Lintner
Tversky
Carhart
Merton

FGV

UFMG

David Alexander | Marcel Zanetti | Pedro Teles

Using Reddit and Markov Chains to Time Bitcoin

David Alexander | Marcel Zanetti | Pedro Teles

FGV

U F *m* G

Motivation

David Alexander | Marcel Zanetti | Pedro Teles

Motivation

1

Volatility Forecast

Camou (2022) shows that a sentiment analysis based on Reddit posts about cryptocurrencies has **predictive power** over crypto **volatility**. However, the results are **mixed** for predicting **returns**.

2

Markov

Poyser (2018) tests the hypothesis that cryptocurrency prices are driven by **herding**. For that, he studies herding behavior under different conditions, including the **Markov-Regime-Switching** approach.
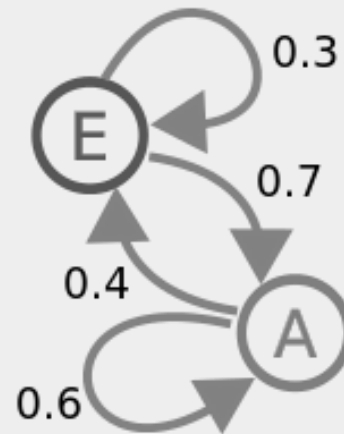
3

In this work, we use **sentiment analysis** and the **Hidden Markov Model** (Baum & Petri (1966)) to get more (less) exposure to Bitcoin when there is a **higher (lower)** chance of the market being more **bullish (bearish)** in the next period.

Data & Methodology
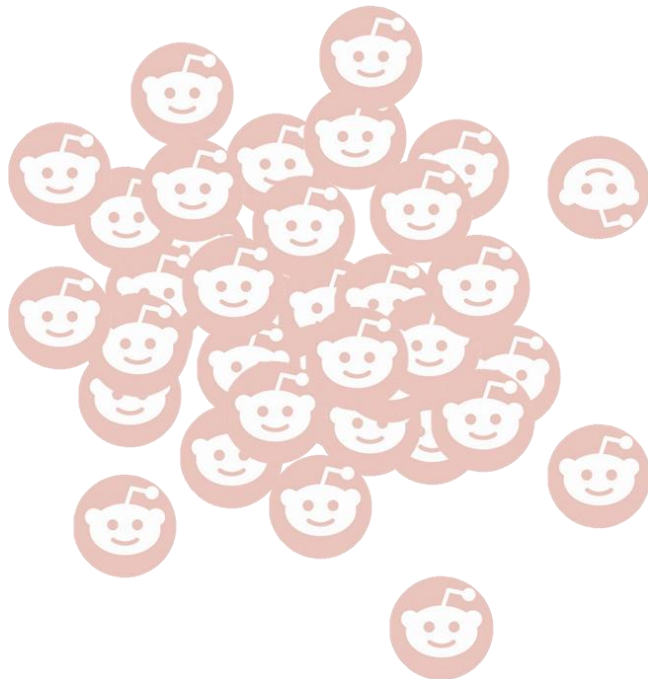
David Alexander | Marcel Zanetti | Pedro Teles
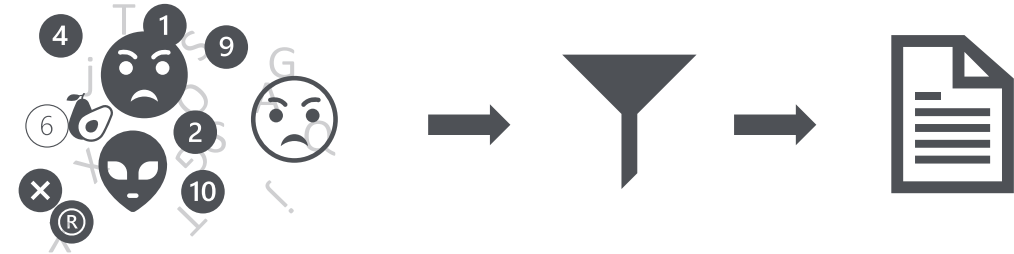
Data & Methodology

1

Collect data

- We collect data regarding **Reddit posts and comments** from **2021-01-01** to **2022-10-01**. In total, we have **169155 text documents**, as well as information about comments and upvotes.

2

Preprocess

First, we preprocess our text data, removing stopwords, punctuations, URLs, numbers, emojis, etc.

