
# Bitcoin and News Around the World in Twenty-Six Languages

Lucia Alessi, Eric Ghysels, Marco Petracco, Zhe Wang¶

marco.Petracco@ec.Europa.eu

**"Big Data and Machine Learning in Finance"**

*Politecnico di Milano, 11 June 2021*

Joint
Research
Centre

# Legal disclaimer

“The content of this presentation represents the views of the authors only and their sole responsibility; it cannot be considered to reflect the views of the European Commission and other institutions the authors work for. The European Commission and other institutions the authors work for do not accept any responsibility for use that may made of the information it contains”

# Contents

# Problem definition

- Cryptocurrencies are famous for their high volatility
- Is volatility driven by fundamentals or by other factors
- Price discovery “noisy” for cryptos wrt other asset classes
- → **What is the role of sentiment in news for cryptos pricing**

# The data

- Time span: April 16, 2014 to August 31, 2020
- High frequency data on BTC quotes from Kaiko
- Sentiment scoring of news extracted from a large variety of sources and languages, from EC's Europe Media Monitor

# The data: Price data (1)

- 6 exchanges: three in Asia (OkCoin in China, Bitfinex in China (Hong Kong), and Quoine in Japan), one in Europe (Bitstamp in Luxembourg), and two in US (Coinbase and Kraken).
- We choose these exchanges because they have long trading history to match with our EMM sample period, and high enough trading volume to alleviate liquidity issues.
- **Returns**
- **Netbuy**

# The data: Europe Media Monitor (EMM) (1)

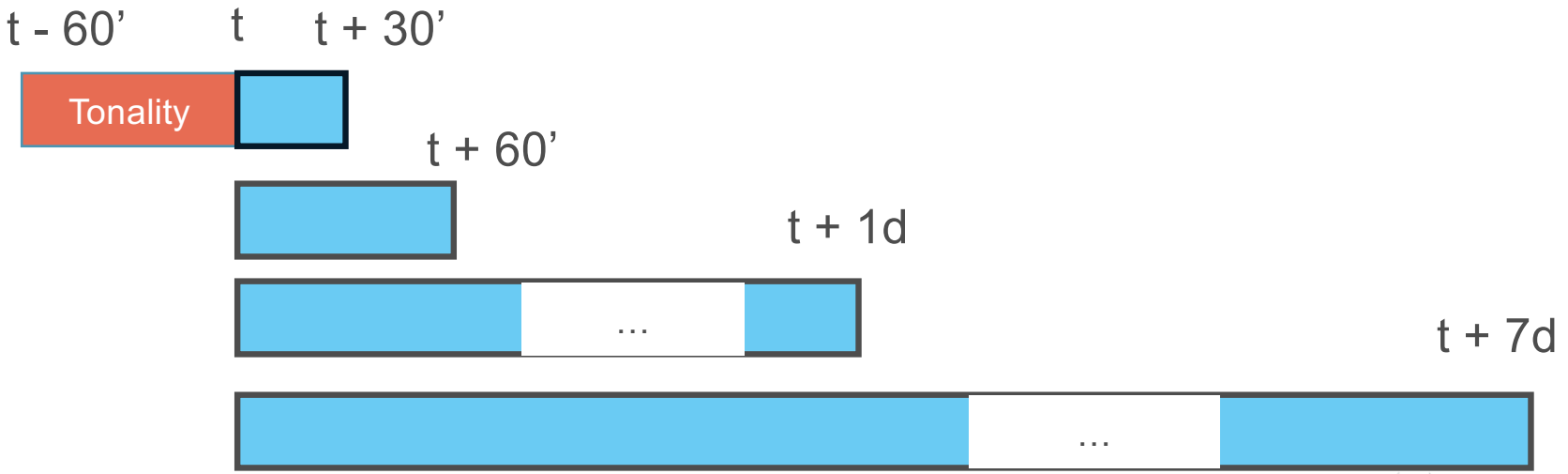
- EMM newsbrief is an automatic system that collects and analyses news media.
- Monitors a selected list of sources for news items, in up to 72 languages.
- Generates structured (meta) data (entities, events, categories …)
- Includes three methods for sentiment, emotion, tonality
- “JRC tonality” algorithm: +/- 1 for slightly positive/negative terms, +/- 4 for strongly positive/negative terms. Normalized by word count.

