ISSN 2078-6115. Вісник Львівського університету. Серія економічна. 2023. Випуск 64.
Visnyk of the Lviv University. Series Economics. 2023. Issue 64.

КІЛЬКІСНІ МЕТОДИ В ЕКОНОМІЦІ. ЦИФРОВА ЕКОНОМІКА

DOI: http://dx.doi.org/10.30970/ves.2023.64.0.6405

УДК 33:339.72:519.86:004.94
JEL C38; G13; C45; Q40; C69

CLUSTER ANALYSIS OF THE TERM STRUCTURE OF CRUDE OIL PRICES USING SELF-ORGANIZING MAPS

Eduard Kolotvin, Iryna Hrabynska

Ivan Franko National University of Lviv,
18 Svobody Ave., Lviv, 79008,
email: eduard.kolotvin@lnu.edu.ua; ORCID: https://orcid.org/0009-0000-6622-411X
email: iryna.hrabynska@lnu.edu.ua; ORCID: https://orcid.org/0000-0001-7390-5626

Abstract. *In this study, we apply self-organizing map (SOM), also referred to as the Kohonen map, clustering to monthly forward curves of crude oil futures prices and ending stock of crude oil at Cushing, OK, for the period from November 2011 to June 2022 in order to investigate the contango-backwardation patterns and their relationship to seasonality and stock levels. In particular, we suggest two approaches to shape-based clustering of forward curves, namely numerical and topological. Our results show that SOM clustering can reveal distinct patterns of futures price term structure across different months, with some months, i.e., January to June, exhibiting precise and others, i.e., July to December, showing somewhat ambiguous behavior. Moreover, we suggest an approach to the comparative analysis of seasonal patterns and underlying fundamentals, namely commercial ending stock levels. In particular, we prove that periods of very low to modest ending stock levels are more likely to exhibit backwardation, whilst periods of high to very high ending stock levels are more likely to exhibit contango. Overall, our study suggests new methods and approaches to analyzing the behavior of crude oil futures prices, namely their term aspect manifested by the configuration of forward curves, highlighting the importance of monthly patterns and seasonal regimes, as well as proves the validity of using machine learning methods and comparative cluster analysis to obtain insights on the dynamics of crude oil futures prices.*
Keywords: *term structure of crude oil futures prices, contango, backwardation, forward curve, shape-based cluster analysis, price clustering, time series, self-organizing map (SOM), Kohonen map, machine learning, artificial neural networks (ANNs), dynamic time warping (DTW), commercial ending stocks of crude oil.*

Introduction. Despite the high pace of the fourth energy transition – rapid growth in the role of alternative energy sources – crude oil remains one of the most critical energy resources, with numerous applications in transportation, heating, and electricity generation, as well as a benchmark for the pricing of other petroleum products, being a key factor in determining the profitability of companies in the energy sector and related industries. Crude oil futures prices are one of the most closely monitored and analyzed indicators in the global energy markets. As a key source of energy, crude oil plays a critical role in the global economy, affecting the prices of other commodities and influencing consumer behavior. Moreover, oil futures prices can affect inflation, trade balances, and geopolitical stability. Hence, they serve as a key indicator of global economic health, influencing decision-making across a wide range of industries and policy areas. Given its importance, understanding the fundamentals of crude oil futures pricing is vital for policymakers, industry leaders, and investors, as it can provide insights into market trends, inform investment decisions, and help mitigate risks associated with price volatility.

The behavior of crude oil futures prices is complex, with various factors influencing their movements. Supply and demand, shifts in technology, global upheavals, and market structure all play a role in shaping crude oil prices. The term structure of oil prices, which describes the relationship between the current and futures prices of crude oil futures contracts, is a critical factor in determining the behavior of crude oil futures prices. Generally, when the futures price is higher than the current price, the market is said to be in contango. On the other hand, when the futures price is lower than the current price, the market is said to be in backwardation. Hereinafter, we refer to these two contrasting extremes of price term structure as the contango-backwardation dichotomy.

In this study, we explore the relationship between monthly seasonality, contango-backwardation futures market structure, commercial ending stocks of oil, and the behavior of crude oil futures prices. Specifically, we apply self-organizing map (SOM) clustering to monthly seasonal forward curves of crude oil futures prices and ending stock levels to investigate how the contango-backwardation dichotomy varies across different months of the year. Our analysis provides insight into the dynamics of crude oil futures prices, suggesting new approaches to analyzing futures prices, such as SOM cluster analysis of forward curves shape and configuration.

Review of related research and publications. The functioning of futures commodity markets as well as relationships between current and future prices, have been a subject of study by such prominent economists as J. M. Keynes, H. Hotelling, H. Working, N. Kaldor, F. H. Weymar, S. J. Turnovsky, and R. S. Pindyck. The peculiarities of the crude oil and other commodity derivatives markets have been examined by B. Simkins, R. Simkins, J. Haubrich, E. S. Shwartz, M. Ludkovski, R. Karmona, J. Kakeu, J. I. Considine, L. H. Ederington and many others.

In this paper, we employ methods of time-series clustering, which is a rather new field of scientific research. Nonetheless, it has been of great use in multiple domains and research areas over the past decades. E. A. Maharaj, P. D'Urso, and J. Caiado, in particular, suggest a profound study of time series clustering and classification, covering the methodology of most common supervised and unsupervised clustering approaches and techniques that are widely used in economics, finance, and other fields such as medicine, astronomy, and environmental science [15]. S. Aghabozorgi et al. [1], as well as T. W. Liao [14], in their turn, provide a broad review of approaches to time-series clustering and discuss most recent developments in time-series clustering across various scientific domains. Interestingly, Y. Huhtala, J. Karkkainen, and H. Toivonen discuss the discovery of non-obvious relationships between financial time series using wavelets and feature-driven clustering [11], while H. Izakian and W. Pedrycz suggest a fuzzy-clustering framework for detecting shape and amplitude anomalies within temporal data [12].

