Predicting Market Reactions to News:
An LLM-Based Approach Using Spanish Business Articles

Jesus Villota Miranda (*CEMFI*)

This paper explores how Large Language Models (LLMs) can enhance market prediction by analyzing Spanish business news during the volatile COVID-19 period. We propose a novel approach that guides LLMs to systematically classify economic shocks in news articles, comparing its effectiveness against a traditional vector-based text analysis method in predicting market reactions

Methodology

Data: Spanish Business Articles referred to IBEX-35 firms | Publisher: DowJones | Period: 07/2020 - 09/2021

Embeddings-Trading

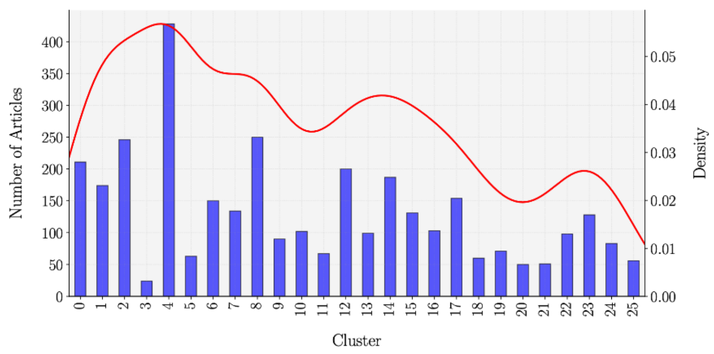
1) Map the wording of an article i into an embedding vector $\mathbf{e}^i \in \mathbb{R}^{512}$
2) Cluster the embeddings using `KMeans` (# clusters = 26)
3) For each article i, extract the set of affected firms $\mathcal{F}^i$ using pattern recognition
4) Long-Short Cluster Trading of each $(i, j),\ j \in \mathcal{F}^i$

LLM-Trading

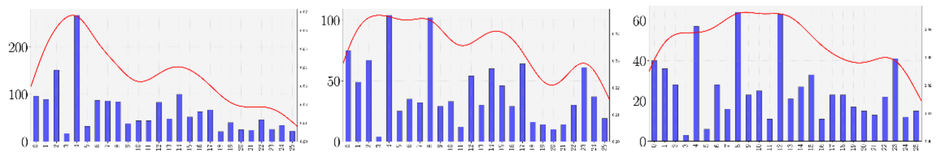
1) For each article i, ask the LLM to identify the set of affected firms $\mathcal{F}^i_{LLM}$
2) For each firm $j \in \mathcal{F}^i_{LLM}$, ask the LLM to classify the shock that the article implies in j
3) Cluster the articles based on the shock classification (# clusters = 20)
4) Long-Short Cluster Trading of each $(i, j),\ j \in \mathcal{F}^i_{LLM}$

Distribution of Articles through Clusters

(A) All data ($\mathcal{D}$)



(B) Training data ($\mathcal{D}^{tr}$) (C) Validation data ($\mathcal{D}^{val}$) (D) Test data ($\mathcal{D}^{test}$)



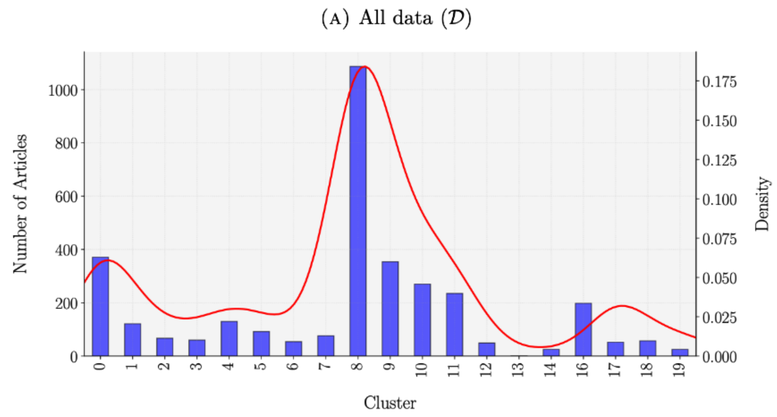
Unstable clustering

*The distribution profile of articles through clusters is **unstable** across data splits*