# The data: Europe Media Monitor (EMM) (2)

- Articles belonging to “Bitcoin” category (inclusion based on appearance of keywords and criteria determined by subject matter experts).
- List of sources include most major online newspapers, and a set of specialised news outlets
- 403,112 articles in 26 languages (184,354 of which in English).
- **Time-stamp** (retrieval), **language**, “JRC **tonality**”
- Aggregate in 1-hour intervals, for English and all other languages Retain all intervals with articles both in English and other languages
- About 30 thousand 1-hr time intervals (67% of available)

# Methods:

- Divide every day in 24 1-hour time intervals t. (t = 0 : 23)
- Calculate tonality over period (t-1hr:t)
- Calculate returns and netbuy of periods (t:t+k), where k = 30’, 60’, 1d, 7d






European Commission

# Methods:

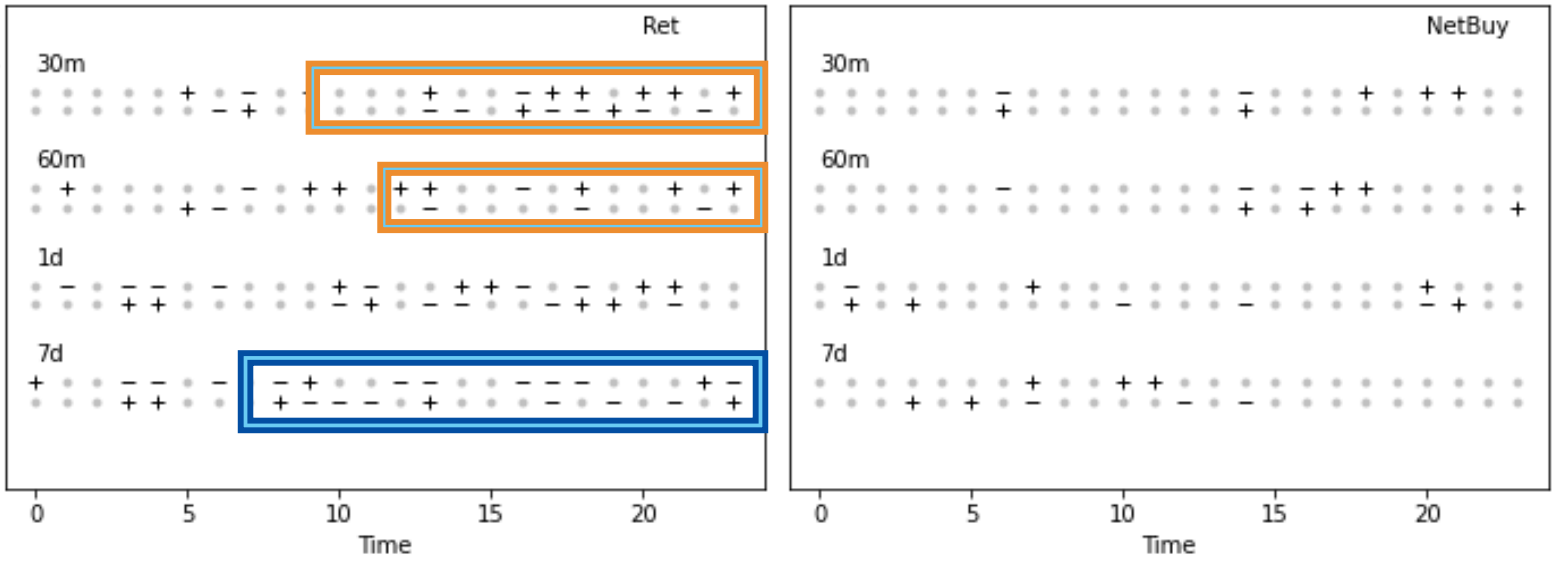
$$\text{Sign}(\text{Ret}_{t+k}) = \text{Logistic}(\alpha + \beta_1 \text{TonalityEN}_t + \beta_2 \text{TonalityOT}_t + \gamma \text{Controls}_t + \epsilon_{t+1})$$

$$\text{NetBuy}_{t+k} = \alpha + \beta_1 \text{TonalityEN}_t + \beta_2 \text{TonalityOT}_t + \gamma \text{Controls}_t + \epsilon_{t+1}$$