When it comes to SOM clustering in economic and financial analysis, this approach has been mainly used to predict the behavior of time series. In particular, G. A. Barreto discusses applications of SOMs to time series prediction, arguing that SOM-based models possess multiple advantages, e.g., easy incorporation of new knowledge in the initialized SOM-driven model [3, p. 135]. S. Dablemont et al. suggest a general method for forecasting financial time series using SOMs, applying it to the DAX30 index prediction [9]. The approach has also found application in banking, as M. Nordlinder proposes using SOM clustering for assessing operational deposit levels on bank accounts, identifying different types of such accounts, and consumer behavior related thereto [16]. J. A. Lee and M. Verleysen justly demonstrate that the SOM algorithm is a flexible tool capable of achieving interesting properties once new parameters, such as neighborhood adaptations, are incorporated [13]. A. Blazejewski and R. Coggins prove that SOMs are a valuable tool for clustering high-frequency financial data, which might be a complicated task even if using sophisticated stochastic models [5].

Of particular interest is an article by J. Baruník and B. Malinská, in which they propose a neural-network-based approach to forecasting the term structure of crude oil prices, rightly noting the rarity of scientific literature devoted to this particular topic [4]. In our study, we contribute to the scarce developments in this area by suggesting a new framework of neural-network-driven analysis of the term structure of crude oil prices, namely SOM clustering of forward curves by their shape and, moreover, comparative cluster analysis with underlying fundamentals.

Rationale and objectives of the research. In this research, we aim to discover the monthly seasonality patterns of crude oil futures prices using the SOM clustering technique to identify patterns in the crude oil futures prices behavior over different periods, as well as analyze the relationship of such patterns with the contango-backwardation dichotomy, applying our findings to the study of oil market fundamentals, namely levels of commercial stocks.

Methods of the research. In this study, we use general scientific and special research methods on financial and economic phenomena and processes in their continuous development and interrelationship. The methods used to accomplish the objectives of the study are as follows: historical – to assess pricing trends formed in the past; induction and deduction – to study individual components and factors of futures pricing processes and to determine cause and effect linkages; abstract and logical analysis and synthesis – to identify and summarize the criteria for clustering forward curves; descriptive analysis and observation – to build and explain tables, graphs, and charts; machine learning (in particular, unsupervised learning) – to determine the dimensionality and construct SOMs; and economic, mathematical, statistical methods as well as algorithmization and programming – to perform calculations and computations, including data normalization, dynamic time warping (DTW), grouping, clustering, and other mathematical transformations.

Main research material. While SOM clustering has found multiple applications across various domains, there has been little to no evidence of previous usage of SOM cluster analysis to study term structures of commodity prices and their dynamics, i.e., forward curves. Such models as the cost of carry model, the convenience yield model, the expectations theory, and their variations, as well as many other deterministic and stochastic models, have been widely used to discover the peculiarities of the futures prices and their behavior. We believe that introducing new approaches and combinations thereof, as well as improved visualization techniques, such as SOM clustering of forward curves and comparative cluster analysis, will benefit the understanding of futures market functioning and contribute to further research of crude oil prices term structure.

Building a SOM for forward curve clustering. In this study, the SOM was implemented using the MiniSOM library [17] in Python 3.9.12. The input data were prices of futures contracts for WTI crude oil on the NYMEX/CME exchange from November 2011 to June 2022, namely observations for 128 months and 12 maturities (see Figure 1). The data were obtained from the market price aggregator *Barchart* [7]. Descriptive statistics of the input data are presented in Table 1.

To perform cluster analysis, we normalized the data using the feature scaling method, i.e., we scaled the input data so that their values were in the [0, 1] range. Next, we determined the dimensionality of the SOM grid by obtaining a square matrix with (4 x 4) dimensions using the following formula:

$$som_x = som_y = \left\lceil \sqrt[4]{len(data)} \right\rceil \tag{1}$$

where som_x is the width of the SOM grid, som_y is the height of the SOM grid, and *len(data)* is the number of observations.

Table 1

Descriptive Statistics of Input Data

M	Count	Mean	Std	Min	25%	50%	**75%**	**Max**
1	128	65.92	20.42	10.01	50.42	61.97	86.38	120.93
2	128	66	20	18.84	50.25	61.82	86.05	118.25
3	128	66.01	19.38	21.85	50.77	61.83	85.75	115.31
4	128	65.97	18.82	24.37	51.22	62.14	85.47	112.62
5	128	65.9	18.36	26.37	51.56	62.4	85.18	110.25
6	128	65.8	17.99	27.67	51.46	62.57	84.93	108.1
7	128	65.67	17.67	28.72	51.45	62.8	84.61	106.16
8	128	65.53	17.42	29.62	51.38	62.44	84.3	104.43
9	128	65.41	17.2	30.38	51.56	62.07	84.04	102.9
10	128	65.28	17.01	31.07	51.59	62.04	83.89	101.51
11	128	65.17	16.85	31.66	51.81	61.97	83.78	100.25
12	128	65.06	16.71	32.17	51.76	61.58	83.7	99.07

Source: calculated by the author based on the input data
Note: M – expiration, months; Count – number of data points; Mean – arithmetic average; Std – standard deviation; Min – minimal value; 25% – first quartile; 50% – median or second quartile; 75% – third quartile; Max – maximal value.

