Clustering commodity markets in space and time: Clarifying returns, volatility, and training regimes through unsupervised machine learning

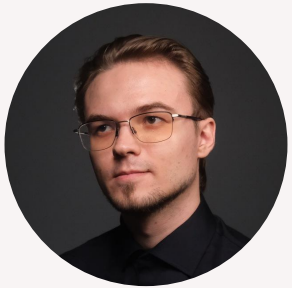
Paper by J.M. Chen, M.U. Rehman and X.V. Vo

Presentation by Dekkusheva A. (212), Kulakov D. (212), Pronina A. (212), Spirina M. (211) and Tomaily T. (211)

TABLE OF CONTENTS

Meet our team

Dekkusheva Alina



Kulakov Denis

Pronina Anna

Tomayly Tatiana

Spirina Mayya

What are commodity markets?

A commodity market is a market that **trades in the primary economic sector** rather than manufactured products.

Commodity markets represent a **quarter** of global trade in goods.

Commodity markets are the **most important source of income** for some of the world's **poorest** countries.

Goal of the research

Use unsupervised machine learning to define crises and commodities comovement

In our case:

- Gold
- Silver
- Platinum
- Palladium
- Copper
- Zinc
- Tin
- Lead
- Nickel
- Aluminum

- Brent
- WTI
- Gasoil
- Gasoline (petrol)
- Wheat
- Corn
- Soybeans
- Coffee
- Cocoa
- Cotton
- Lumber

Raw prices of commodities

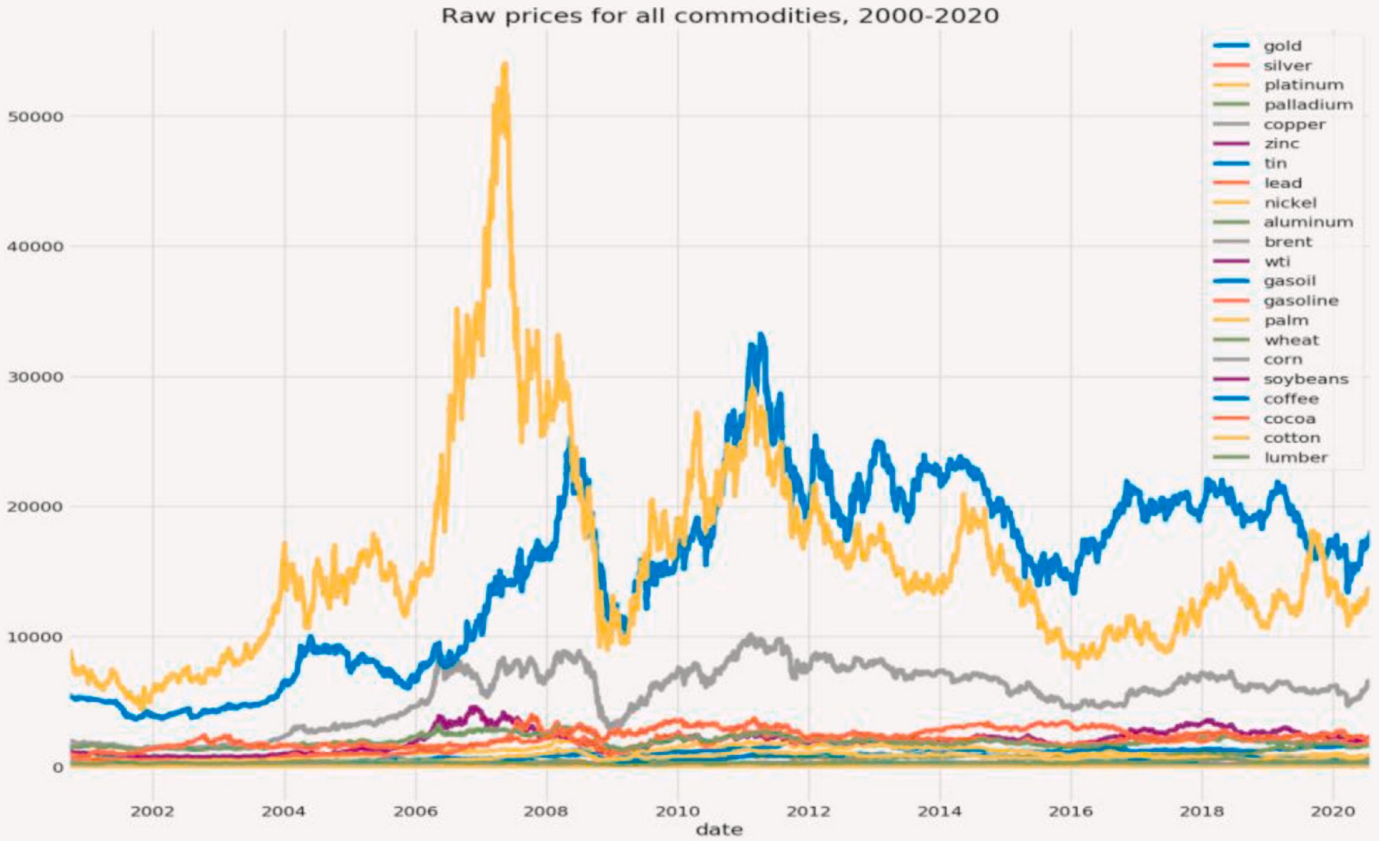


Fig. 1. Commodities pricing.