- Ret(t+k) and NetBuy (t+k) with k = 30 mins, 60 mins, 1 day, and 7 days, starting from the end of the time interval t. 24, 1-hour time intervals.
- TonalityEN(t) is the average tonality of all English news over time period t. TonalityOT(t) is the average of tonality of news published in other languages over the same time interval.
- Controls include a macroeconomic sentiment index constructed by the Federal Reserve Bank of San Francisco.
- For each time block, dummies for “excess difference” between EN and OT tonality.
- Also estimated with fixed effects by Exchange, and with different slopes by exchange
- Mainly interested in the **SIGN** of significant coefficients for news sentiment in EN and OT

# Results: returns and netbuy (pooled and FE)

$\beta_1$ (EN)
$\beta_2$ (OT)

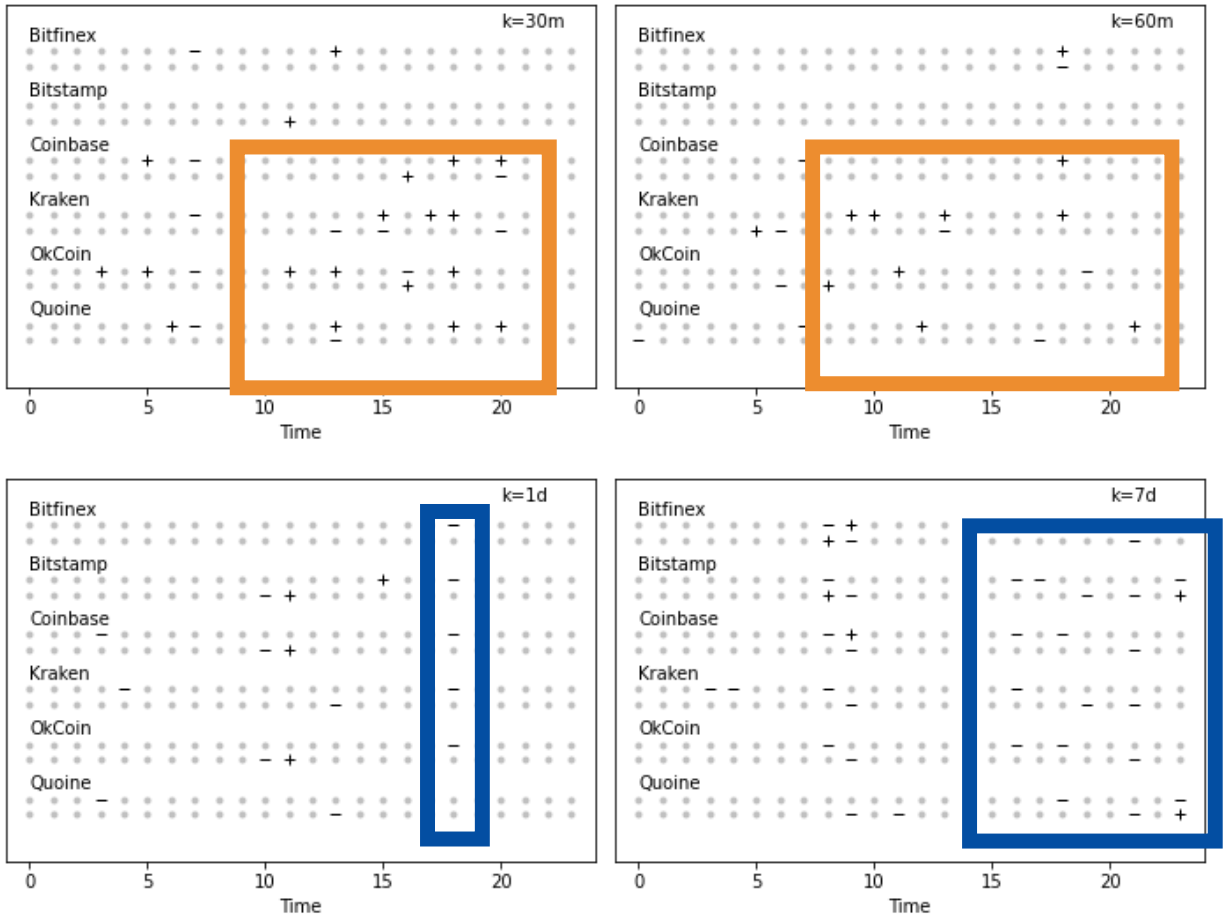



- 24 time slots; **+**, significant, positive; **-** , significant, negative; top line (EN), bottom line (OT)
- On average, returns positively correlated with news tonality at short time intervals, and negatively correlated at the longest time interval
- Netbuy volume more muted relation.