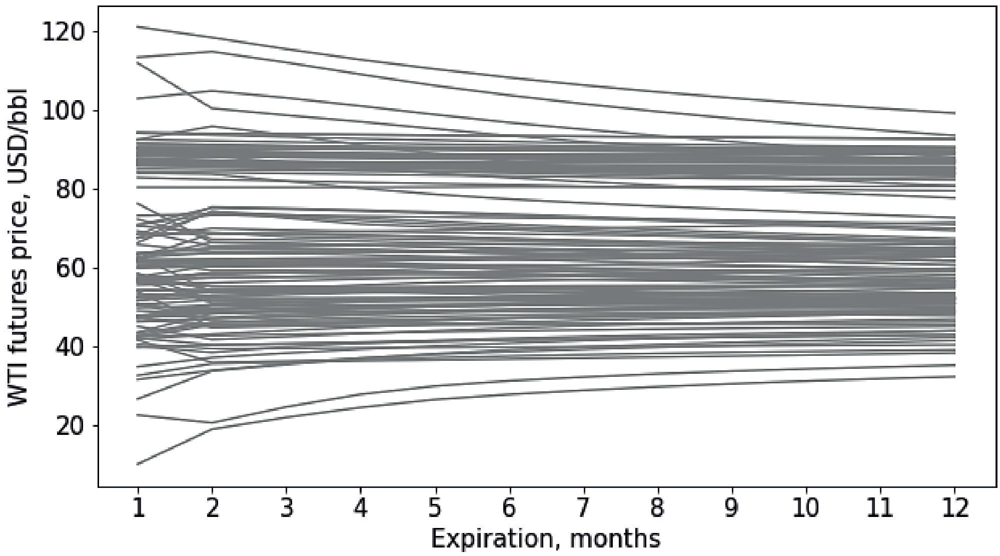


Figure 1. WTI Forward Curves, November 2011 – June 2022
Source: constructed by the author based on the input data

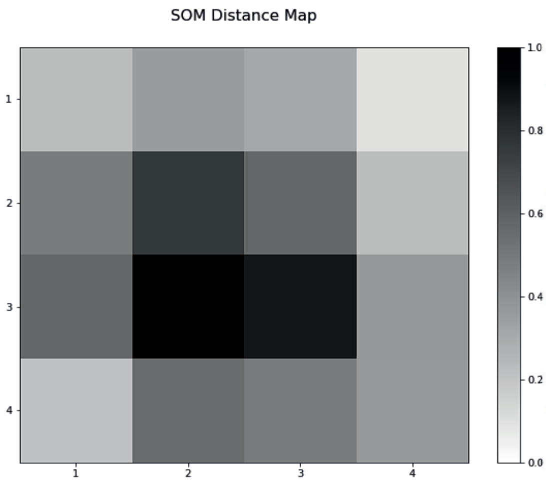
Next, we set the SOM grid parameters:

$$SOM = SOM(som_x, som_y, len(data), \sigma = 1, \alpha = 2, \theta = 'Gauss', s = 0) \quad (2)$$

where som_x is the width of the SOM grid, som_y is the height of the SOM grid, *len(data)* is the number of observations, σ is the spread parameter of the neighborhood function, α is the learning coefficient, θ is the type of neighborhood function, and s is the value of the random seed.

After obtaining the SOM grid, we set the number of training iterations over the input data to be 100,000. The initial weights of the grid nodes for the input dataset were identified randomly. The map of neural distances obtained after all training iterations, i.e., the transformation of a set of one-dimensional input vectors into a two-dimensional set of neurons, is shown in Figure 2. The activation map of the SOM network (transposed) and the corresponding degrees of neuronal activation are shown in Figure 3.

In this study, we performed shape-based clustering of the input data using two criteria: quantitative and topological. Both clustering variants are based on the same SOM, so the distribution of clusters across neurons is identical in both cases (Figure 4). It should be noted that according to the features of the input data and the parameters of the constructed SOM, one cluster appeared to be empty. The clustering of the input data is obviously different for the two criteria. To group the forward curves and calculate the relative distances, we used the DTW method to avoid possible inaccuracies in constructing typical curves for each cluster compared to Euclidean distances. We determined the degree of similarity of the curves by comparing the DTW distances between them.



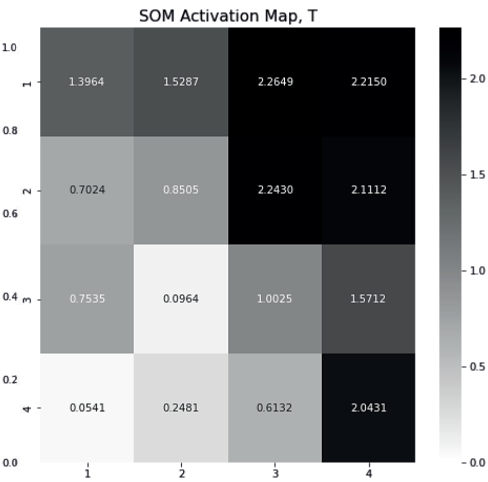


Figure 2 (left). SOM Distance Map **Figure 3 (right). SOM Activation Map, Transposed**

Source: developed by the author

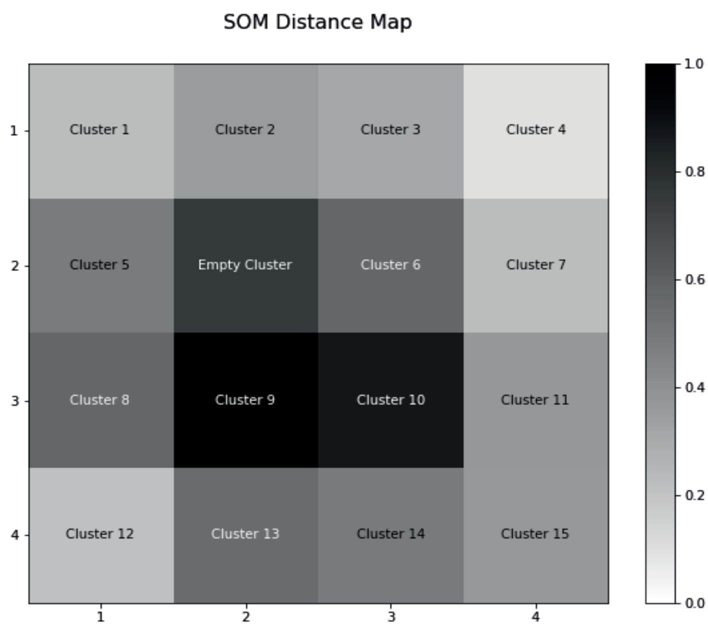


Figure 4. SOM Distance Map and Respective Clusters
Source: developed by the author

Quantitative clustering of forward curves. To cluster the normalized forward curves by the quantitative criterion, we grouped the curves by the degree of similarity of their shapes and sorted them in descending order of observation frequency. Accordingly, Cluster 1 has the largest number of observations (26 observations), and Cluster 15 has the lowest number (1 observation). The result of quantitative shape-based clustering is shown in Figures 5-6.