Descriptive statistics

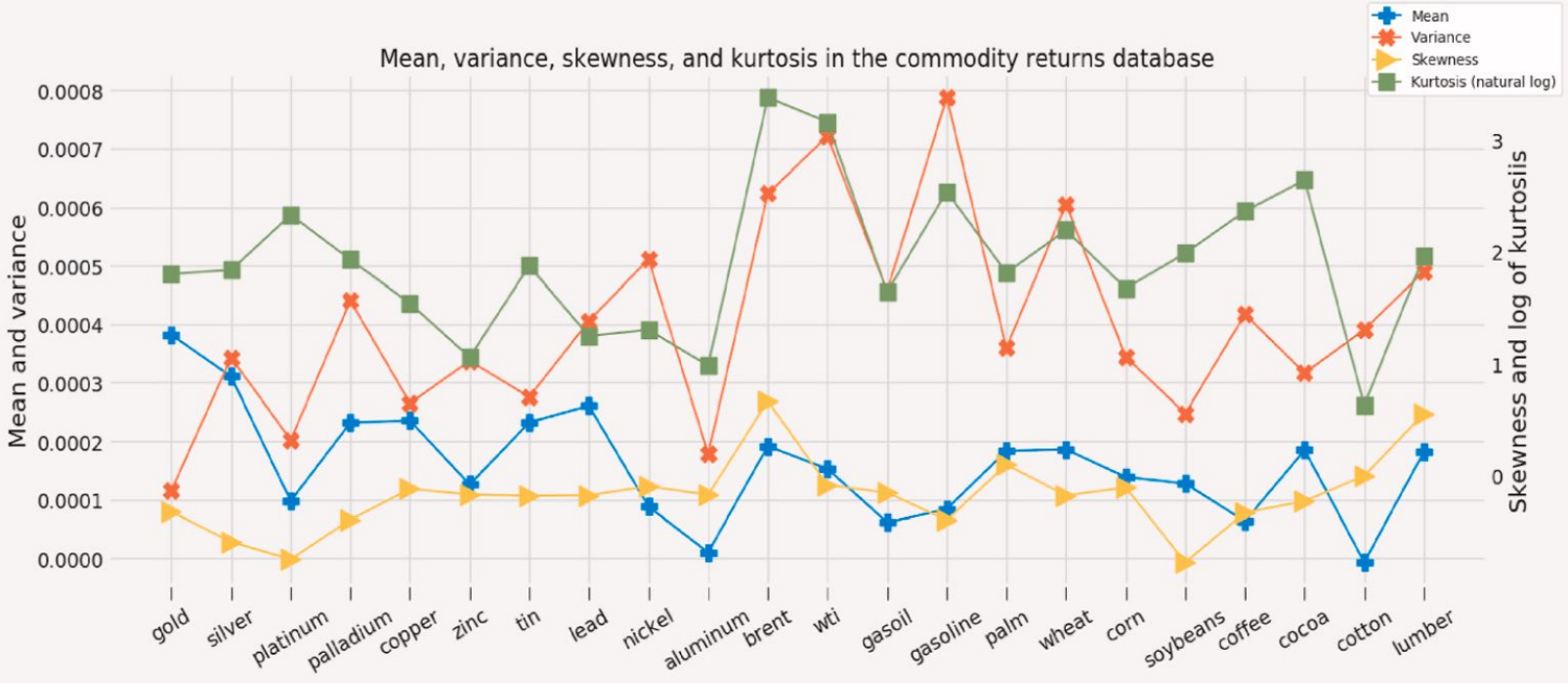


Fig. 3. Descriptive statistics.