European Commission

# Results: returns (by exchange)

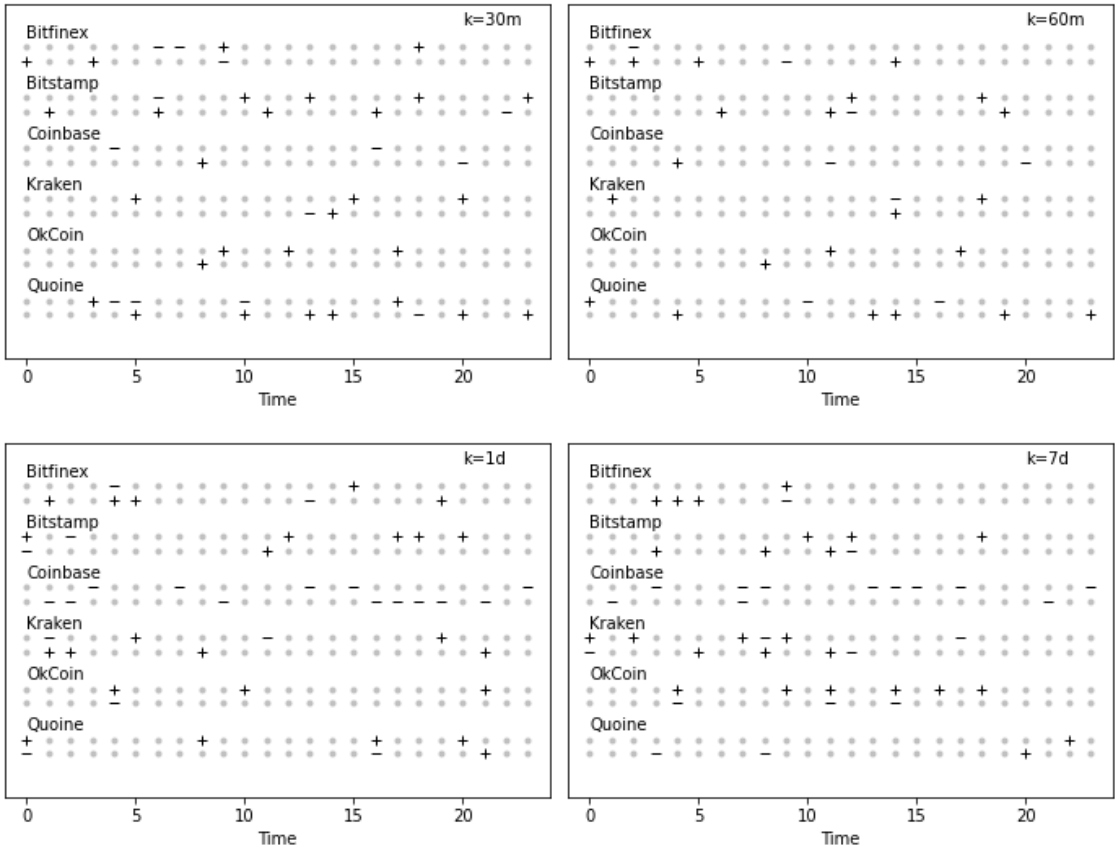



- On average, returns positively correlated with news tonality at short time intervals, and negatively correlated at the longest time interval
- For bitfinex and Bitstamp short term relationship disappears



European Commission

# Results: netbuy (by exchange)




- More significant coefficients when exchange specific intercepts are allowed
- Some exchanges seem to have a prevalence of one sign … artefact or …



European Commission

# Conclusions:

- Positive correlation between tonality of news articles in English and return on BTC over the next 30' and 60'
- agree with previous research pointing to a role for sentiment in determining returns in crypto-currencies, and with the fact that news sentiment could affect short term and intraday returns
- evidence of a reversal of these effects over longer time horizons, especially 7 days . This is suggestive of the fact that these dynamics could be tied to extensive noise trading on BTC markets.
- further supported by the analysis of results at individual exchange level, where the effects seems to disappear on Bitfinex, which is reportedly a more efficient BTC exchange.
- NetBuy dynamics more muted at aggregate level, but different exchanges seem to respond more to news.

# Further research:

- Relationships seems to be inverted for non-English language news. This result could point to a different use of news by noise traders in different countries, or to the existence of dynamics in news tonalities across different languages and will need to be investigated further.

- Netbuy dynamics might be compatible with interest and noise driven dynamics, and/or with different roles for different exchanges, or BTC trading for other purposes

- Subject of further investigation to understand if more extreme cases of market “one sidedness” could be the object of prediction.