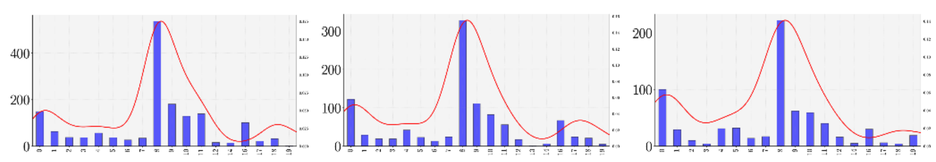
Stable clustering

*The distribution profile of articles through clusters is **stable** across data splits*

(A) All data ($\mathcal{D}$)



(B) Training data ($\mathcal{D}^{tr}$) (C) Validation data ($\mathcal{D}^{val}$) (D) Test data ($\mathcal{D}^{test}$)



Trading Signal by Cluster

	Cluster	Greedy	Stable
0	Miscellaneous (Colonial, Acciona, Amadeus, Grifols, Endesa, IAG, Bankinter...)	SHORT	
1	Quarterly & Semi-Annual Earnings Reports	SHORT	
2	BBVA & Sabadell: Financial Performance & Strategic Movements	SHORT	
3	Telefónica & Cellnex: Telecommunications Tower Sales & Market Dynamics	LONG	LONG
4	CaixaBank: Mergers and Strategic Moves in the Banking Sector		
5	Telefónica, Indra, & MásMóvil: Regulatory and Strategic Moves in Telecom	LONG	
6	Siemens Gamesa: Supply Agreements, Profitability Targets in Renewable Energy	SHORT	
7	Cellnex: Strategic Acquisitions and Financial Moves in Telecom Infrastructure	LONG	
8	Acciona, Endesa, Enagás & Naturgy: Strategic Moves & Regulatory Developments in the Energy Sector	LONG	
9	Repsol: Strategic Moves and Challenges in the Energy Sector	LONG	
10	Ferrovial, Acciona: Strategic Expansions and Financial Maneuvers in Infrastructure	SHORT	SHORT
11	Solaria: Strategic Moves and Market Challenges in Renewable Energy	LONG	LONG
12	Iberdrola: Strategic Collaborations and Renewable Energy Developments	SHORT	
13	IAG: Financial Performance	LONG	
14	Santander & CaixaBank: Financial Moves and Sustainability Initiatives	SHORT	
15	ACS & Acciona: Strategic Movements and Infrastructure Projects	SHORT	SHORT
16	Telefónica: Financial Performance and Strategic Moves	LONG	
17	Meliá and Spanish Tourism Sector: Challenges Amidst the Pandemic	SHORT	
18	Takeover Bids for Naturgy and MásMóvil	SHORT	
19	Naturgy: Financial Performance	SHORT	SHORT
20	PharmaMar, Grifols: Regulatory Approvals and Market Moves in the Pharmaceutical Sector	LONG	LONG
21	Repsol: Financial Performance	LONG	LONG
22	Aena: Financial Performance	LONG	LONG
23	Enagás, Endesa, Iberdrola, Red Eléctrica: Regulatory and Market Challenges in the Energy Sector	SHORT	
24	BBVA, CaixaBank, Banco Sabadell: Layoffs and Restructuring	LONG	LONG
25	Inditex, Acerinox: Market Performance and Strategic Developments in the Post-Covid Context	SHORT	SHORT

2 cluster-selection algorithms:

- **Greedy:** maximize *average Sharpe Ratio* in Validation set
- **Stable:** maximize *cluster-rank correlation* between Training & Validation sets