Having divided the curves into clusters, we calculated the probability of the *i*-th cluster being the most typical forward curve configuration for the *j*-th month. Based on these indicators, we constructed a heatmap of the probability distributions by month and cluster number (Figure 7).

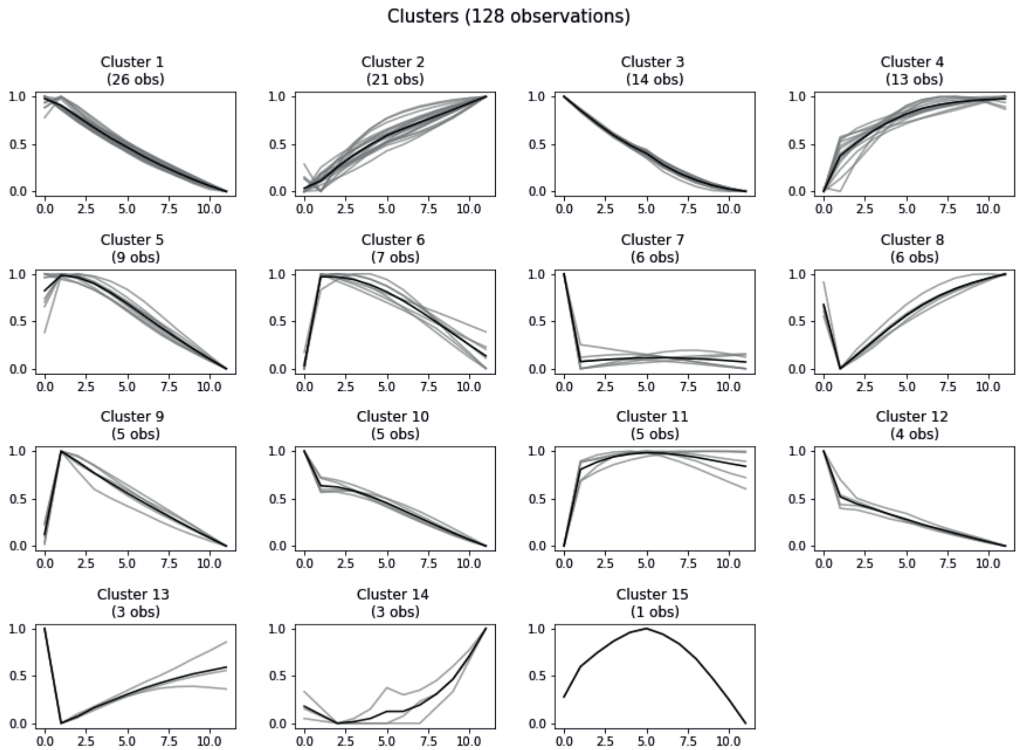


Figure 5. Graphical representation of forward curve distribution by clusters, quantitative approach

Source: developed by the author

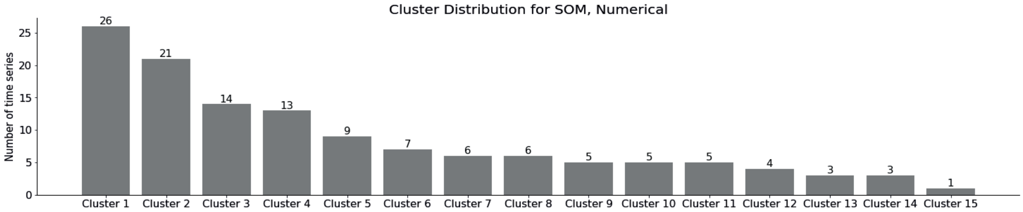


Figure 6. Distribution of forward curves by clusters, quantitative approach

Source: developed by the author

From the configuration of typical curves for the clusters (Figure 5) and the scattering characteristics on the heatmap (Figure 7), we can see that pure contango and backwardation are most typical for the studied period (Clusters 1-5). Interestingly, the backwardation is stronger in the first half of the year (Clusters 1, 3, 5), while both contango and backwardation characterize the second half. Cases of «humped» contango and backwardation (Clusters 6-15) are less likely scenarios for all months except for mid-summer (July) and the end of the year (December). For July, the most likely scenarios are backwardation and «humped» contango – short-end backwardation with long-end contango. For December, on the contrary, it's short-end contango with long-end backwardation.

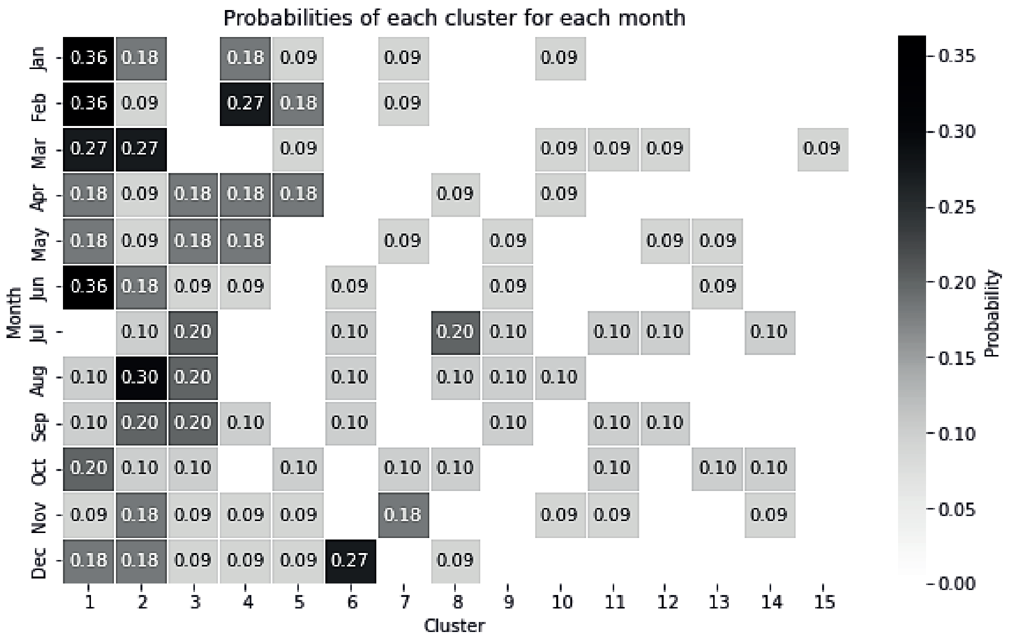


Figure 7. Heatmap of cluster probability distribution by month, quantitative approach
Source: developed by the author

Topological clustering of forward curves. To perform shape-based clustering of the normalized forward curves by the topological criterion, we grouped and sorted the curves by the degree of similarity of their configurations in descending order: from the most to the least similar. Accordingly, Cluster 1 is the benchmark (26 observations), and Cluster 15 is the least similar cluster to it, having the reverse configuration (13 observations). We can see that Cluster 1 represents backwardation, and Cluster 15 represents contango. The result of topological shape-based clustering is shown in Figures 8-9.