Unconditional time-invariant correlation

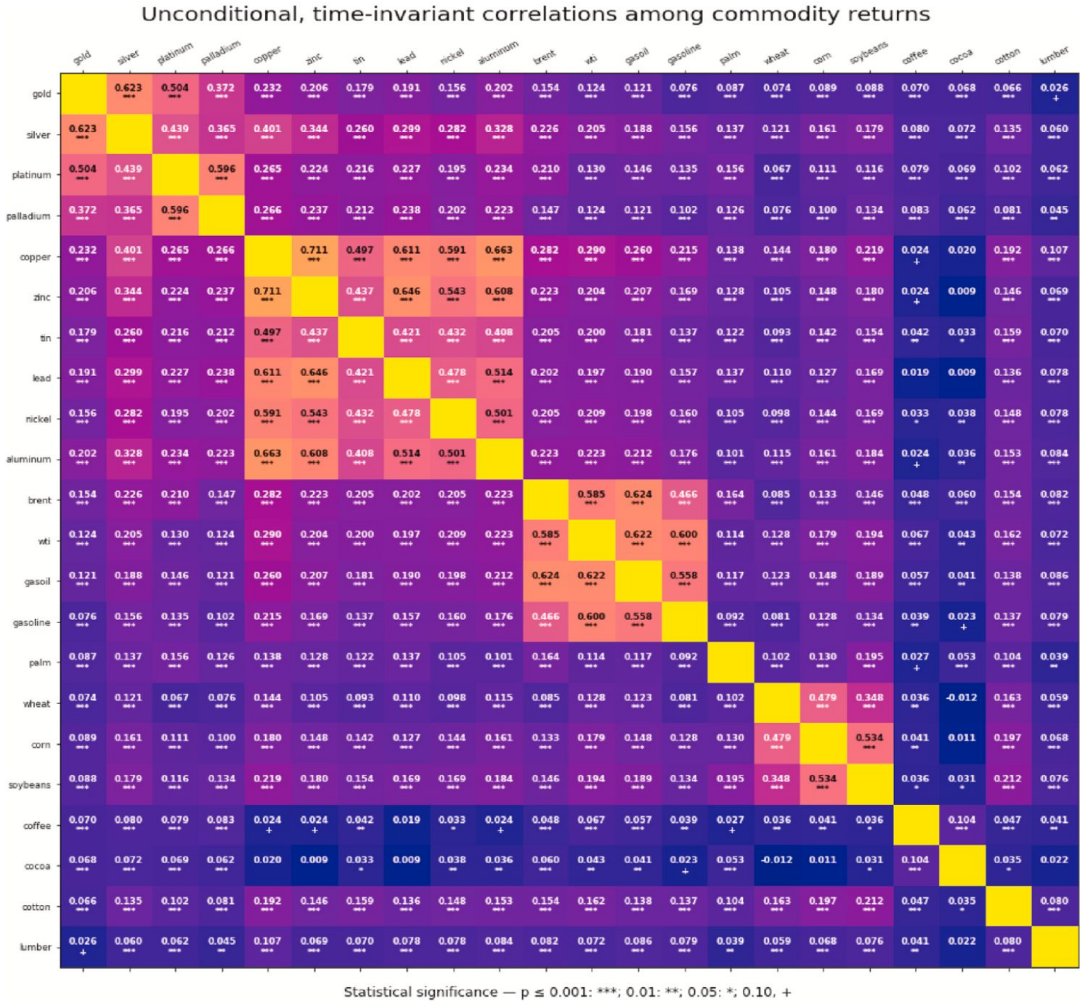


Fig. 4. Unconditional time-invariant correlation.

Cumulative log returns

```
Logarithmic Return
= ln(Present Value / Past Value)
```

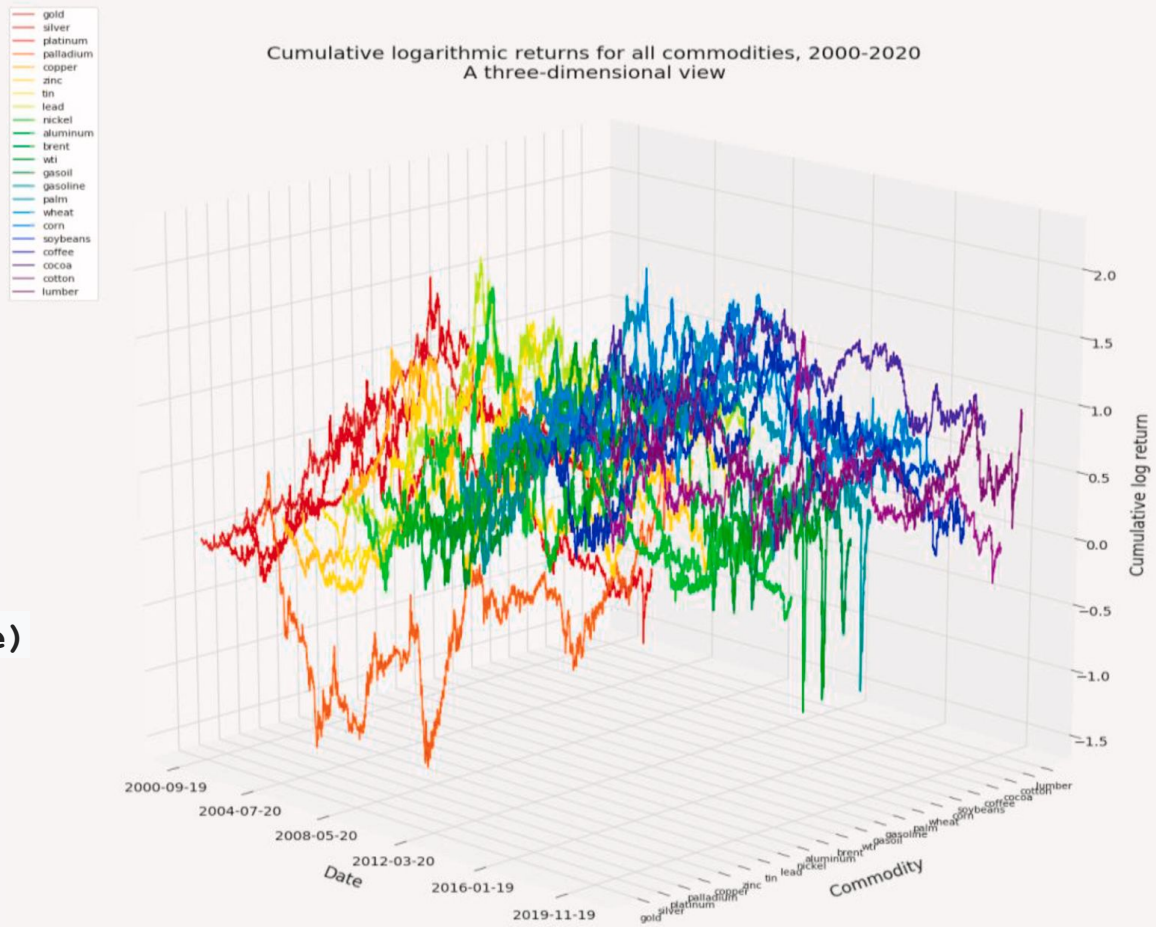


Fig. 2. Cumulative log returns.

Conditional volatility with GJR-GARCH(1, 1, 1)

Volatility refers to the **level of risk** in a financial or economic time series.

Conditional volatility means that the volatility is allowed to **change over** time **based on past observations.**

The most common method for modeling conditional volatility is through the use of **GARCH** models.

Table 1
GJR-GARCH model results.