Short-lived Signals

*High reliance of the signal on the **past performance** of a cluster (e.g: financial performance topics)*

***Topics change through time** (e.g: Covid topics are outdated now!)*

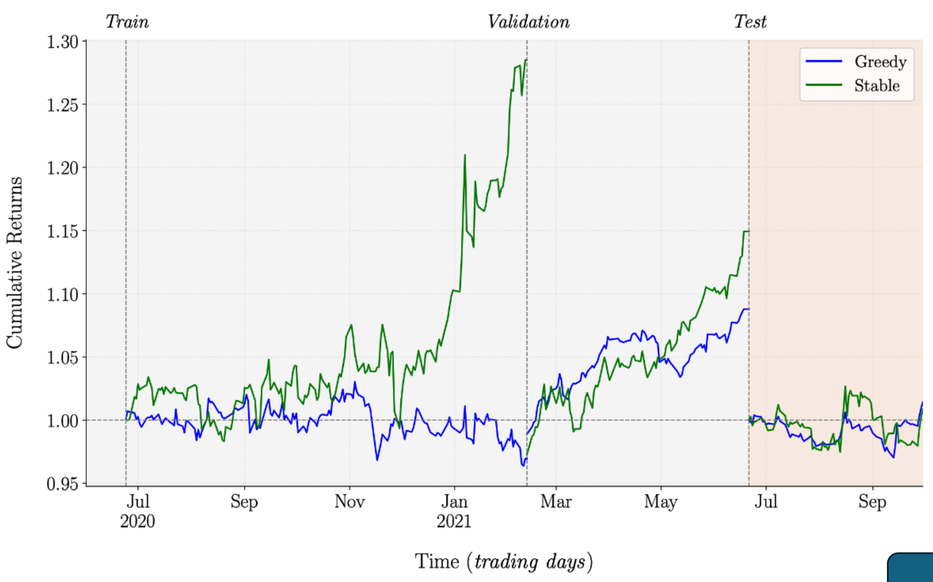
Long-lasting Signals

*Trading Signals depend on the **nature of the shock** implied on the affected firm*

Clean** & **Interpretable** trading signals that **persist over time

	Cluster	Greedy	Stable
0	(demand, minor, positive)		
1	(demand, minor, negative)		SHORT
2	(demand, major, positive)	SHORT	SHORT
3	(demand, major, negative)	LONG	LONG
4	(supply, minor, positive)	LONG	
5	(supply, minor, negative)	SHORT	
6	(supply, major, positive)	LONG	
7	(supply, major, negative)	SHORT	
8	(financial, minor, positive)	LONG	LONG
9	(financial, minor, negative)		SHORT
10	(financial, major, positive)	LONG	
11	(financial, major, negative)	SHORT	
12	(technology, minor, positive)	LONG	
13	(technology, minor, negative)		
14	(technology, major, positive)	SHORT	
15	(technology, major, negative)		
16	(policy, minor, positive)	SHORT	SHORT
17	(policy, minor, negative)	SHORT	SHORT
18	(policy, major, positive)	SHORT	SHORT
19	(policy, major, negative)	SHORT	SHORT

Returns to the Trading Strategies



Negligible out-of-sample profitability

*Topic-based trading achieves Sharpe Ratios of **0.7** & **0.2** respectively for the **Greedy** & **Stable** algorithms*

Robust out-of-sample profitability

*LLM-based trading achieves Sharpe Ratios of **4.30** & **4.39** respectively for the **Greedy** & **Stable** algorithms*

Conclusion

Our findings demonstrate that LLMs, when guided by appropriate economic frameworks, can effectively predict market reactions to news through systematic classification of financial narratives. By identifying and classifying firm-specific economic shocks embedded in business articles, our LLM approach achieves robust temporal stability and consistent profitability, significantly outperforming traditional embedding-based methods that fail to capture lasting market signals.

cemfi

jesus.villota@cemfi.edu.es | Jesus Villota-Miranda