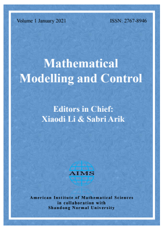
Mathematical Modelling and Control

https://www.aimspress.com/journal/mmc



MMC, 4(4): 459–469.
DOI: 10.3934/mmc.2024036
Received: 22 August 2024
Revised: 07 September 2024
Accepted: 13 September 2024
Published: 20 December 2024

Research article

On a data model associated with antitrust behaviors

Weiguang Zhou[1], Guangxin Wang[1], Jixiao Lu[1], Hongxin Ruan[2], Jun Wang[1] and Rui Zhang[3,*]

[1] Shandong Monitoring Center for Market Regulation, Jinan 250013, China

[2] Information Office, Yankuang Energy Group Company Limited, Zoucheng 273599, China

[3] School of Mathematics and Information Science, Shandong Technology and Business University, Yantai 264005, China

* **Correspondence:** Email: zhangruimath@163.com.

Abstract: In this paper, we established a big data model based on the data analysis method and the endogenous structure mutation theory, and judged from the dimensions of time and space. Additionally, we gave a detailed inspection and analysis among enterprises producing cement of a certain province. The results indicated that this model can identify monopolistic behaviors from multiple dimensions and, thus, improve regulatory efficiency.

Keywords: data analysis; big data model; structural mutation; antitrust behaviors

1. Introduction

With the development of economy, it is of great significance to prevent unfair competitions such as antitrust behaviors. At present, there is a lack of knowledge about the identification of monopoly agreement due to the fact that there exists a variety of industries, enterprises, and products.

The characteristics of product price fluctuations in the market are complex. The changes in external conditions such as rising cost prices, policy changes, epidemics, and financial crises all exhibit sudden changes in price data. For the detection of mutation points, Bai and Perron proposed the Bai-perron structural mutation test method and the sequential test method in [1, 2], respectively. These two methods are based on the minimal residual sum of squares based on the principle of dynamic programming. When detecting structural mutation points, data and errors with different structures can be identified, so as to confirm the date of structural mutation within the confidence interval, that is, to build a test by determining the critical value through the different distribution of data.

Bollerslev [3] proposed the generalized autoregressive conditional heteroskedasticity (GARCH) model, which is able to characterize the volatility characteristics in financial time series, such as "peak fat tail" disribution, volatility agglomeration, volatility persistence, etc. Subsequently, many scholars developed some methods for mining outliers in the GARCH model (for example, see [4, 5]). If there are sudden changes in product prices, a price time series GARCH model can be constructed based on the intervals divided by the sudden changes, and endogenous structural breakpoint testing can be performed. When the GARCH model shows that product prices have volatility clustering, it often means that market price fluctuations are influenced by external shocks and previous period fluctuations, and have volatility memory properties. Therefore, we can determine the corresponding price change period.

Various mathematical models play important roles in many research areas. In 2023, Yavuz et al. [6] established a mathematical model for the tuberculosis epidemic under the consciousness effect. In 2024, Chavada et al. [7] presented a mathematical model for smoking-related cancer. In this paper, we focus on the mathematical model of economic behaviors. We apply the big data model and machine learning algorithm to build an "antitrust big data perception model", which integrates dynamic monitoring, scientific analysis, risk early warning, decision-making assistance, and other functions. Based on actual investigation and verification, analysis results of our model are basically consistent with the actual situation.

The organization of this paper is as follows: In Sections 2 and 3, we give the establishment and verification of our model, and the conclusions, respectively.

2. The establishment and verification of our model

To determine the monopolistic behavior of enterprises on product prices, in this paper, based on the theory of structural mutation, we study the heterogeneity of product price fluctuation characteristics from dimensions of time and space.

In the time dimension, we first establish a time series model of enterprise product prices, apply endogenous multiple mutation testing to conduct structural mutation testing on price fluctuations, dynamically search and identify structural change points of product price data, thereby revealing the aggregation and risk of product price fluctuations in different sample periods, and further establish a GARCH model to explore the heterogeneity of product price fluctuation characteristics. Meanwhile, the fluctuation of the enterprise product price is not only influenced by internal factors, but may also be influenced by other enterprise product price. Therefore, it is necessary to construct a spatial econometric model in the spatial dimension to study the spatial correlation, aggregation, and heterogeneity of product prices among different enterprises. Then, we examine the spatial correlation, aggregation, and heterogeneity of product prices in both the "price stability period" and "price fluctuation period" from a holistic and local perspective in order to reveal whether enterprises have behaviors such as "monopoly", "collusion", and "price leadership follow".

Now we apply above methods to identify whether there are monopolistic behaviors among 33 enterprises producing cement in a certain province in China. For convenience, in the following, we denote the names of these 33 enterprises by ENTi, where $i = 1, 2, 3, \ldots, 33$. Due to the fact that the research objects have many feature variables, we first divide the research objects into different groups so that the research objects in the same group have similar features. As an unsupervised machine learning process in statistical data analysis, cluster analysis can automatically classify the research objects according to their respective characteristics without prior knowledge. In particular, K-means is a clustering algorithm which discovers K clusters of a given dataset.