3

VADER analysis

After that, as in Camou (2022), we apply **VADER analysis** (Hutto & Gilbert (2014)) to assign an **intensity score** to every observation, indicating if it is positive or negative. As the documentation states, VADER (Valence Aware Dictionary and sEntiment Reasoner) is specifically **attuned** to sentiments expressed in **social media**.

Data & Methodology

1

Post score

- Following, for **every week** in our sample, we take the **weighted average** of the **VADER score** based on the **post score** (difference between upvotes and downvotes). We start the predictions on the 31st week, leaving 30 observations for training the first model.
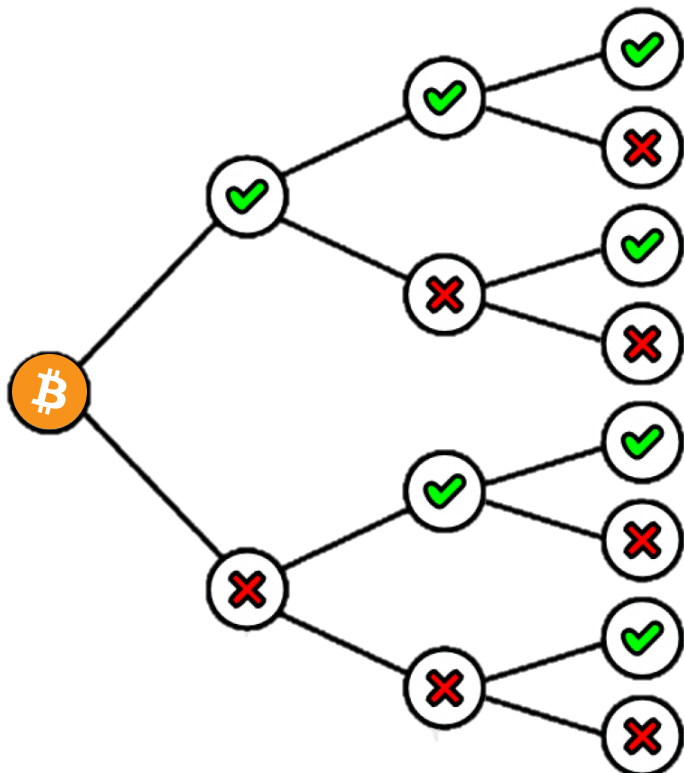
Data & Methodology

1

Hidden states

Thenceforth, we use the percentage difference in this weekly score to model the Bitcoin returns' **hidden states**. We specify that there are **two hidden states**: a **bullish** and a **bearish** one.