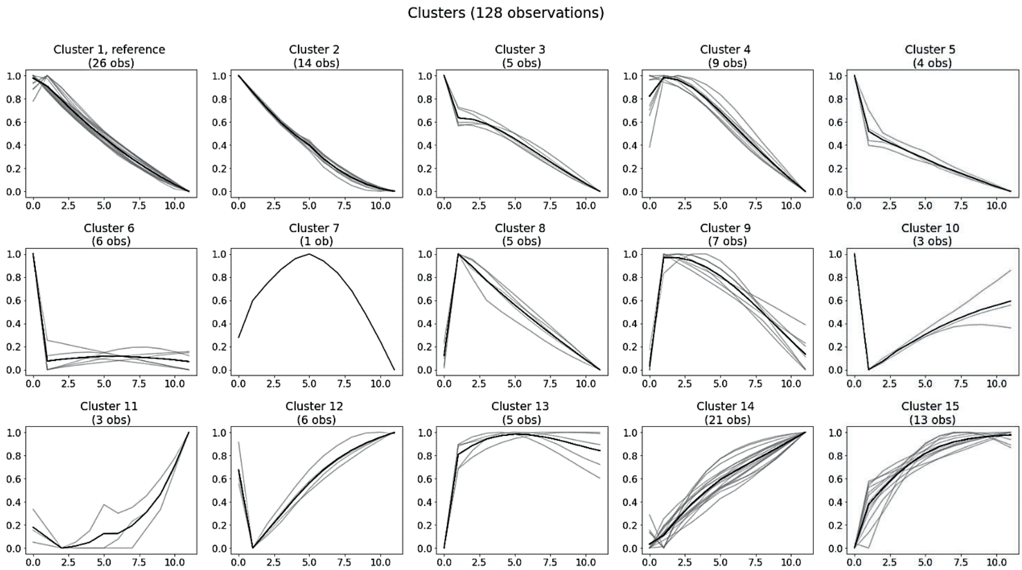


Figure 8. Graphical representation of forward curve distribution by clusters, topological approach
Source: developed by the author

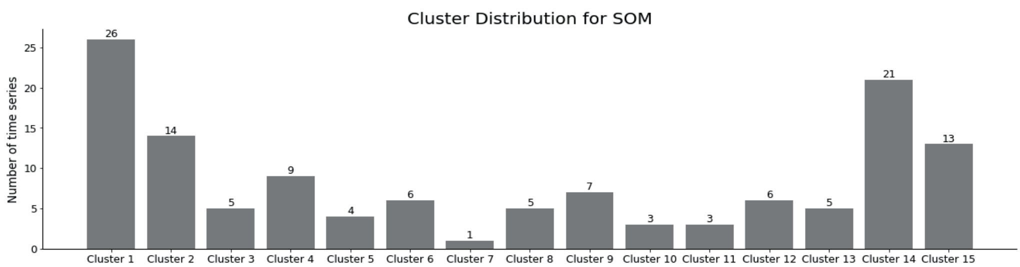


Figure 9. Distribution of forward curves by clusters, topological approach
Source: developed by the author

Similarly to the quantitative approach, we calculated the probability with which the *i*-th cluster is the most typical forward curve configuration for the *j*-th month. Based on this, we determined which cluster type is the most (least) typical for a given month (see Table 2). Additionally, we generated a heatmap of the probability distribution by month and cluster number (see Figure 10).

Table 2

Most to Least Common Clusters for Each Month

Month	**Most Common Cluster → → → → → → → → → → → → Least Common Cluster**							
Jan	1	14	15	4	6	3	-	-	-
Feb	1	15	4	14	6	-	-	-	-
Mar	1	14	13	4	7	3	5	-	-
Apr	1	2	15	4	14	12	3	-	-
May	1	2	15	14	10	5	6	8	-
Jun	1	14	2	15	8	9	10	-	-
Jul	2	12	11	14	13	5	9	8	-
Aug	14	2	1	12	8	3	9	-	-
Sep	2	14	1	15	13	9	5	8	-
Oct	1	2	11	14	12	13	6	4	10
Nov	14	6	11	1	2	15	4	3	13
Dec	9	1	14	2	15	12	4	-	-