Dep. Variable:	nickel		
Mean Model:	Zero Mean	R-squared:	0.000
		Adj. R-squared:	0.000
Vol Model:	GJR-GARCH	Log-Likelihood:	−11084.2
Distribution			
Standardized Student's t		AIC:	22178.5
Method:			
Maximum Likelihood		BIC:	22211.2
No. Observations:	5182		
Df Residuals: 5177			
Df Model:	5		

Volatility Model				
Coefficient	*Std. err*	*t*	*P>|t|*	*95.0% Conf. Int.*
*O*mega	0.0474	1.890e-02	2.507	1.216e-02 [1.035e-02,8.444e-02]
alpha [1]	0.0412	9.343e-03	4.413	1.019e-05 [2.292e-02,5.954e-02]
gamma [1]	1.9531e-03	8.061e-03	0.242	0.809 [-1.385e-02,1.775e-02]
beta [1]	0.9485	1.172e-02	80.943	0.000 [0.925, 0.971]

Distribution				
coefficient	*Std. err*	*t*	*P>|t|*	*95.0% Conf. Int.*
Nu	6.8175	0.602	11.327	9.640e-30 [5.638, 7.997]

Spatial clustering by commodity

K-means clustering

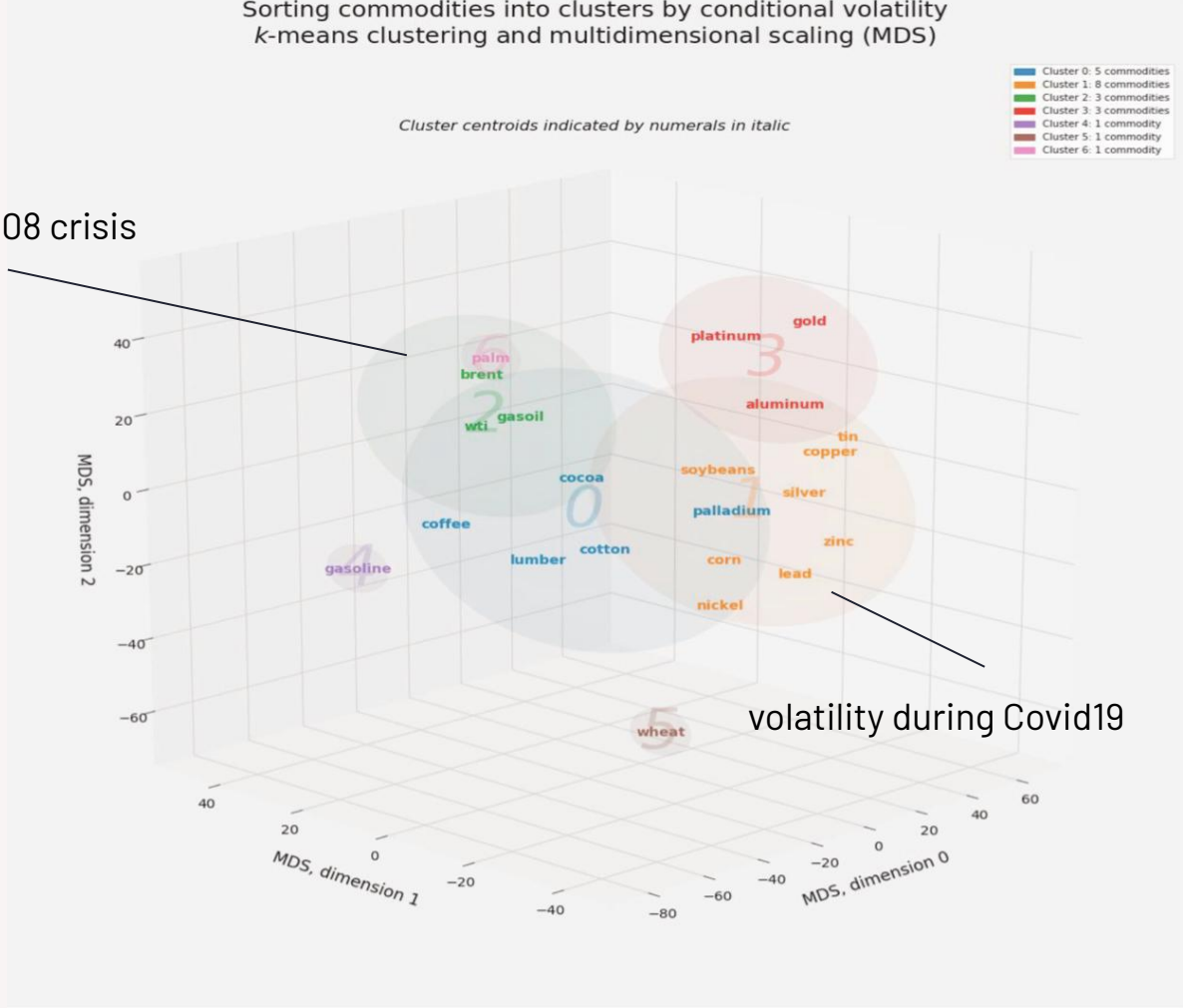
an algorithm that **depends on distances** among volatility/return series **for different commodities**

Multidimensional scaling (MDS)

an algorithm that **projects clusters** into **three dimensions** capable of human perception

k-means + MDS on conditional volatility

volatility during 2008 crisis



volatility during Covid19

Fig. 8. *K*-means clustering and MDS- conditional volatility.