2.1. *Enterprise clustering*

By the K-means clustering algorithm, we conduct cluster analysis on these 33 cement producing enterprises according to mill production capacity, cellar production capacity, and annual sales indicators. Before doing this, we first introduce the K-Nearest Neighbor algorithm. For the sake of simplicity, here we abbreviate the K-Nearest Neighbor algorithm to the KNN algorithm. The KNN algorithm is one of the important methods in data processing. With the help of the KNN algorithm, we can approximate each sample of the data by its KNNs.

Since the missing indicator data could effect the results of clustering analysis and all three indicators are not missed simultaneously, thus we apply the KNN algorithm to approximate the initial data before using the clustering algorithm by a similar argument to Batista and Monard [8]. The KNN algorithm is a classification algorithm that does not require training parameters. Therefore, when dealing with the problem of missing values in multiple indicators, the KNN algorithm is generally used to approximate the missing data by using the mean of the closest sample point to the missing sample. The specific steps are as follows:

(1) Apply Matlab to filter out abnormal indicator data;

(2) Apply the other two indicators of the sample point and calculate the Euclidean distance from other points. Select the two closest approximation points (x_1, y_1) and (x_2, y_2) to the sample point, that is, the parameter $k = 2$;

(3) Take the mean of the nearest two approximate points as the approximate estimation of missing data.

The calculate results of the KNN algorithm are in the following Table 1, where red data is the filled data.

Table 1. The calculate results of KNN algorithm.

Order number	Pseudonym enterprise	Mill capacity	Cellar capacity	Average annual sales share
1	ENT24	109.2	465.2	4214433.54
2	ENT28	83.2	412	1071451.74
3	ENT32	83.2	156.7	573049.95
4	ENT14	41.6	156.7	352137.45
5	ENT15	41.6	156.7	321374.23
6	ENT16	90.6	156.7	283913.02
7	ENT17	41.6	156.7	183234.23
8	ENT18	44.8	339.5	1831714.85
9	ENT19	45.5	156.7	579466.09
10	ENT20	49	156.7	279711.79
11	ENT22	91.2	532	1335535.18
12	ENT23	45.5	156.7	200609.68
13	ENT9	76.7	691.2	8235664.68
14	ENT27	83.2	339.5	1676329.33
15	ENT30	41.6	156.7	119906.53
16	ENT33	83.2	156.7	108544.07

Next, we take *K*-means clustering analysis on all cement producing enterprises in the province according to preprocessed indicators of mill capacity, cellar capacity, and average annual sales shares. The specific steps are as follows:

(1) Cluster all enterprises based on the above three indicators, set the model parameter $K = 2$, that is, divide the enterprises into two categories, and appropriately select the initial centers of the two categories;

(2) Calculate the distance between the sample points composed of three indicators of the company and the centers of the two categories, and divide the company closest to the two categories into two categories;

(3) Recalculate the center points of two categories as the new class center;

(4) Repeat above steps (2) and (3) until the clustering center no longer changes, and then the algorithm stops and outputs the results.

The calculate results of *K*-means clustering analysis are in the following Table 2.

Table 2. Calculate results of K-means clustering analysis.

Enterprise group	Pseudonym enterprise	Quantity
1	ENT1 ~ ENT12	12
2	ENT13 ~ ENT33	21

Based on the grouping results, we calculate the standardized mill capacity, cellar capacity, and average annual sales shares of these two groups. Then, we can determine the enterprises that need to be focused on according to the average of these three indicators. The calculation results are as in the Table 3 below. In Table 3, the SMC, SCC, and SAASS are the abbreviations of standardized mill capacity, standardized cellar capacity, and standardized average annual sales shares, respectively.

Table 3. Standardized data.

Enterprise group	SMC	SCC	SAASS
1	0.584027778	0.790920608	0.474523412
2	0.39525463	0.418913054	0.099185023

By analyzing the mill capacity, cellar capacity, and average annual sales shares, these enterprises are clustered into two categories: high impact group and low impact group. Being compared to the low influence group, the enterprises in the high influence group are more likely to have "collusion" suspicion. Therefore, we focus on observing and testing enterprises within the high impact group.

The visualization results after grouping are shown in the following Figure 1, where the *x*-axis represents the cellar capacity, the *y*-axis represents the mill capacity, and the size of bubbles represents average annual sales shares.

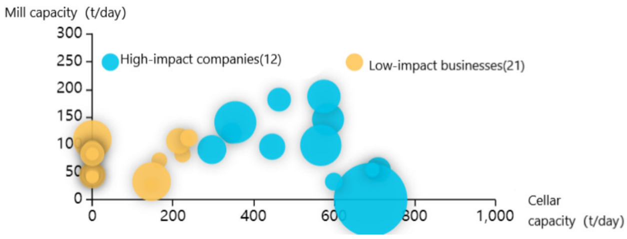


Figure 1. Legend of the figure.