Source: developed by the author

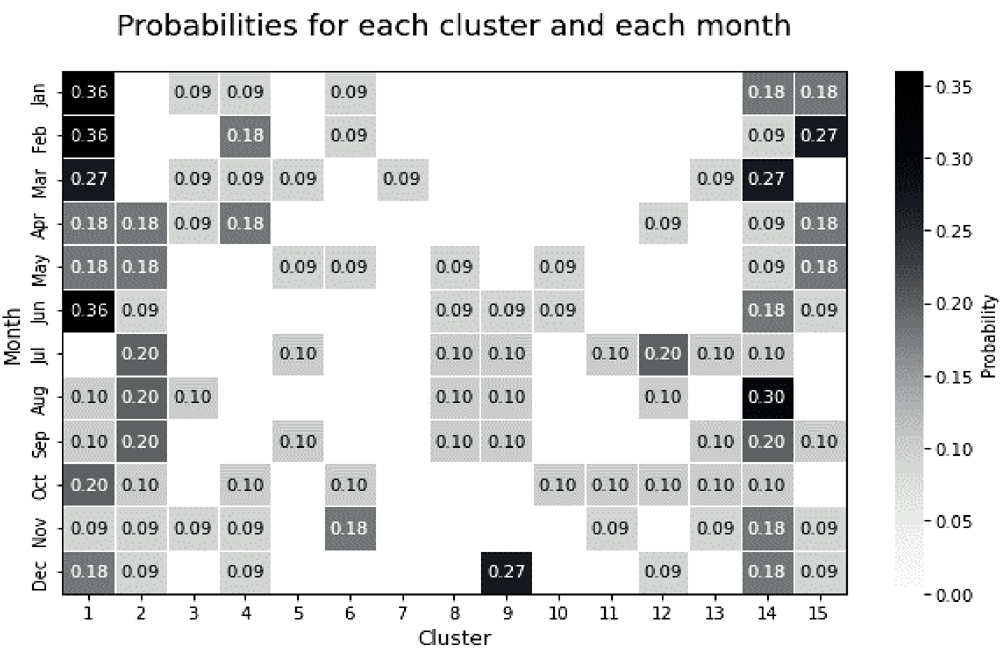


Figure 10. Heatmap of cluster probability distribution by month, topological approach
Source: developed by the author

From the configuration of typical curves for the clusters (Figure 8) and the scatterplots in the heatmap (Figure 10), we can see that for January – June, the most likely market scenario is pure backwardation (Clusters 1-6). At the same time, for July – December, the prevalence of backwardation (Clusters 1-6) or contango (Clusters 13-15) is rather unclear, as intermediate options such as «humped» contango and backwardation (Clusters 7-12) also have a relatively high probability. The intermediate pattern has the highest probability in December when short-end contango with long-end backwardation is most common. In July, intermediate variations – namely, «humped» contango – are also more typical.

The results indicate that the prevalence of a particular market situation in a given month in the mid-term is seasonal, and thus follows specific periodic patterns. The predominance of pure backwardation in the first half of the year, as well as relatively high probabilities of «humped» contango in July and «humped» backwardation in December, in our opinion, are the result of a combination of several factors, namely changes in weather and temperature conditions, expectations of industrial oil consumers about the demand for their products, as well as the cyclical accumulation and consumption of commercial ending stock of oil and the peculiarities of production processes. Our assumption is consistent with the commonly accepted view of the factors determining the crude oil market conditions, including their seasonal components, such as supply and demand balances, physical production considerations, and regional temperatures (e.g., see [2]). In the next part of the study, we will analyze the cluster relationship between the prevalence of market situations and the cyclicality of accumulation and consumption of commercial oil stock to assess our assumption.

Clustering of commercial ending stock levels. To analyze the processes of accumulation and consumption of commercial oil stocks, we used monthly data on commercial crude oil ending stocks at the Cushing hub in OK, US, from November 2011 to June 2022, in thousand barrels, obtained from the EIA database [8]. Commercial stocks at the Cushing, OK, hub are «the nexus of oil fundamentals», as fairly described by D. Brusstar and R. Karas [6], that serves as the core element of the global market for the benchmark NYMEX WTI contracts, providing the delivery mechanism for such contracts and offering large storage and transporting facilities for oil producers, refiners, traders, and other market participants. Descriptive statistics of the input data are presented in Table 3.

Table 3

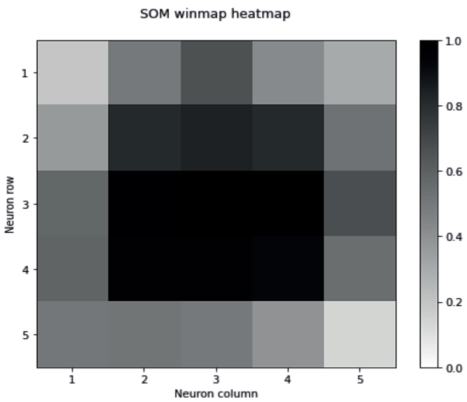
Descriptive Statistics of Input Data

	Count	Mean	Std	Min	25%	50%	75%	Max
Value	128.00	45,225.98	13,805.39	17,946.00	35,230.75	46,163.00	56,928.50	69,414.00

Source: developed by the author

To perform the cluster analysis of the commercial stock levels, we normalized the data using the feature scaling method, i.e., we scaled the input data so that their values were in the range [0, 1]. To obtain a more accurate clustering result, we artificially increased the dimensionality of the SOM grid so that $som_x = som_y = 5$. Next, we set the SOM grid parameters by increasing the learning factor to 5 (see Formula 3), which is a purely empirical adjustment. The number of training iterations was left unchanged, i.e., 100,000. The resulting SOM neuron distance map and graphical interpretation of clustering are shown in Figures 11 and 12, respectively.

$$SOM = SOM(som_x, som_y, len(data), \sigma = 1, \alpha = 5, \theta = 'Gauss', s = 0) \quad (3)$$



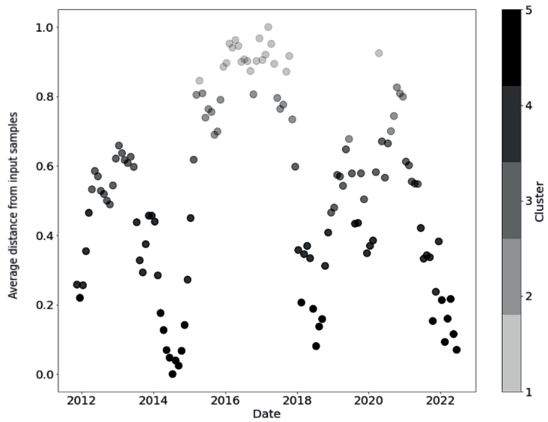


Figure 11 (left). SOM Distance Map

Figure 12 (right). Cushing Ending Stocks of Crude Oil Clusters

Source: developed by the author

The commercial oil ending stock levels were distributed into 5 clusters. Normalization of the input data allows us to nominally characterize the clusters as follows: Cluster 1 – very high level of stocks, Cluster 2 – high level of stocks, Cluster 3 – moderate level of stocks, Cluster 4 – low level of stocks, Cluster 5 – very low level of stocks (see Table 4).