k-means + MDS on logarithmic returns

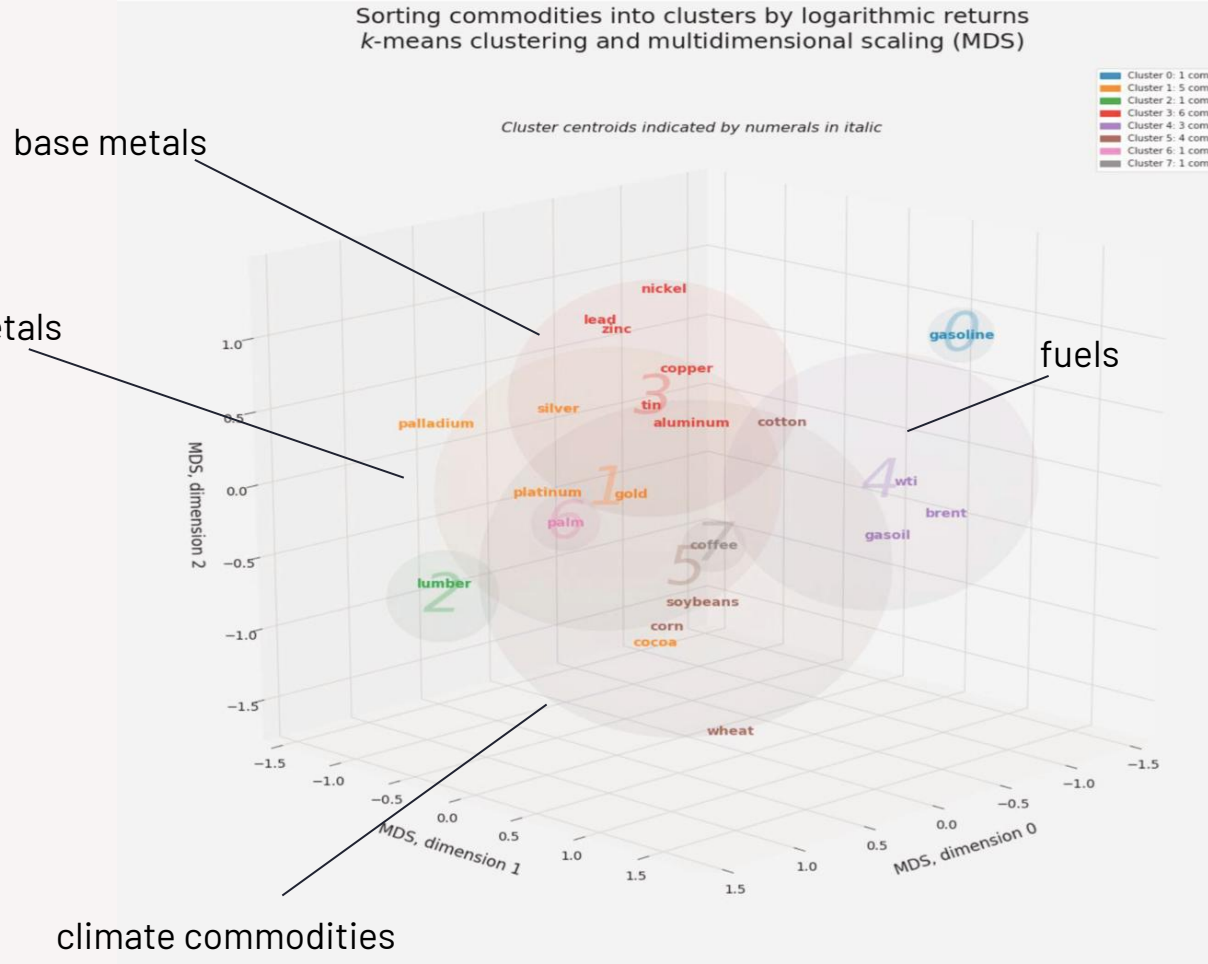
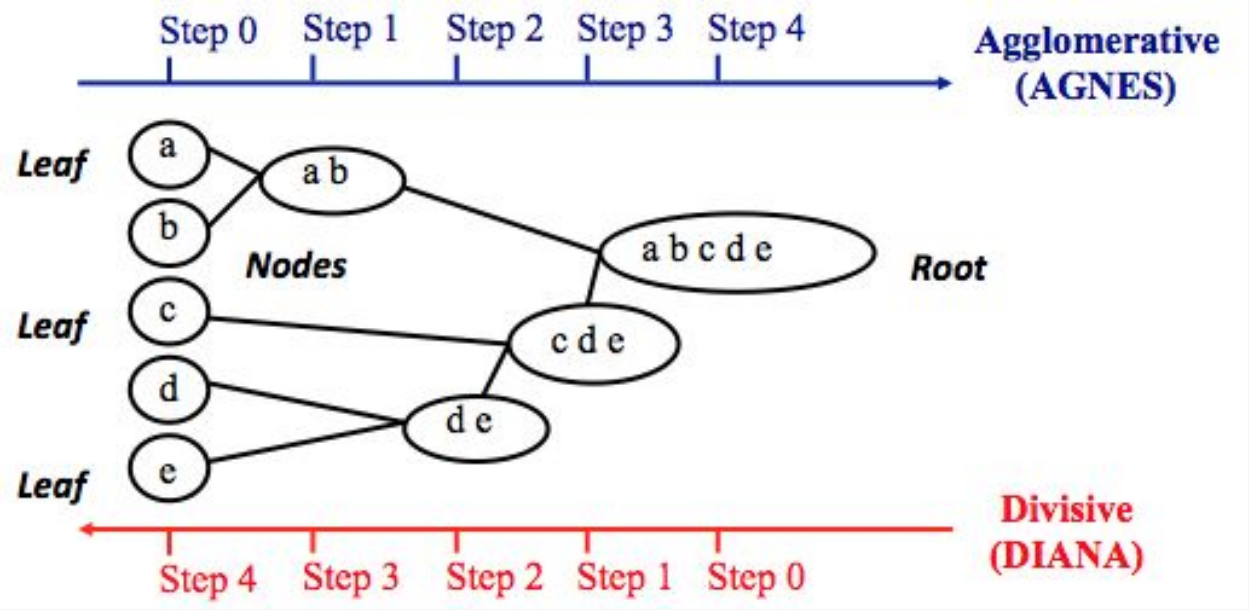
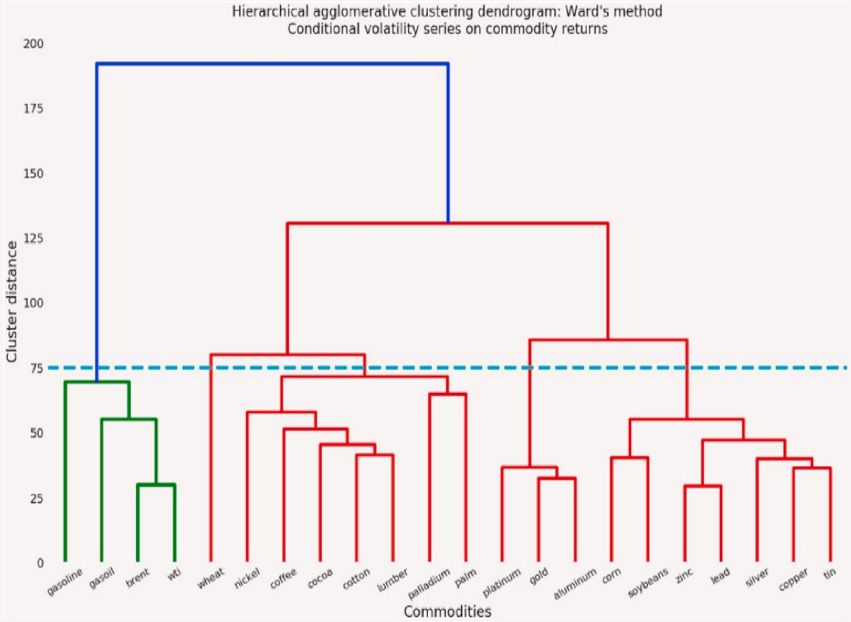