2.2. Analysis from time dimension

In the time dimension, we establish a price change period identification model. In detail, a price trend chart is generated for the product prices of all enterprises based on changes in price trends. Then, we apply the endogenous structural mutation method to conduct structural mutation by testing price changes and dynamically searching for mutation points. If there are mutation points, we then establish a GARCH model of product price time series based on the intervals divided by the mutation points to measure the clustering characteristics of product price fluctuations.

We use Python's ruptures package to detect structural mutation points, and apply dynamic programming and auto-regressive loss function, which is a built-in function of Python, for implementation. The following Table 4 is the structural mutation points of the output of ruptures.

Table 4. Structural mutation points.

Order number	Structural mutation
1	27 Dec. 2015
2	30 May 2016
3	11 Dec. 2017
4	27 Oct. 2018
5	31 May 2019

For each interval divided by structural mutation points, we establish a GARCH (1,1) model based on price time series for clustering analysis. The similar applications can be found in [9–11]. The following Tables 5 and 6 are the GARCH model results for each interval divided by structural mutation points.

Table 5. Results of GARCH (1,1) model: part 1.

Order number	Period	Stationarity test	Correlation test	ARIMA model	ARCH test
1	6 May 2015 to 27 Dec. 2015	✓	✓	✓	✓
2	27 Dec. 2015 to 30 May 2016	✓	×	–	–
3	30 May 2016 to 11 Dec. 2017	✓	✓	✓	✓
4	11 Dec. 2017 to 27 Oct. 2018	✓	✓	✓	✓
5	27 Oct. 2018 to 31 May 2019	✓	✓	✓	✓

Table 6. Results of GARCH (1,1) model: part 2.

Order number	Period	GARCH model		
		α and β	coef	$P > \|t\|$
1	6 May 2015 to	α	0.1805	0.0296
	27 Dec. 2015	β	0.3410	0.0008
2	27 Dec. 2015 to	-	-	-
	30 May 2016	-	-	-
3	30 May 2016 to	α	0.2483	0.0097
	11 Dec. 2017	β	0.5138	0.0002
4	11 Dec. 2017 to	α	0.1012	0.0552
	27 Oct. 2018	β	0.8735	0.0000
5	27 Oct. 2018 to	α	0.2921	0.0691
	31 May 2019	β	0.4929	0.0307

In this model, we take the condition that

$$\alpha + \beta > 0.7$$

as a threshold of the sum of model coefficients close to 1. From the values in Table 6, we can get the following three decisions:

(1) 6 May 2015 to 27 Dec. 2015,

$$\alpha + \beta = 0.5215.$$

The interval coefficient sum is not close to 1, which indicates that the conditional variance fluctuation has no sustained memory.

(2) 27 Dec. 2015 to 30 May 2016. The correlation test in this period does not pass, which indicates that the sequence has no statistic significance.

(3) 30 May 2016 to 11 Dec. 2017,

$$\alpha + \beta = 0.7621;$$

and 11 Dec. 2017 to 27 Oct. 2018, $\alpha + \beta = 0.9747$; 27 Oct. 2018 to 31 May 2019, $\alpha + \beta = 0.7850$.

In these three periods, the β coefficient is large, and has passed the significance test, which indicates that the price fluctuation has long-term memory, that is, the past price fluctuation is a long-term price fluctuation and has a certain impact. The α coefficient is greater than 0, which indicates the price has volatility aggregation. $\alpha + \beta$ approaches 1, which indicates that conditional variance fluctuations have sustained memory, and in the market, a shock at a certain moment will have a sustained effect. This effect is difficult to be eliminated in a short period of time.

In summary, the results of the price fluctuation period identification model indicate that in the periods 30 May 2016 to 11 Dec. 2017, 11 Dec. 2017 to 27 Oct. 2018, and 27 Oct. 2018 to 31 May 2019, the price shows a volatility aggregation effect, and this volatility has a long-term memory. Thus, these three periods are price change periods.

2.3. Analysis on price and cost

Changes of cost are possible causes of changes of price. Thus, it is necessary to exclude the impact of costs before determining a monopoly by the enterprise. We establish the price-cost resonance model, and then judge whether the cost change is one of the reasons of the price mutation through the Granger causality test. By a similar argument to the works [12,13], we carefully check whether the trend of price changes is consistent with the trend of cost changes.

2.3.1. Test of Granger causality

To eliminate the time series mutation caused by factors of increased prices caused by increased cost, and to avoid solely judging monopolistic behavior of enterprises according to structural mutations in prices, a price-cost resonance model is established.

We study the price and cost of 6 enterprises: ENT2, ENT6–ENT8, ENT10, and ENT11, and find that there is a certain correlation between the trends of price and cost changes. We dispose the price and cost series of these 6 companies by the first order differences to eliminate the trend effect. After the series was stable, we performed the Granger causality test on this stationary series. The test results are shown in the following Table 7.

Table 7. Results of the granger causality test.