Table 4

Descriptive Statistics of Commercial Oil Ending Stock Clustering Results

Cluster	Count	Mean	Std	Min	25%	50%	75%	Max
1	21	0.92	0.04	0.85	0.90	0.91	0.95	1.00
2	19	0.76	0.05	0.68	0.74	0.76	0.80	0.83
3	35	0.58	0.05	0.47	0.54	0.58	0.61	0.67
4	31	0.36	0.07	0.24	0.33	0.36	0.43	0.46
5	22	0.12	0.07	0.00	0.07	0.13	0.17	0.22

Source: developed by the author

To evaluate the relationship between the results of shape-based clustering of forward curves (using the topological approach) and commercial ending stock levels, we calculated the value of the determination coefficient, R^2, between the corresponding cluster numbers for each date within the studied period. The value of the coefficient is 0.67, which indicates a rather significant interdependence of the variables. When constructing heatmaps using the two clustered indicators, we see that for periods with very low, low, and moderate levels of ending stocks (Clusters 3-5), a situation of backwardation (Clusters 1-6) is more typical, while for periods with high and very high levels of ending stocks (Clusters 1-2), a situation of «humped» contango (Cluster 12) or pure contango (Clusters 13-15) is rather typical. The results confirm the existence of a considerable interdependence between the studied variables.

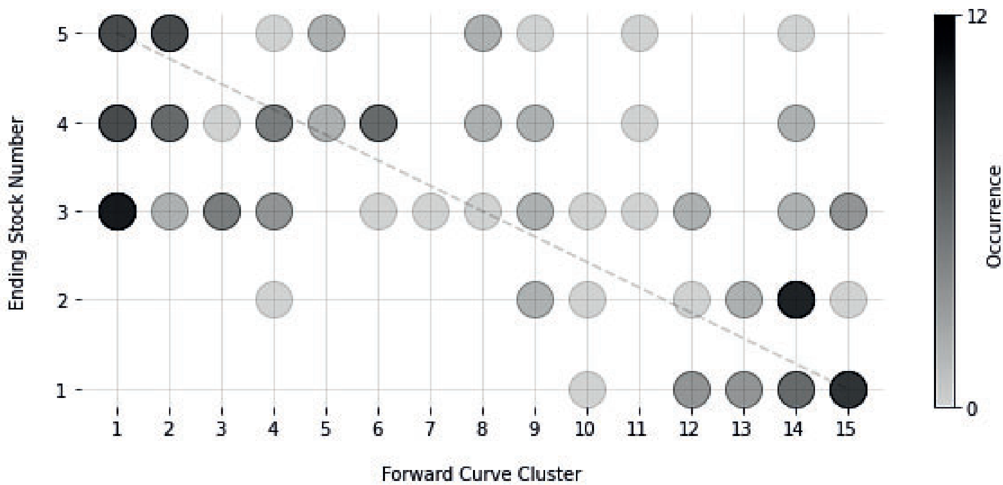


Figure 13. Forward Curve and Ending Stock Cluster Analysis Comparison
Source: developed by the author

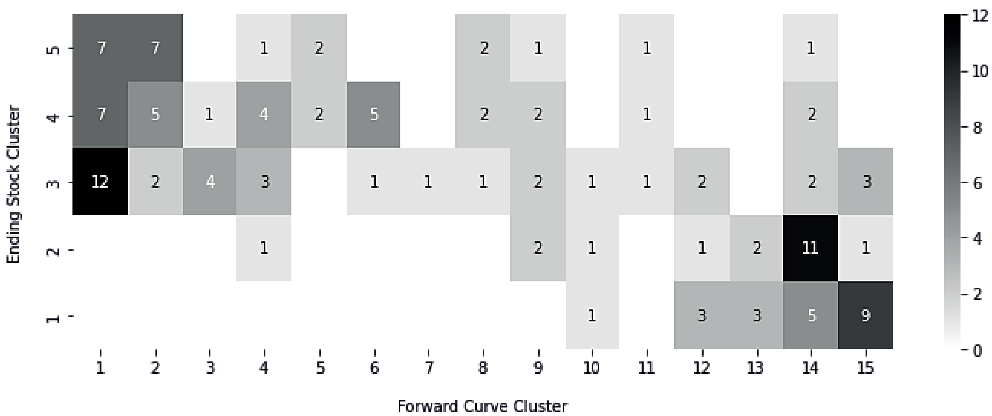


Figure 14. Forward Curve and Ending Stock Cluster Analysis Comparison
Source: developed by the author

Conclusions and prospects for further research. Our study suggests a new approach to the analysis of the crude oil prices term structure, namely cluster analysis of forward curves based on machine learning. In this study, we identified seasonal features of term pricing in the crude oil market using SOM-based cluster analysis. Using two approaches to shape-based clustering, namely quantitative and topological, we demonstrated the existence of cyclical patterns in the term structure of oil futures prices and traced their dependence on commercial ending stock levels.

The application of cluster analysis using artificial neural networks (ANNs) provides a wide range of opportunities to study the basic patterns and phenomena of term pricing, particularly for crude oil and other commodities. Identification of cyclical, seasonal patterns of futures price behavior and assessment of fundamental market interdependencies allows applying new and improving existing predictive models to make rational and effective investment and management decisions, explaining the behavior of consumers and producers in the crude oil and petroleum products market. Moreover, we believe that our approach to comparative clustering can be used in further examination of terms structures of other commodity and financial markets in relation to the peculiarities of interest rates, stock indices, exchange rate dynamics, and pricing of alternative investment assets.

We are of the opinion that further research should focus on the changes in seasonal patterns and patterns over time to trace the evolution of pricing trends in the oil futures market in the context of the contango-backwardation dichotomy, as well as to better understand the fundamental drivers of such changes.