2

Rebalance

Finally, we go **long X%** in **Bitcoin** and **1 − X% long** in **cash** or the **risk-free** rate. In this case, X is the **probability of being in the bullish state** in the next week.
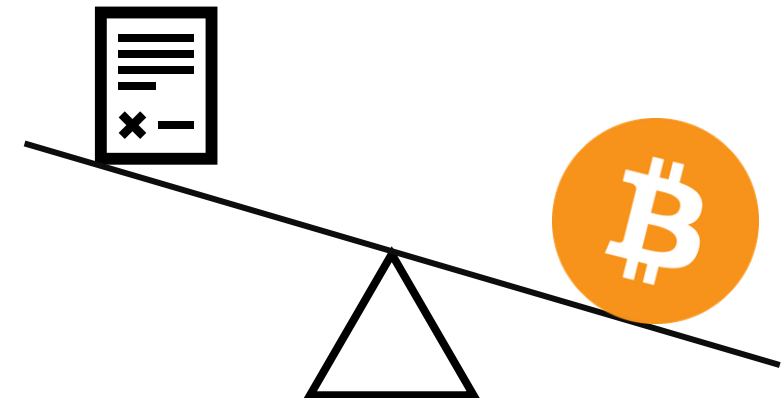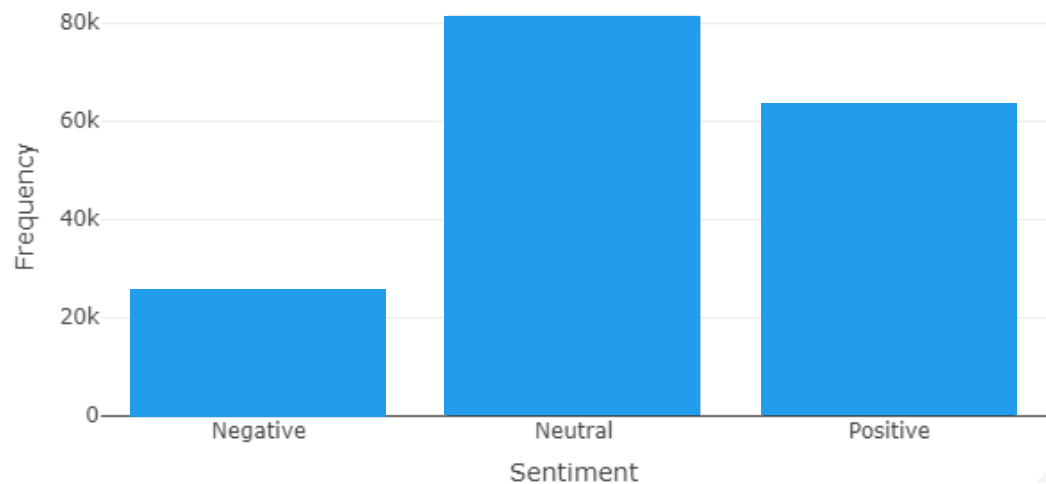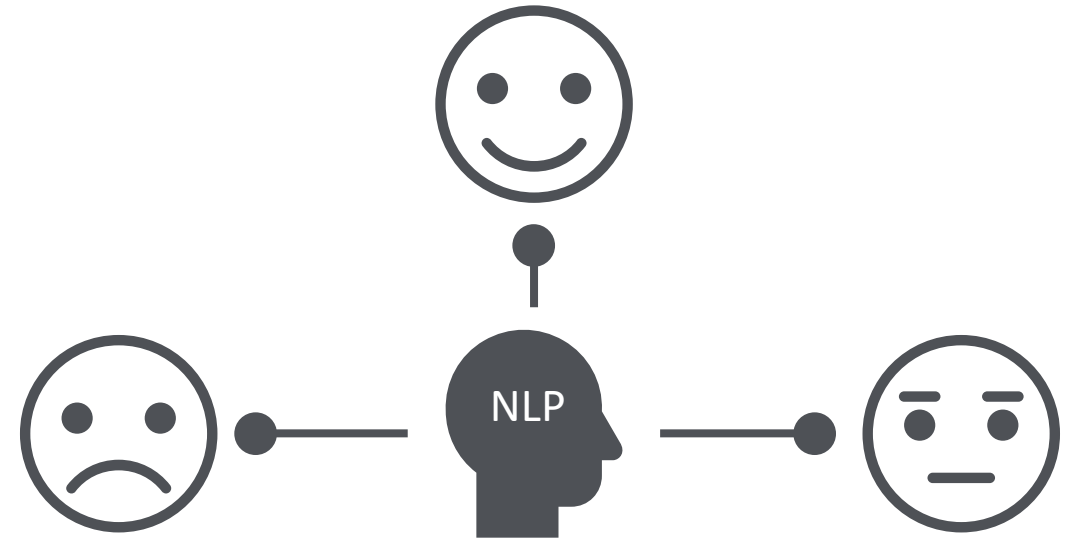
NLP Analysis

1 Posts Sentiment Frequency

2 NLP Classification

Neutral posts dominates. However, in this period, there were more positive than negative comments.