Order number	Enterprise	Test statistic	Critical value	p-value	df
1	ENT2	9.521e-06	5.991	1.000	2
2	ENT6	1.094e-05	5.991	1.000	2
3	ENT7	1.392e-06	5.991	1.000	2
4	ENT8	6.896e-06	5.991	1.000	2
5	ENT10	1.026e-06	5.991	1.000	2
6	ENT11	3.094e-06	5.991	1.000	2

The p-values of the test results are significantly greater than 0.05, which indicates that the original hypothesis is not rejected at a 95% confidence level, and there is a causal relationship between cost and price. Therefore, the reasons for the price fluctuation periods of ENT2, ENT6–ENT8, ENT10, and ENT11 include cost variation factors. Before determining a monopoly, it is necessary to eliminate the cost factor in the price fluctuations.

2.3.2. Elimination of cost factors in price fluctuations

To eliminate the impact of cost factors on prices, before using spatial dimension analysis, the cement price of the enterprise is subtracted from the cost to obtain the profit of the enterprise. Further analysis of profit correlation and grey correlation is conducted to determine whether the enterprise is suspected of collusion. We analyze the profits of 6 enterprises: ENT2, ENT6–ENT8, ENT10, and ENT11, and find that the profits of these 6 enterprises rise almost coherently.

After eliminating the factors of changes in enterprise production costs, we turn to the analysis from spatial dimension.

2.4. Analysis from space dimension

The fluctuation of an enterprise's product profit is not only influenced by internal factors but also by the products of other enterprises, especially for enterprises that have reached monopoly agreements. Therefore, further analysis is needed from space dimension after determining the mutation point. In detail, a spatial econometric model is established for enterprises in the high impact group entering the price change periods. According to different groups obtained from cluster analysis, we calculate the intra-class correlation coefficient and interclass correlation coefficient of each enterprise's group. Through comparing the correlation coefficients, we can identify whether there is a similar change trend between two time series.

2.4.1. Spatial correlation model

We first perform spatial correlation analysis on three price change periods: 30 May 2016 to 11 Dec. 2017, 11 Dec. 2017 to 27 Oct. 2018, and 27 Oct. 2018 to 31 May 2019, separately. To do this, we apply the first order temporal correlation method to determine the degree of correlations between these enterprises. The correlation coefficients are defined as

$$\mathrm{CORT}(X_T, Y_T) = \frac{\sum_{t=1}^{T}(X_{t+1} - X_t)(Y_{t+1} - Y_t)}{\sqrt{\sum_{t=1}^{T-1}(X_{t+1} - X_t)^2}\sqrt{\sum_{t=1}^{T-1}(Y_{t+1} - Y_t)^2}}.$$

Here, we let X_T be the time series of product profits for a single enterprise in the high impact group, and T be the corresponding time range of price fluctuations. If the purpose is to calculate the intra group correlation coefficient, then Y_T represents the time series of product profits of other single enterprises in the high impact group. If the purpose is to calculate the inter group correlation coefficient, then Y_T represents the time series of product profits of other single enterprises in the low impact group. We calculate intra group correlation coefficients and inter group correlation coefficients in the spatial correlation analysis by taking the data of the price change period 30 May 2016 to 11 Dec. 2017 as an example. The results are in the following Tables 8 and 9, where some enterprises have blank results due to missing original data or not meeting constraints.

Table 8. Intra group correlation coefficient (30 May 2016 to 11 Dec. 2017).

	ENT1	ENT2	ENT3	ENT4	ENT5	ENT6	ENT7	ENT8	ENT9	ENT10	ENT11
ENT1	1	0.899	0.452	0.828	0.641	0.652	0.901	0.755	0.844	0.64	0.64
ENT2	0.899	1	0.492	0.805	0.827	0.503	0.856	0.679	0.947	0.603	0.755
ENT3	0.452	0.492	1	0.373	0.39	0.367	0.398	0.353	0.451	0.374	0.228
ENT4	0.828	0.805	0.373	1	0.633	0.782	0.842	0.81	0.789	0.714	0.589
ENT5	0.641	0.827	0.39	0.633	1	0.372	0.659	0.55	0.856	0.459	0.439
ENT6	0.652	0.503	0.367	0.782	0.372	1	0.66	0.912	0.478	0.716	0.142
ENT7	0.901	0.856	0.398	0.842	0.659	0.66	1	0.773	0.785	0.819	0.639
ENT8	0.755	0.679	0.353	0.81	0.55	0.912	0.773	1	0.679	0.667	0.373
ENT9	0.844	0.947	0.451	0.789	0.856	0.478	0.785	0.679	1	0.502	0.719
ENT10	0.64	0.603	0.374	0.714	0.459	0.716	0.819	0.667	0.502	1	0.412
ENT11	0.64	0.755	0.228	0.589	0.439	0.142	0.639	0.373	0.719	0.412	1
Mean value	0.750	0.761	0.443	0.742	0.621	0.599	0.757	0.686	0.732	0.628	0.540

Table 9. Intra group correlation coefficient (30 May 2016 to 11 Dec. 2017).