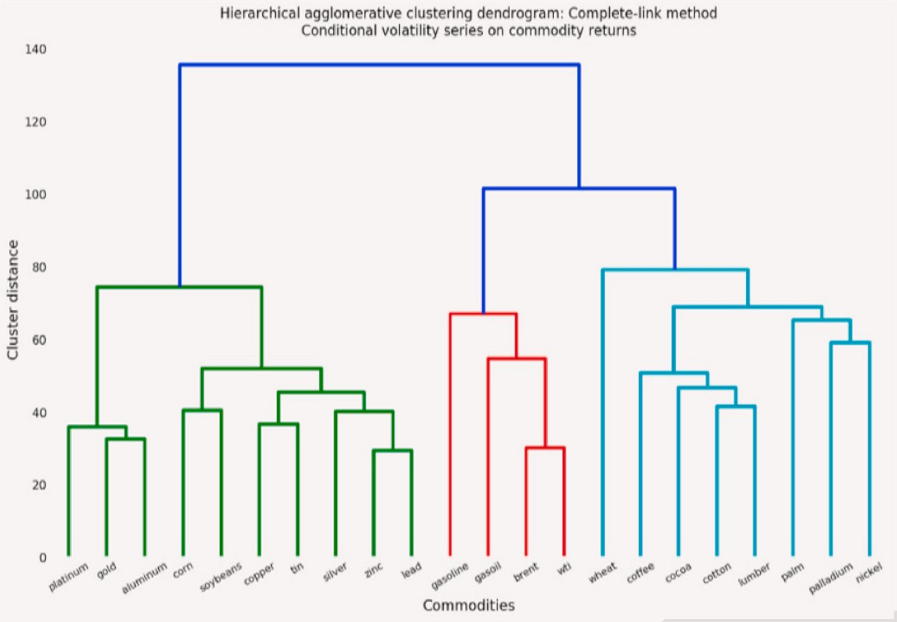
Fig. 9. *K*-means clustering and MDS- log returns.