- Introduce social media sentiment (e.g. Reddit, Twitter) …

# Thank you





European Commission

# Keep in touch

EU Science Hub: ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub – Joint Research Centre

EU Science, Research and Innovation

EU Science Hub



European Commission

# The data: Europe Media Monitor (EMM) (3)

**Table 1: Summary Statistics of Tonality**

| Hourly Block | TonalityEN | TonalityOT | Hourly Block | TonalityEN | TonalityOT |
|---|---|---|---|---|---|
| 0 | -2.638 | -0.119 | 12 | -1.186 | 0.357 |
| 1 | -3.408 | 0.364 | 13 | -1.182 | 0.385 |
| 2 | -3.259 | 0.022 | 14 | -1.592 | 0.337 |
| 3 | -3.226 | 0.218 | 15 | -1.945 | 0.251 |
| 4 | -2.596 | 0.232 | 16 | -1.661 | 0.257 |
| 5 | -2.023 | 0.511 | 17 | -2.152 | 0.294 |
| 6 | -2.246 | 0.495 | 18 | -2.285 | 0.026 |
| 7 | -2.274 | 0.432 | 19 | -2.134 | -0.104 |
| 8 | -1.839 | 0.445 | 20 | -2.563 | 0.275 |
| 9 | -1.626 | 0.458 | 21 | -2.451 | -0.061 |
| 10 | -1.797 | 0.476 | 22 | -2.868 | 0.033 |
| 11 | -1.860 | 0.866 | 23 | -3.029 | 0.139 |
| | | | | | |
| | | | All | -2.164 | 0.292 |

Notes: Entries to the table are the average tonality of all hourly time blocks.

# The data: Price data (2)

**Table 2: Summary Statistics of Return and NetBuy**

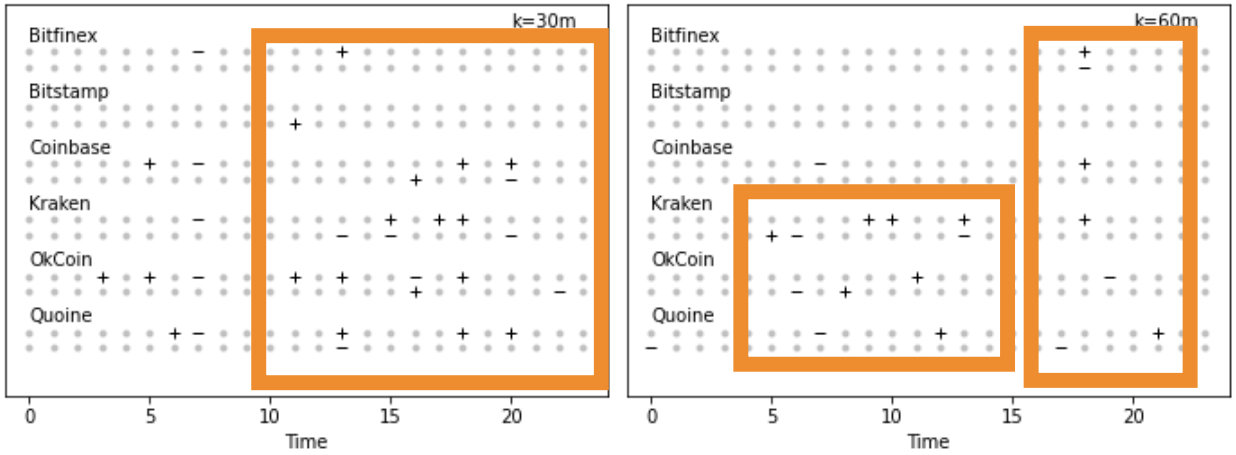
| Ret | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 30 mins | | 60 mins | | 1 day | | 7 days | |
| | N | Mean | N | Mean | N | Mean | N | Mean |
| Bitfinex | 30148 | 2.68E-05 | 30146 | 7.84E-05 | 30098 | 1.75E-03 | 29930 | 0.016 |
| Bitstamp | 30103 | 2.29E-05 | 30100 | 8.15E-05 | 30036 | 1.85E-03 | 29802 | 0.017 |
| Coinbase | 28479 | 9.80E-06 | 28463 | 6.81E-05 | 28360 | 2.22E-03 | 28237 | 0.018 |
| Kraken | 29439 | 3.12E-05 | 29369 | 5.56E-05 | 28989 | 2.02E-03 | 28795 | 0.017 |
| OkCoin | 26106 | 7.20E-05 | 26091 | 6.93E-05 | 25943 | 1.52E-03 | 25621 | 0.014 |
| Quoine | 27257 | -8.95E-05 | 27167 | 6.27E-05 | 26681 | 2.39E-03 | 26519 | 0.019 |