	ENT1	ENT2	ENT3	ENT4	ENT5	ENT6	ENT7	ENT8	ENT9	ENT10	ENT11
ENT13	0.732	0.78	0.564	0.645	0.785	0.331	0.611	0.648	0.895	0.47	0.493
ENT14	0.39	0.451	0.095	0.411	0.515	0.192	0.486	0.592	0.656	0.365	0.619
ENT15	0.233	0.005	-0.23	0.13	-0.26	0.273	0.154	0.331	0.06	-0.224	0.057
ENT18	0.868	0.801	0.323	0.902	0.643	0.785	0.817	0.877	0.812	0.552	0.522
ENT20	0.218	0.022	-0.08	0.267	0.107	0.195	0.166	0.199	-0.02	0.165	-0.024
ENT21	0.909	0.813	0.371	0.934	0.626	0.85	0.918	0.902	0.798	0.759	0.522
ENT22	0.844	0.947	0.451	0.789	0.856	0.478	0.785	0.679	1	0.502	0.719
ENT23	0.936	0.961	0.124	0.925	0.58	0.408	0.834	0.587	0.971	0.459	0.971
ENT26	0.53	0.589	0.261	0.502	0.49	0.328	0.535	0.592	0.731	0.295	0.574
ENT28	0.652	0.503	0.367	0.782	0.372	1	0.66	0.912	0.478	0.716	0.142
ENT29	0.722	0.581	0.463	0.794	0.457	0.951	0.755	0.955	0.522	0.712	0
ENT30	0.528	0.703	0.533	0.742	0.784	0.637	0.585	0.67	0.7	0.394	-0.343
ENT31	0.911	0.942	0.364	0.836	0.757	0.537	0.925	0.713	0.883	0.666	0.713
ENT32	0.923	0.884	-0.26	0.842	0.566	0.846	0.829	0.825	0.873	0.829	0.877

Due to the continuous price change periods: 30 May 2016 to 11 Dec. 2017, 11 Dec. 2017 to 27 Oct. 2018, and 27 Oct. 2018 to 31 May 2019, we calculate the average intra group and inter group correlation coefficients corresponding to the product profits of each enterprise in the high impact group separately during the three price change periods. The results of the respective tests were used to comprehensively determine the correlation characteristics. The results are shown in the Tables 10–12 below.

Table 10. Correlation characteristics in high impact group (30 May 2016 to 11 Dec. 2017).

Enterprise	Intra group mean value	Inter group mean value	Difference	Correlation characteristics
ENT1	0.750	0.671	0.079	Intra > Inter
ENT2	0.761	0.642	0.119	Intra > Inter
ENT3	0.443	0.238	0.205	Intra > Inter
ENT4	0.742	0.679	0.064	Intra > Inter
ENT5	0.621	0.519	0.101	Intra > Inter
ENT6	0.599	0.558	0.041	Intra > Inter
ENT7	0.757	0.647	0.110	Intra > Inter
ENT8	0.686	0.677	0.009	Intra > Inter
ENT9	0.732	0.668	0.064	Intra > Inter
ENT10	0.628	0.476	0.152	Intra > Inter
ENT11	0.540	0.417	0.122	Intra > Inter

Table 11. Correlation characteristics in high impact group (11 Dec. 2017 to 27 Oct. 2018).

Enterprise	Intra group mean value	Inter group mean value	Difference	Correlation characteristics
ENT1	0.644	0.672	-0.028	Intra < Inter
ENT2	0.502	0.295	0.207	Intra > Inter
ENT3	0.516	0.558	-0.042	Intra < Inter
ENT4	0.624	0.651	-0.027	Intra < Inter
ENT5	0.505	0.525	-0.020	Intra < Inter
ENT6	0.702	0.557	0.144	Intra > Inter
ENT7	0.614	0.539	0.074	Intra > Inter
ENT8	0.511	0.491	0.020	Intra > Inter
ENT9	0.604	0.433	0.171	Intra > Inter
ENT10	0.162	-0.051	0.213	Intra > Inter
ENT11	0.695	0.559	0.136	Intra > Inter

Table 12. Correlation characteristics in high impact group (27 Oct. 2018 to 31 May 2019).

Enterprise	Intra group mean value	Inter group mean value	Difference	Correlation characteristics
ENT1	0.934	0.944	-0.010	Intra < Inter
ENT2	0.949	0.940	0.008	Intra > Inter
ENT3	0.867	0.830	0.037	Intra > Inter
ENT4	0.791	0.738	0.052	Intra > Inter
ENT5	0.944	0.944	0.000	Intra = Inter
ENT6	0.950	0.929	0.021	Intra > Inter
ENT7	0.948	0.944	0.004	Intra > Inter
ENT8	0.931	0.903	0.028	Intra > Inter
ENT9	0.947	0.947	0.000	Intra = Inter
ENT10	0.909	0.884	0.025	Intra > Inter
ENT11	0.911	0.867	0.045	Intra > Inter

Combining with the actual market situation of these enterprises, considering that many enterprises have strong systematic correlation and noting that the inter group correlation shows the characteristics of "high value-low value", we preliminarily believed that the enterprises whose intra group correlation coefficient is greater than the inter group correlation coefficient in the three price change periods are suspected of dominating the market product prices.