Hierarchical agglomerative clustering

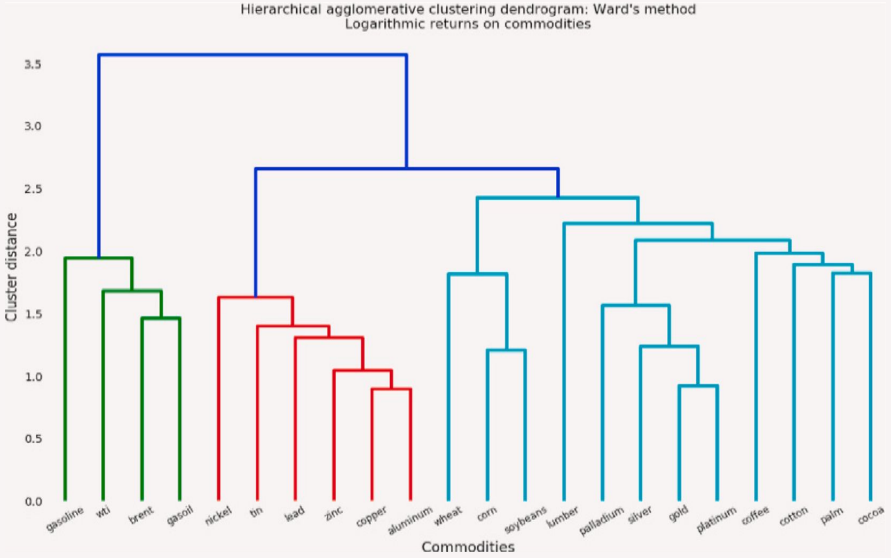


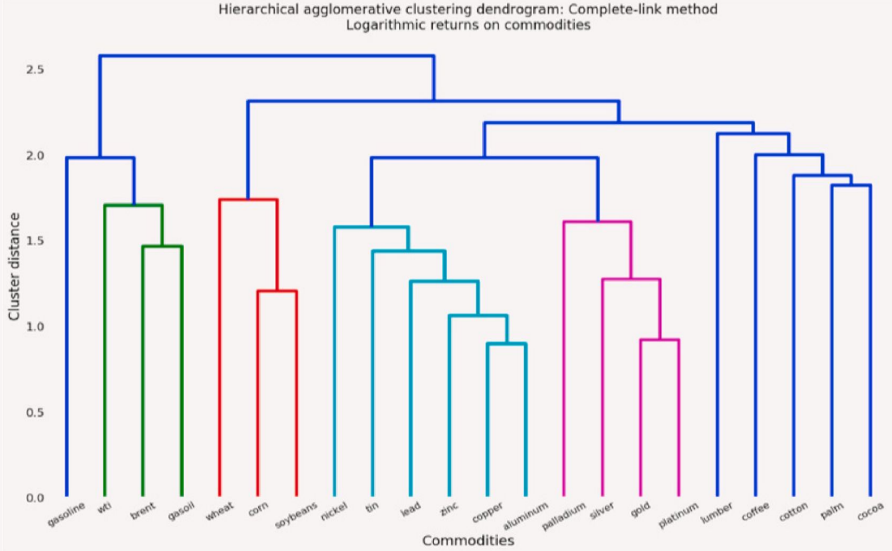
Conditional volatility





Logarithmic returns





Temporal clustering of trading days

Transposition of the conditional volatility matrix

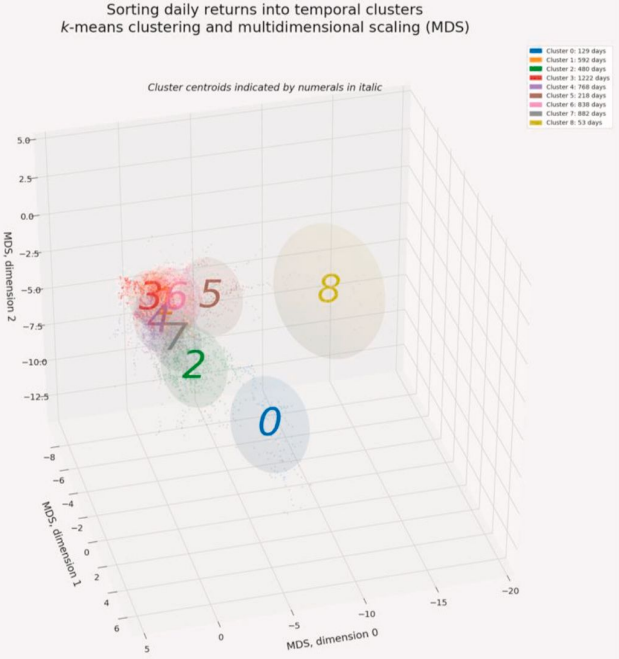


Fig. 14. *K*-means clustering and MDS- returns.

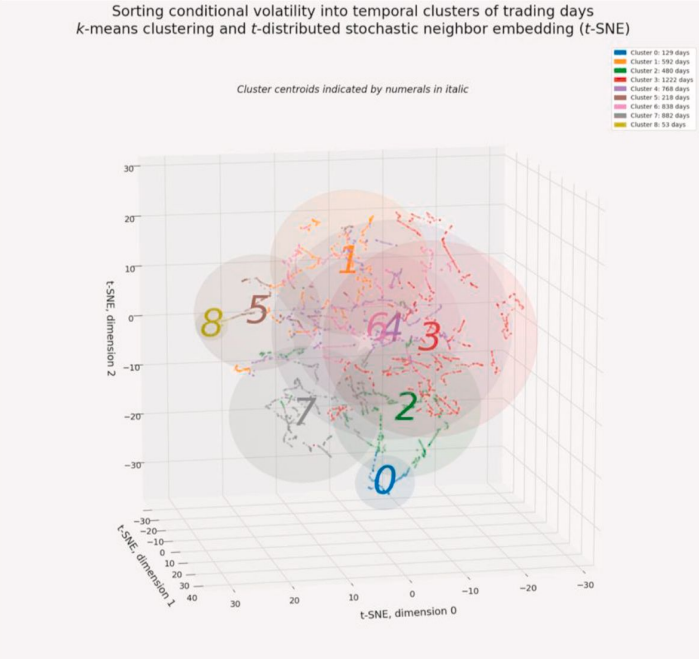


Fig. 15. *K*-means clustering and t-SNE-trading days.