NLP Analysis

Reddit Posts WordCloud

Hidden Markov Model

David Alexander | Marcel Zanetti | Pedro Teles

Hidden Markov Model

Transition matrix

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9

Hidden Markov Model

1 Hidden Markov Model
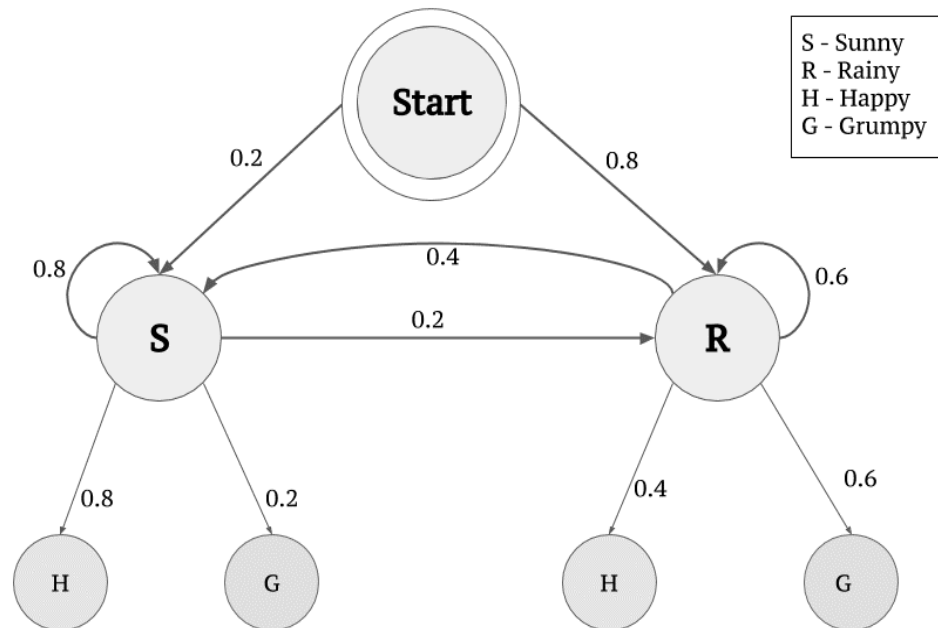
2 OBSERVABLE STATES REDDIT SENTIMENT

S - Sunny
R - Rainy
H - Happy
G - Grumpy

Hidden Markov Model
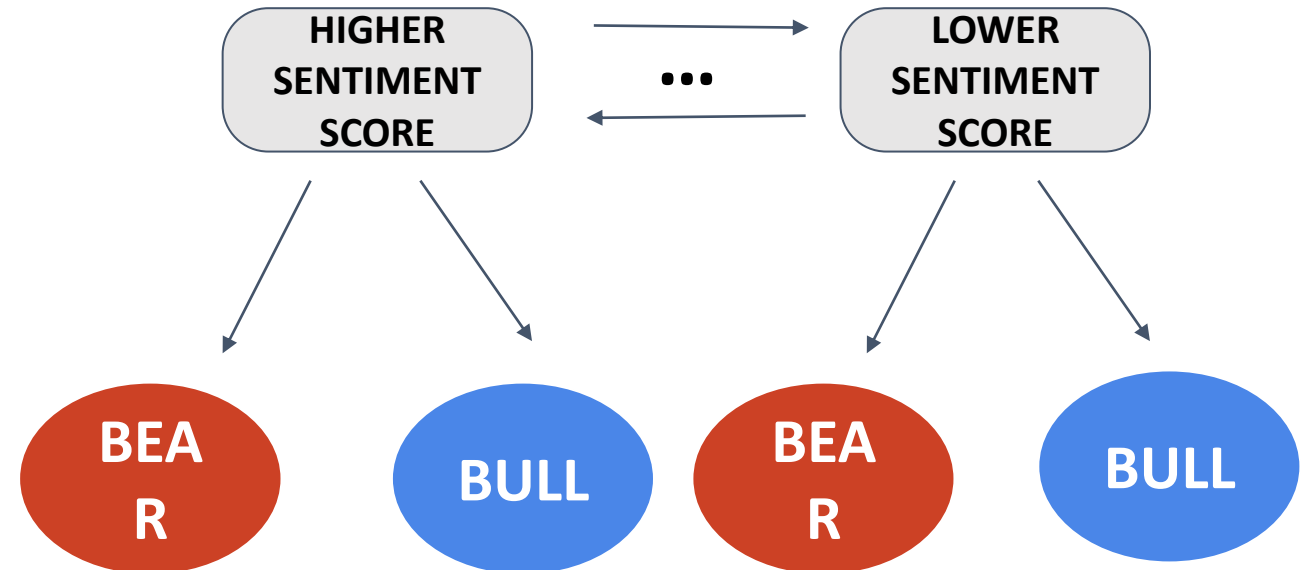
Bitcoin Return and HMM Hidden States