Next, we explore the inter relationships among high impact enterprises. We study the correlation coefficients of the product profits of each enterprise and get that there is a strong correlation among multiple high impact enterprises, which indicates that their enterprises' profit volatility has the same trend of change. We further verify this strong correlation through the product price sequence diagram of high impact group enterprises during the price change period, which is shown in the following Figure 2.

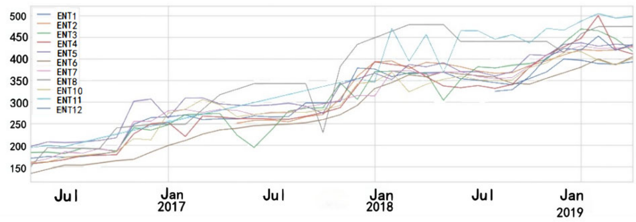


Figure 2. The product price sequence diagram.

From the above Figure 2, it can be seen that the trends of the product prices of ENT6, ENT8, ENT10, and ENT11 are relatively consistent. The enterprises ENT8 and ENT11 have relatively high products prices and a strong influence in the market. Based on the analysis of the product profit correlation of enterprises, there is a clear correlation between enterprises ENT6, ENT8, ENT10, and ENT11. The possibility of "price leadership" is high, with the characteristic of price "collusion".

In the three periods of ENT2, ENT6–ENT8, ENT10, and ENT11, the intra group correlation was greater than the inter group correlation, and data analysis showed that ENT6, ENT8, ENT10, and ENT11 were strongly correlated with each other. Therefore, it is preliminarily determined that the above six enterprises are suspected of monopoly. Next, further analysis will be conducted through spatial dimension.

From the following Figure 3, we find that there are four enterprises located in S city, while the distribution of the other enterprises is relatively scattered. Considering the regional clustering of monopolistic behavior, it is considered that ENT2, ENT6–ENT8, ENT10, and ENT11 have significant suspicion of collusion, and it is necessary to continue to carry out antitrust identification.

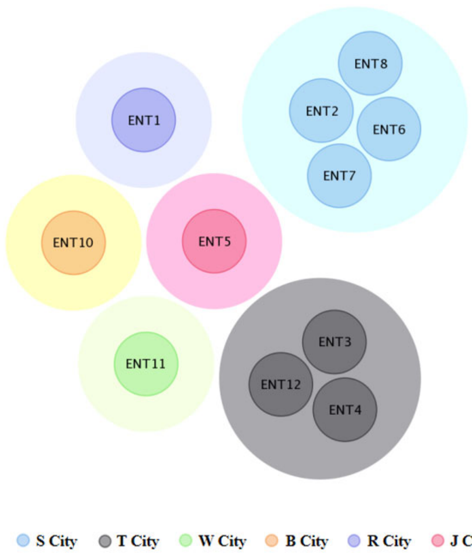


Figure 3. Geographical distribution of enterprises.

2.4.2. Grey correlation model

The grey correlation model is a measure of the degree of correlation between two factors over time. If the changing trends of two factors in the data are basically consistent, it indicates a high degree of synchronous change and a high degree of correlation between these two factors. The grey correlation coefficient is a measure of the degree of grey correlation, which is based on the similarity or dissimilarity of the development between factors, also known as the "grey correlation degree". As a method of measuring the degree of correlation between factors, from [14] we know that the closer the value of correlation is to 1, the better the correlation is. Now we begin to verify the above model conclusion by calculating the grey correlation degree of the enterprise profit time series data, and the specific steps are as follows:

(1) Due to the different units of influencing factors in the reference sequence and comparison sequence, standardization is carried out to eliminate dimensionality;

(2) Calculate the grey correlation coefficients of reference sequence and comparison sequence. The correlation coefficient between each comparative sequence x_i and the reference sequence x_0 at various times can be calculated by the formula

$$\zeta_{0i} = \frac{\Delta(\min) + \rho\Delta(\max)}{\Delta_{0i}(k) + \rho\Delta(\max)},$$

where the distinguishing coefficient $\rho = 0.5$, $\Delta(\min)$ is the second level minimum difference, $\Delta(\max)$ is the second level maximum difference, and $\Delta_{0i}(k)$ is the absolute difference between each point on each x_i curve and each point on the x_0 curve.

We apply the grey correlation model to calculate the profit correlation degree of high impact group enterprises, which is shown in the following Figures 4–6.