| NetBuy | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 30 mins | | 60 mins | | 1 day | | 7 days | |
| | N | Mean | N | Mean | N | Mean | N | Mean |
| Bitfinex | 29934 | -0.004 | 30061 | -0.009 | 30150 | -0.020 | 30140 | -0.023 |
| Bitstamp | 25893 | 0.074 | 25897 | 0.073 | 25976 | 0.054 | 26090 | 0.048 |
| Coinbase | 28363 | 0.084 | 28388 | 0.078 | 28638 | 0.068 | 28604 | 0.063 |
| Kraken | 27407 | 0.009 | 28041 | 0.003 | 30143 | -0.011 | 30153 | -0.017 |
| OkCoin | 25120 | -0.055 | 25615 | -0.056 | 26240 | -0.068 | 26439 | -0.061 |
| Quoine | 24189 | -0.028 | 25379 | -0.034 | 28347 | -0.019 | 28977 | -0.019 |

Notes: Entries are the statistics of $Ret_{t+k}$ and $NetBuy_{t+k}$ for $k$ =30 mins, 60 mins, 1 day and 7 days.

# Results: returns (by exchange – 30’ and 60’)

**Table 5: Tonality Predicts Homogeneous Future Returns at Different Exchanges.**

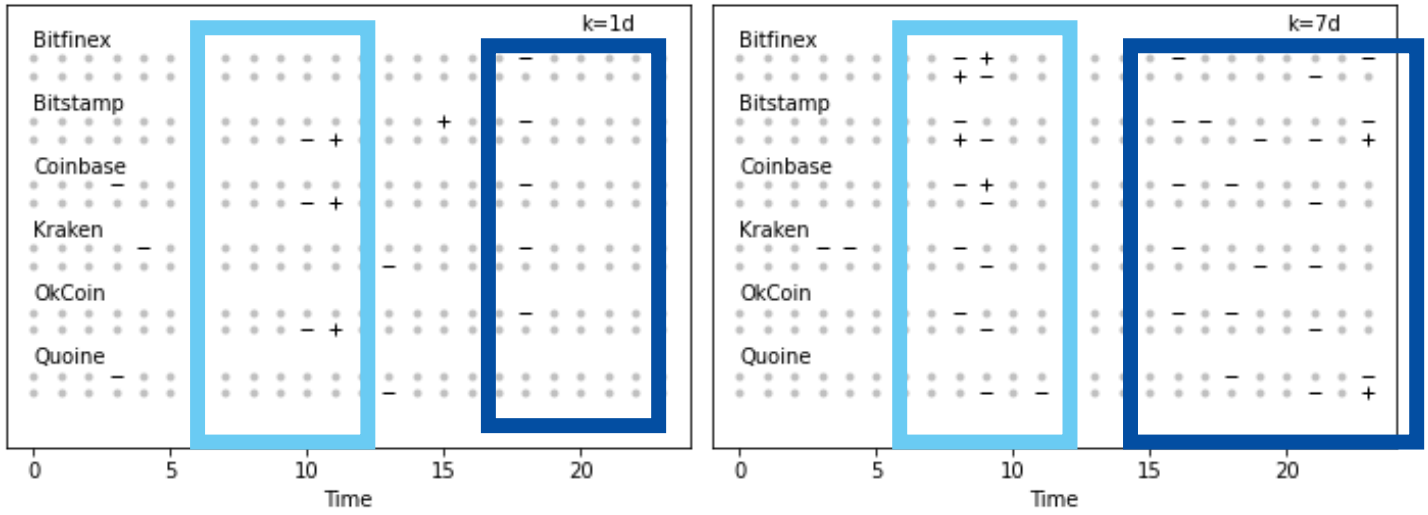



- On average, returns positively correlated with news tonality at short time intervals, and negatively correlated at the longest time interval
- Netbuy volume more muted relation.



European Commission

# Results: returns (by exchange – 1d and 7d)




- On average, returns positively correlated with news tonality at short time intervals, and negatively correlated at the longest time interval
- Netbuy volume more muted relation.



European Commission