554	0	0	0	556
	0.5T	844	721	831	880	880	748	884	885
	0.8T	0	0	0	532	0	0	0	553

Table 6: Estimating the location of the change point for the panel data in 1/1000 accuracy

7. Conclusions

A precise change point identification allows practitioners to conduct an effective root cause analysis and remove the unnatural cause that affected the process. Change point analysis is started after the process panel data has been affected by a special cause and has shifted to an unnatural condition. Hence identifying the change point of a panel data is evaluated as an important issue in process management and data analysis issues. In this research, a new statistic is proposed for detecting the location of the change point in the panel data case. The proposed method allows practitioners to estimate the change point effectively. The comparative numerical performance analysis indicated that the capability of the proposed method is superior compared to the existing models of the literature. The investigation of two

real panel case studies in this paper indicated that the proposed method is an effective approach to analysing different panel cases.

References

- Atashgar, K., and Rafiee, N. (2019). Identifying the change point of panel data using simultaneously EWMA and CUSUM methods. *Journal of Industrial Engineering*, 52 (4), 471-481. https://doi.org/10.22059/JIENG.2019.272591.1661
- Atashgar, K., Rafiee, N. and Karbasian, M. (2022). A new hybrid approach to panel data change point detection. *Journal of Communications in Statistics Theory and Methods*, 51(5), 1318-1329. https://doi.org/10.1080/03610926.2020.1760298
- Atashgar, K., and Rafiee, N. (2020). Identification of the Automotive Manufacturing Change Point Approaching Panel Data. International Conference on Industrial Engineering and Operations Management Dubai, UAE.
- Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics*, 157 (1), 78–92. https://doi.org/10.1016/j.jeconom.2009.10.020
- Chen, Z., and Hu, Y. (2017). Cumulative sum estimator for change-point in panel data. Statistical Papers, 58 (3), 707–728. https://doi.org/10.1007/s00362-015-0722-y
- Cho, H. (2016). Change-point detection in panel data via double CUSUM statistic. Electronic Journal of Statistics, 10(2), 2000–2038. https://doi.org/10.48550/arXiv.1611.08631
- De Wachter, S., and Tzavalis, E. (2012). Detection of structural breaks in linear dynamic panel data models. Computational Statistics & Data Analysis, 56(11), 3020–3034. https://doi.org/10.1016/j.csda.2012.02.025
- Enomoto, T., and Nagata, Y. (2016). Detection of change points in panel data based on the Bayesian MT method. Total Quality Science, 2(1), 36–47. https://doi.org/10.17929/tqs.2.36
- Horváth L., Hušková, M. (2012). Change-point detection in panel data. Journal of Time Series Analysis, 33(4), 631–648. https://doi.org/j.1467-9892.2012.00796.x
- Li, F., Tian, Z., Xiao, Y., and Chen, Z. (2015). Variance change-point detection in panel data models. Economics Letters, 126, 140–143. https://doi.org/10.1016/j.econlet.2014.12.005
- Maciak, M., Pestova, B., and Pesta, M. (2018). Structural breaks in dependent, heteroscedastic, and external panel data. Kybernetika, 54, (6), 1106–1121. https://doi.org/10.14736/kyb-2018-6-1106
- Pestova, B., and Pesta, M. (2015). Testing structural changes in panel data with small fixed panel size and bootstrap. Metrika, 78(6), 665–689. https://doi.org/10.48550/arXiv.1509.01291
- Pestova, B., and Pesta, M. (2017). "Change point estimation in panel data without boundary issue". Risks, 5(1), 1-22. https://doi.org/10.3390/risks5010007
- Pettitt, A.N. (1979). A non-parametric approach to the change-point problem. Applied Statistics, 28(2), 126-135. https://doi.org/10.2307/2346729
- Taguchi, G. (2002), Technological development in the MT system. Japan Standards Association. (In Japanese)
- Zhu, X., Li, Y., Liang, C., Chen, J. and Wu, D. (2013). Copula-based change point detection for financial contagion in Chinese banking. Information Technology and Quantitative Management, 17, 619–626. https://doi.org/10.1016/j.procs.2013.05.080