Bitcoin Return and HMM Hidden States
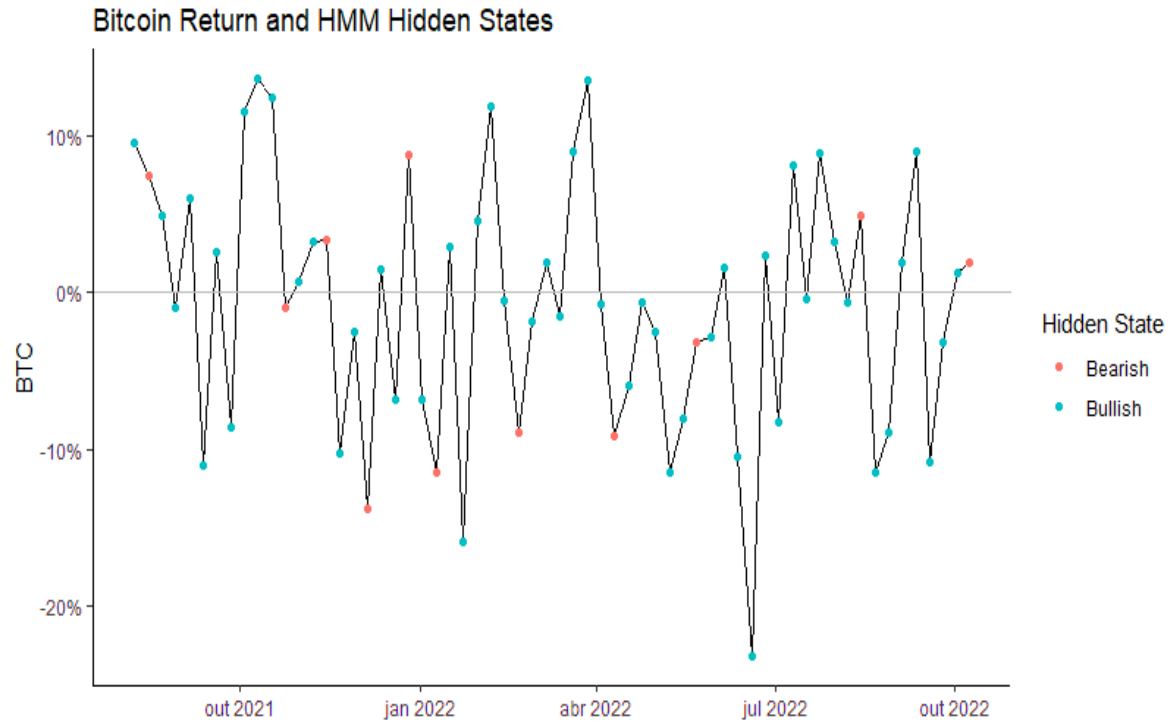
Results