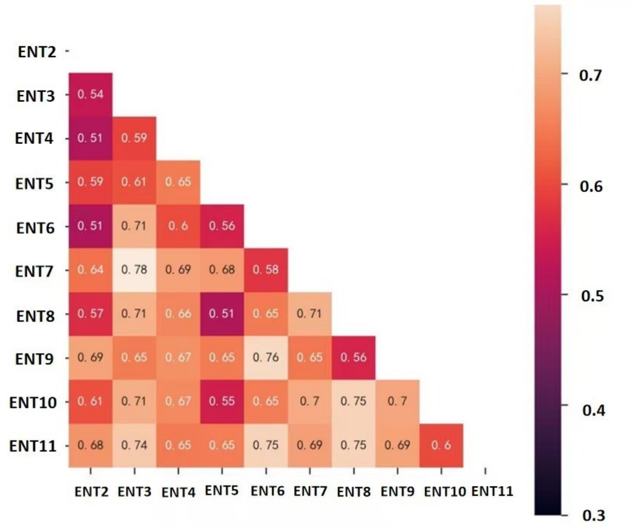


Figure 4. 30 May 2016 to 11 Dec. 2017.

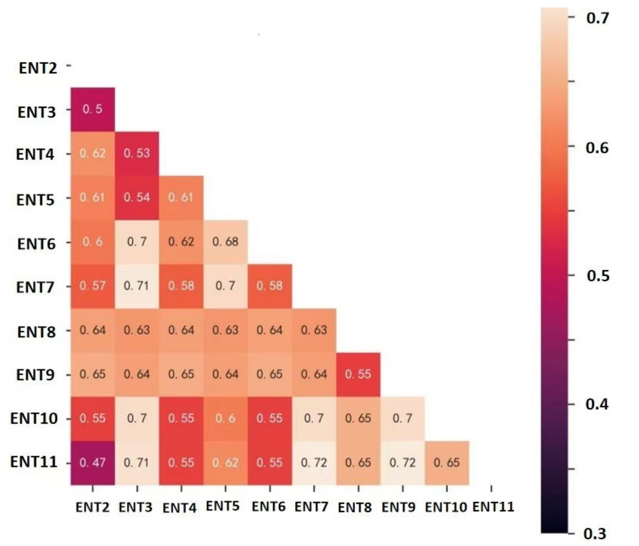


Figure 5. 11 Dec. 2017 to 27 Oct. 2018.

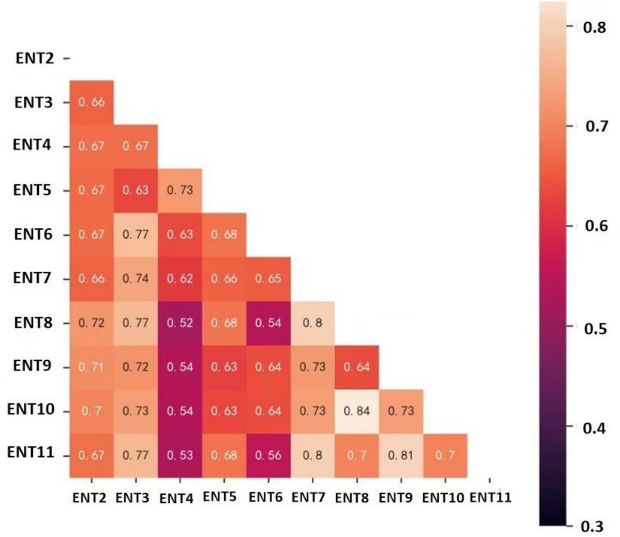


Figure 6. 27 Oct. 2018 to 31 May 2019.

Then, the mean value of grey correlation degree of these six enterprises obtained from the spatial correlation model over three price change periods is shown in the below Table 13.

Table 13. The mean value of grey correlation degree.

Enterprise	The mean value of grey correlation degree		
	30 May 2016 to 11 Dec. 2017	11 Dec. 2017 to 27 Oct. 2018	27 Oct. 2018 to 31 May 2019
ENT1, ENT3, ENT4, ENT5, ENT9	0.62015136	0.606047602	0.641702521
ENT2, ENT6, ENT7, ENT8, ENT10, ENT11	0.714674243	0.691330024	0.761702107

The results of grey correlation analysis show that the grey correlation between enterprises ENT2, ENT6–ENT8, ENT10, and ENT11 is significantly higher than that of other enterprises. The trend of profit changes in these six enterprises is basically consistent. It indicates a high degree of synchronous change and a high degree of correlation among these six enterprises. Therefore, the above results validate the conclusion of the spatial correlation model, and it can be deduced that these six enterprises mentioned above have suspicion of monopoly.

2.4.3. The mutation point test of operating income of monopolistic enterprises

Based on these six companies with suspected monopoly identified in the previous sections, we further verified whether their mutation points of main operating income are consistent with market price fluctuations by applying the Pettitt mutation point test (see [15]). If there is a significant point of change in the main operating income of an enterprise, and the date of the change occurs before the period of market price fluctuations, it is considered that the enterprise has suspicion of manipulating market prices.

The algorithm of the Pettitt mutation point test is as follows. For the time series X_t of the main operating income of the enterprise, we calculate the statistic U_t as

$$U_t = U_{t-1} + V_t, \quad t \in [2, N],$$

where

$$V_t = \sum_{j=1}^{N} \operatorname{sgn}(x_t - x_j), \qquad U_1 = V_1.$$

We take the value K_t, which is the largest absolute value of U_t, as the most significant mutation point, and calculate the P-value of the statistic corresponding to K_t. If the P-value satisfies a level of significance less than the given level, then it indicates the existence of statistically significant mutation points.