References

1. Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015). Time-series clustering – A decade review. *Information Systems*, *53*, 16–38. https://doi.org/10.1016/j.is.2015.04.007.
2. Al Sadoon, H., Ward, C., Considine, J., & Al Dayel, A. (2019). *A Short-Term Forecasting Model for Brent Oil Prices*. King Abdullah Petroleum Studies and Research Center. https://doi.org/10.30573/ks--2019-mp08.
3. Barreto, G. A. (2007). Time Series Prediction with the Self-Organizing Map: A Review. У *Perspectives of Neural-Symbolic Integration* (с. 135–158). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73954-8_6.
4. Baruník, J., & Malinská, B. (2016). Forecasting the term structure of crude oil futures prices with neural networks. *Applied Energy*, *164*, 366–379. https://doi.org/10.1016/j.apenergy.2015.11.051.
5. Blazejewski, A., & Coggins, R. (2004). Application of self-organizing maps to clustering of high-frequency financial data. *Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation*, *Vol. 32*, 85–90. https://dl.acm.org/doi/10.5555/976440.976452.
6. Brusstar, D., & Karas, R. (2020, 28 вересня). *Why The Cushing Storage Hub Matters to Oil*. U.S. https://www.reuters.com/article/sponsored/storage-hub-matters-to-oil.
7. Crude Oil WTI Prices and Crude Oil WTI Futures Prices. *Barchart*. https://www.barchart.com/futures/quotes/CL*0/futures-prices.
8. Cushing, OK, Ending Stocks of Crude Oil (Thousand Barrels). *U.S. Energy Information Administration (EIA).* https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=MCRST_YCUOK_1&f=M.
9. Dablemont, S., & et al. (2003). Time series forecasting with SOM and local non-linear models - Application to the DAX30 index prediction. *WSOM'2003 proceedings - Workshop on Self-Organizing Maps Hibikino (Japan)*, 340–345. https://perso.uclouvain.be/michel.verleysen/papers/wsom03sd.pdf.
10. Hammer, B., & Hitzler, P. (Eds.). (2007). Perspectives of neural-symbolic integration (Vol. 77). Springer. https://www.researchgate.net/publication/242506480.
11. Huhtala, Y., Karkkainen, J., & Toivonen, H. T. (1999). Mining for similarities in aligned time series using wavelets. *Proceedings, SPIE, Data Mining and Knowledge Discovery: Theory, Tools, and Technology*, *Vol. 3695*. https://doi.org/10.1117/12.339977.
12. Izakian, H., & Pedrycz, W. (2013). Anomaly Detection in Time Series Data using a Fuzzy

C-Means Clustering. *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, 1513–1518. https://doi.org/10.1109/IFSA-NAFIPS.2013.6608627.
13. Lee, J. A., & Verleysen, M. (2002). Self-organizing maps with recursive neighborhood adaptation. *Neural Networks*, *15*(8-9), 993–1003. https://doi.org/10.1016/s0893-6080(02)00073-4.
14. Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern Recognition*, 38(11), 1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025.
15. Maharaj, E. A., D'Urso, P., & Caiado, J. (2019). *Time Series Clustering and Classification*. Taylor & Francis Group. https://doi.org/10.1201/9780429058264.
16. Nordlinder, M. (2021). *Clustering of Financial Account Time Series Using Self Organizing Maps* [Master's Thesis, KTH Royal Institute of Technology]. https://www.diva-portal.org/smash/get/diva2:1537700/FULLTEXT01.pdf.
17. Vettigli, G. (2018). *MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map*. https://github.com/JustGlowing/minisom/.

КЛАСТЕРНИЙ АНАЛІЗ СТРОКОВОЇ СТРУКТУРИ ЦІН НА НАФТУ З ВИКОРИСТАННЯМ КАРТ КОХОНЕНА

Едуард Колотвін, Ірина Грабинська

Львівський національний університет імені Івана Франка,
79008, м. Львів, просп. Свободи, 18,
email: eduard.kolotvin@lnu.edu.ua; ORCID: https://orcid.org/0009-0000-6622-411X
email: iryna.hrabynska@lnu.edu.ua; ORCID: https://orcid.org/0000-0001-7390-5626

Анотація. У цьому дослідженні ми застосовуємо кластеризацію на основі самоорганізаційної карти (SOM), також відомої як карта Кохонена, до щомісячних форвардних кривих ф'ючерсних цін на сиру нафту та залишкових запасів сирої нафти на хабі в Кушингу, штат Оклахома, за період з листопада 2011 року по червень 2022 року, аби дослідити шаблони режимів контанго-беквордації та їхній зв'язок із сезонністю та рівнями запасів. Зокрема, ми пропонуємо два підходи до кластеризації форвардних кривих залежно від їх форми, а саме чисельний і топологічний. Наші результати показують, що за допомогою кластеризації методами SOM можна виявити характерні особливості строкової структури ф'ючерсних цін у розрізі місяців, причому деякі місяці, наприклад, з січня по червень, демонструють чітку поведінку, а інші, наприклад, з липня по грудень, – досить неоднозначну. Крім того, ми пропонуємо підхід до порівняльного аналізу сезонних тенденцій та фундаментальних чинників, що лежать в їх основі, а саме рівнів кінцевих запасів. Зокрема, ми доводимо, що для періодів, коли комерційні запаси перебувають на дуже низькому або помірному рівні характернішою є ситуація беквордації, тоді як для періодів із високими або дуже високими запасами характернішою є ситуація контанго. В цілому, наше дослідження пропонує нові методи та підходи до аналізу поведінки ф'ючерсних цін на нафту, а саме їх строкового аспекту, що проявляється в конфігурації форвардних кривих, наголошуючи на важливості місячних шаблонів та сезонних режимів, а також доводить обґрунтованість використання методів машинного навчання та порівняльного кластерного аналізу для розуміння динаміки ф'ючерсних цін на сиру нафту.

Ключові слова: строкова структура ф'ючерсних цін на сиру нафту, контанго, беквордація, форвардна крива, кластерний аналіз на основі геометричних форм, кластеризація цін, часові ряди, самоорганізаційна карта Кохонена (SOM), машинне навчання, штучні нейронні мережі (ШНМ), динамічна трансформація часової шкали (DTW), кінцеві комерційні запаси нафти.

Стаття надійшла до редакції 23.02.2023
Прийнята до друку 29.06.2023