Temporal clustering of trading days

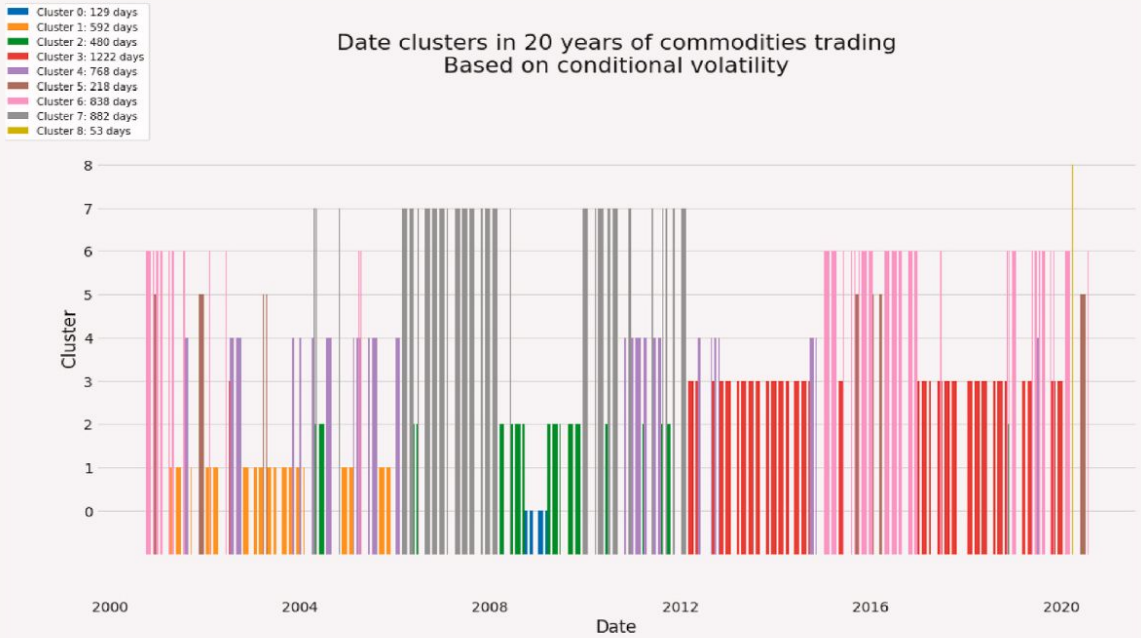
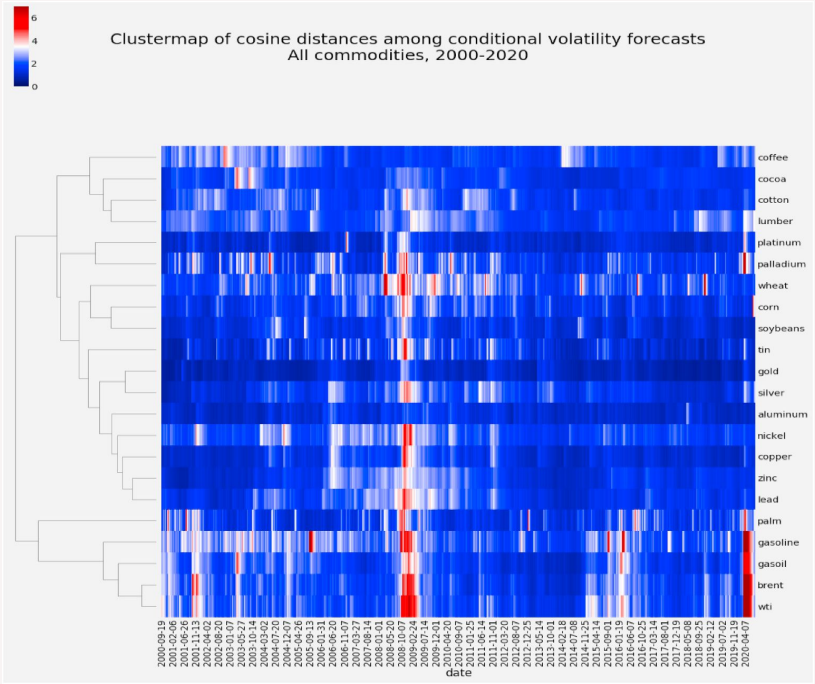
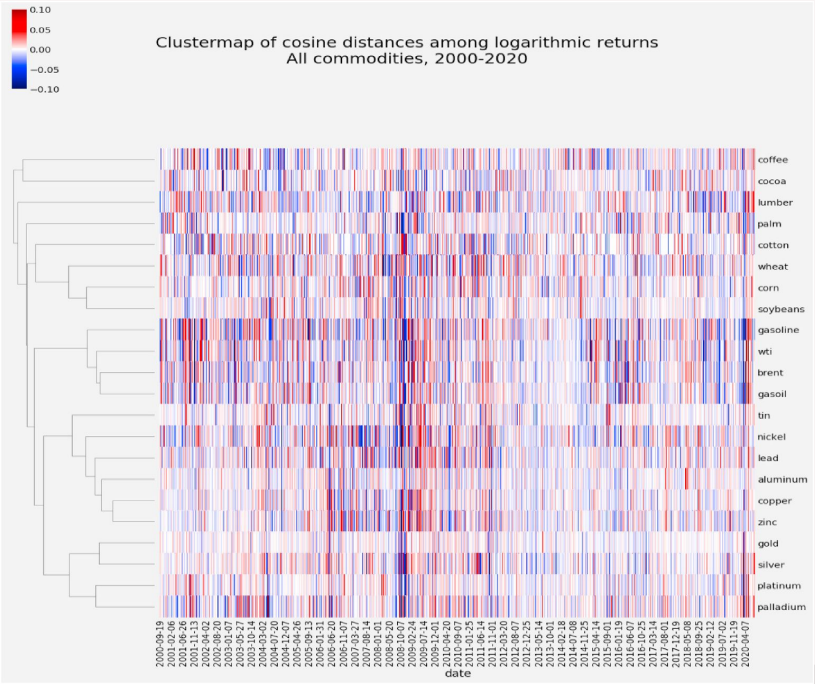


Fig. 16. Data clusters of commodities-conditional volatility.

Cosine distance clustermaps





Conclusions

- Unsupervised learning, particularly hierarchical clustering, can be used to create a structure or "ontology" of how commodities relate in financial space. This helps in understanding how different commodities move together or independently.

- The text emphasizes the importance of such clustering for risk management. It helps in containing risks related to the simultaneous movement of commodities and the spillover of volatility. Identifying commodities that don't move together is crucial.

- Differences in results obtained using clustering based on logarithmic returns and clustering based on conditional volatility provide valuable insights into commodity market dynamics.

- The volatility-based clustering method is highlighted as particularly valuable in understanding comovement in commodities. It helps identify distinct regimes during critical periods like financial crises.

Recommendations and further prospects

- Exploring other unsupervised learning methods like kernel density estimation and Gaussian mixture modeling. Additionally, there's the idea of combining clustering with time-series forecasting to enhance decision-making.
- The text suggests that unsupervised learning should be used to analyze higher-moment phenomena in financial data. For example, clustering could be applied to conditional skewness and kurtosis to capture more complex relationships in data.
- Lastly, the text proposes evaluating measures like expected shortfall and correlation during known or suspected critical periods. This can help identify distinct trading regimes during times of calm and extreme disruption in commodity markets.

THANKS!

TELEGRAM

@ALINA_DEKK

@L_ANNAPRONINA_L

@KULAKOV_DENIS

@MAYYASP

@TTLY25