We apply Matlab software to organize the business data of these six enterprises, and then calculate the mutation point of each enterprise's main operating income. The calculation results are in the following Table 14.

Table 14. Mutation points of main operating income.

Enterprise	ENT2	ENT6	ENT7	ENT8	ENT10	ENT11
Date of mutation point	30 Sep. 2017	31 May 2017	31 Aug. 2015	31 May 2017	31 Nov. 2015	31 Aug. 2017
Recent price changes point	11 Dec. 2017	11 Dec. 2017	27 Dec. 2015	11 Dec. 2017	27 Dec. 2015	11 Dec. 2017

Through the mutation point test results, it can be seen that the enterprises can be divided into two parts by their mutation points of main operating income.

The first part consists of ENT2, ENT6, ENT8, and ENT11. The mutation points of the main operating income of these four enterprises are around the second price change point (11 Dec. 2017), and the date of which it occurred is before market price changes. Therefore, we believe that these four enterprises have a suspicion of monopoly and manipulate market prices through collusion.

The second part consists of ENT7 and ENT10. The mutation points of main operating income of these two enterprises are around the first price change point (27 Dec. 2015), and the date of which it occurred is before market price changes. Therefore, we believe that these two enterprises have a suspicion of monopoly and manipulate market prices through collusion.

3. Conclusions

The results of our model indicate that the research object of this study is suitable for the analysis of monopolistic suspicion in the type of monopolistic agreement reached by operators. The results of analysis from time dimension indicate that the abnormal change period of market prices is from 30 May 2016 to 31 May 2019. The results of analysis from spatial dimension indicate that there are 6 suspected monopolistic enterprises involved, which is further verified by using the grey correlation model.

Based on the economic model and various government data, we apply the big data perception model to identify and monitor the monopoly behavior of enterprises from the two dimensions of time and space. Through practical studies and the model we established in this paper, the vast majority of monopolistic behaviors can be more accurately identified.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors would like to express their gratitude to the referee for a careful reading of the manuscript and many valuable suggestions which highly improve the quality of this paper. This is supported by the Technological Foundation of Administration for Market Regulation of China (Grant No. 2022MK075).

Conflict of interest

The authors declare no conflicts of interest.

References

1. J. Bai, P. Perron, Estimating and testing linear models with multiple structural changes, *Econometrica*, **66** (1998), 47–78. https://doi.org/10.2307/2998540

2. J. Bai, P. Perron, Computation and analysis of multiple structural change models, *J. Appl. Econometrics*, **18** (2003), 1–22. https://doi.org/10.1002/jae.659

3. T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *J. Econometrics*, **31** (1986), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

4. J. A. Doornik, M. Ooms, Multimodality in the GARCH regression models, *Int. J. Forecast.*, **24** (2008), 432–448. https://doi.org/10.1016/j.ijforecast.2008.06.002

5. A. Charles, O. Darné, Outliers and GARCH models in financial data, *Econ. Lett.*, **86** (2005), 347–352. https://doi.org/10.1016/j.econlet.2004.07.019

6. M. Yavuz, F. Özköse, M. Akman, Z. T. Taştan, A new mathematical model for tuberculosis epidemic under the consciousness effect, *Math. Modell. Control*, **3** (2023), 88–103. https://doi.org/10.3934/mmc.2023009

7. A. Chavada, N. Pathak, S. R. Khirsariya, A fractional mathematical model for assessing cancer risk due to smoking habits, *Math. Modell. Control*, **4** (2024), 246–259. https://doi.org/10.3934/mmc.2024020

8. G. E. A. P. A. Batista, M. C. Monard, A study of K-nearest neighbour as a model-based method to treat missing data, *Proceedings of Argentine Symposium on Artificial Intelligence*, 2001.

9. R. Engle, Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *J. Bus. Econ. Stat.*, **3** (2002), 339–350. https://doi.org/10.1198/073500102288618487

10. H. N. Huang, W. Zhong, Evaluation on volatility forecasting performance of GARCH-type models, *Chin. J. Mana. Sci.*, **15** (2007), 13–19. https://doi.org/10.3321/j.issn:1003-207x.2007.06.003
11. P. Kokoszka, R. Leipus, Change-point estimation in ARCH models, *Bernoulli*, **6** (2000), 513–539. https://doi.org/10.2307/3318673
12. C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37** (1969), 424–438. https://doi.org/10.2307/1912791
13. C. A. Sims, Money, income, and causality, *Amer. Econ. Rev.*, **62** (1972), 540–552.
14. S. F. Liu, Z. G. Fang, Y. Lin, Study on a new definition of degree of grey incidence, *J. Grey Syst.*, **9** (2006), 115–122. https://doi.org/10.30016/JGS.200612.0005
15. C. Truong, L. Oudre, N. Vayatis, Selective review of offline change point detection methods, *Signal Process.*, **167** (2020), 107299. https://doi.org/10.1016/j.sigpro.2019.107299