David Alexander | Marcel Zanetti | Pedro Teles

FGV

U F *m* G

Results

Cumulative Returns

Results
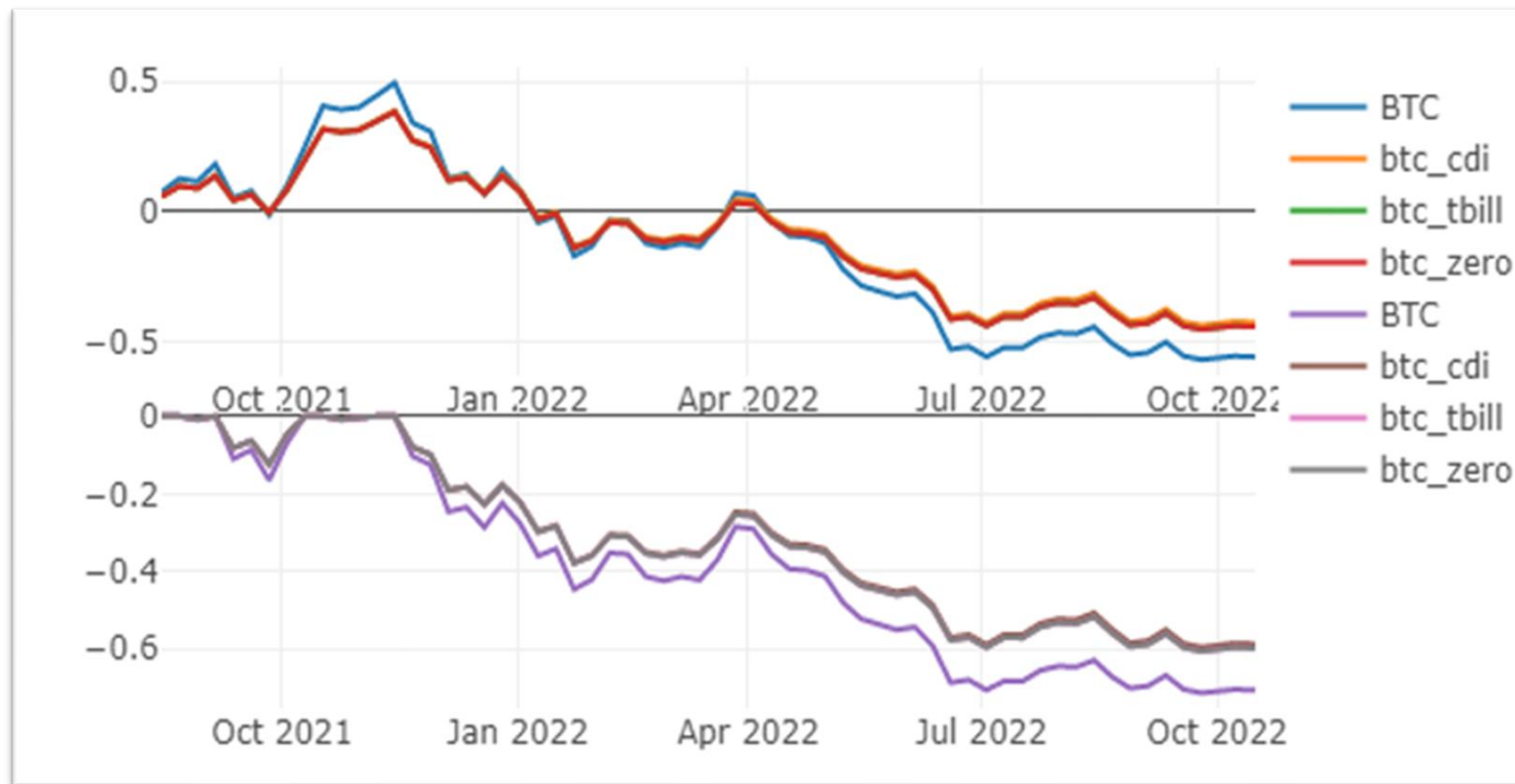
1

Table 1 – Portfolio Return Statistics

	BTC-CDI	BTC-Tbill	BTC-Cash	BTC	T-Bill
Annualized Return	-89,60%	-90,80%	-90,80%	-96,40%	2,20%
Annualized Volatility	93,90%	93,90%	93,90%	125,30%	
Modified Sharpe Ratio	-0,84	-0,85	-0,85	-1,21	
Max. Drawdown	59,50%	60,50%	60,50%	71,30%	
CVaR	-13,10%	-13,20%	-13,20%	-17,70%	
Skewness	-0,25	-0,25	-0,25	-0,24	
Kurtosis	- 0,32	-0,32	- 0,32	-0,23	

2

- The difference between the BTC-CDI and the BTC Modified Sharpe Ratio is highly significative: **z-stat = 10,60** (Jobson & Korkie (1981) & Memmel (2003))

Conclusion

David Alexander | Marcel Zanetti | Pedro Teles

FGV

U F M G

Conclusion

- Since the scrapping process (API limits) and the NLP analysis (preprocessing) are slow, we had to limit our sample to just under two years. On account of this fact, the inferences are limited. Despite this warning, using sentiment analysis and the HMM, we were able to form better portfolios in comparison to the plain Bitcoin one.

- Interested researchers can improve the model by adding more data (longer time horizon) and by exploring different structures. The work could also be expanded by adding other cryptocurrencies and getting data from other social platforms, like Twitter.

Conclusion

1

Future Developments

- For example, we could construct Long & Short portfolios based on the Markov output that goes long (short) cryptocurrencies with a higher (lower) probability of being bullish in the next period.

- In another example, Lewin and Campani (2020) consider four different hidden states to model the equity market. In addition, the Reddit data could be grouped in days or months, instead of weeks. Finally, one could attempt to use other information from the Markov Model to further improve the results.

References

- CAMOU, L. A. L. Reddit as a prediction tool for crypto-assets. Brazilian Review of Finance, v. 20, n. 1, 2022.

- POYSER, O. Herding behavior in cryptocurrency markets. arXiv preprint arXiv:1806.11348, 2018.

- BAUM, L. E.; PETRIE, T. Statistical inference for probabilistic functions of finite state Markov chains. The annals of mathematical statistics, JSTOR, v. 37, n. 6, p. 1554–1563, 1966.

- Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model forsentiment analysis of social media text,Proceedings of the InternationalAAAI Conference on Web and Social Media, Vol. 8.

- ISRAELSEN, C. L. et al. A refinement to the sharpe ratio and information ratio. Journal of Asset Management, v. 5, n. 6, p. 423–427, 2005.

- M. Lewin and C. H. Campani. Gestão de carteiras sob múltiplos regimes: Estratégias que performam acima do mercado. Revista de Administração Contemporânea, 24:300–316, 2020.

- Jobson, J. Dave, and Bob M. Korkie. "Performance hypothesis testing with the Sharpe and Treynor measures." Journal of Finance (1981): 889-908.

- Memmel, Christoph. "Performance hypothesis testing with the Sharpe ratio." Available at SSRN 412588 (2003)

Confira nosso trabalho no Github

SCAN ME

Thank you!

David Alexander | Marcel Zanetti | Pedro Teles

FGV

U F M G