Data Integration and Graphical Models for Cryptocurrencies

Luciana Dalla Valle

University of Plymouth

September 21, 2022

Joint work with Claudia Tarantola (University of Pavia)

Aims

- Our project aims at exploiting different sources of high quality information, yielding more accurate and timely predictions of financial prices than those produced by existing methodologies.

Aims

- Our project aims at exploiting different sources of high quality information, yielding more accurate and timely predictions of financial prices than those produced by existing methodologies.
- We leverage **crypto-asset prices** as well as related social media information with a novel **data integration** methodology based on graphical and dependence models producing accurate predictions and assessments of financial risks.

Motivations

- Statistical models for financial asset pricing and forecasting might generate incomplete and **not accurate enough results**, if built on a single source of information.

Motivations

- Statistical models for financial asset pricing and forecasting might generate incomplete and **not accurate enough results**, if built on a single source of information.
- On the one hand, due to the recent events affecting the world population and politics (i.e. the Covid-19 pandemic and lockdown, the war in Ukraine, etc.) market unpredictability is making financial forecasts based on historical asset prices **less reliable**.

Motivations

- Statistical models for financial asset pricing and forecasting might generate incomplete and **not accurate enough results**, if built on a single source of information.
- On the one hand, due to the recent events affecting the world population and politics (i.e. the Covid-19 pandemic and lockdown, the war in Ukraine, etc.) market unpredictability is making financial forecasts based on historical asset prices **less reliable**.
- On the other hand, social media data are generated by users on a voluntary basis and may **not capture** information about the **entire population**.

Methodology

- Our approach is based on graphical and dependence models, which allow us to **integrate** asset prices with textual information gathered from social media platforms.

Methodology

- Our approach is based on graphical and dependence models, which allow us to **integrate** asset prices with textual information gathered from social media platforms.
- In contrast to other “black-box” approaches, graphical and dependence models allow a **transparent and immediate interpretation** of results.

Graphical models

- Graphical models are probabilistic tools expressing the **conditional dependence structure** between random variables.

Graphical models

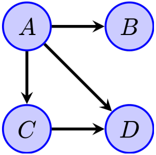
- Graphical models are probabilistic tools expressing the **conditional dependence structure** between random variables.
- Graphs are an intuitive way of **representing and visualising** the relationships between many variables.

Graphical models

- Graphical models are probabilistic tools expressing the **conditional dependence structure** between random variables.
- Graphs are an intuitive way of **representing and visualising** the relationships between many variables.
- A graph allows us to abstract out the **conditional independence relationships** between the variables from the details of their parametric forms.

Graphical models

- Graphical models are probabilistic tools expressing the **conditional dependence structure** between random variables.
- Graphs are an intuitive way of **representing and visualising** the relationships between many variables.
- A graph allows us to abstract out the **conditional independence relationships** between the variables from the details of their parametric forms.
- Graphical models allow us to define general message-passing **algorithms** that implement probabilistic inference efficiently (Maathuis, 2018).



Dependence models

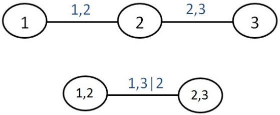
- Dependence models (specifically, vine copulas) are mathematical tools that allow the separation between the marginal distributions and their dependence structure and, in some particular cases, they can be represented via graphical models.

Dependence models

- Dependence models (specifically, vine copulas) are mathematical tools that allow the separation between the marginal distributions and their dependence structure and, in some particular cases, they can be represented via graphical models.
- Vine copulas use bivariate copulas as building blocks to define highly flexible multivariate distributions that are represented via graphical models as nested set of connected trees.

Dependence models

- Dependence models (specifically, vine copulas) are mathematical tools that allow the separation between the marginal distributions and their dependence structure and, in some particular cases, they can be represented via graphical models.
- Vine copulas use bivariate copulas as building blocks to define highly flexible multivariate distributions that are represented via graphical models as nested set of connected trees.
- The flexibility of vine copulas allows us to overcome many of the issues associated with commonly used distributions by allowing different complex **asymmetric dependencies** and tail behaviours to be modelled (Czado, 2019).



Data Integration Methodology

- We apply the methodology to **integrate crypto-asset prices** and social media data, extracted from platforms such as Twitter and Google Trends, producing precise predictions and measuring financial risk accurately.

Data Integration Methodology

- We apply the methodology to **integrate crypto-asset prices** and social media data, extracted from platforms such as Twitter and Google Trends, producing precise predictions and measuring financial risk accurately.
- Bitcoin data **time horizon**: February–June 2021

Data Integration Methodology

- We apply the methodology to **integrate crypto-asset prices** and social media data, extracted from platforms such as Twitter and Google Trends, producing precise predictions and measuring financial risk accurately.
- Bitcoin data **time horizon**: February–June 2021
- We implement time series to model the data dynamics of **cryptocurrencies** and **online** gathered **information**.

Data Integration Methodology

- We apply the methodology to **integrate crypto-asset prices** and social media data, extracted from platforms such as Twitter and Google Trends, producing precise predictions and measuring financial risk accurately.
- Bitcoin data **time horizon**: February–June 2021
- We implement time series to model the data dynamics of **cryptocurrencies** and **online** gathered **information**.
- Subsequently, we use graphical and dependence models, such as vine copulas, to capture the **dependence structure** between variables.

Preliminary results: Google trends

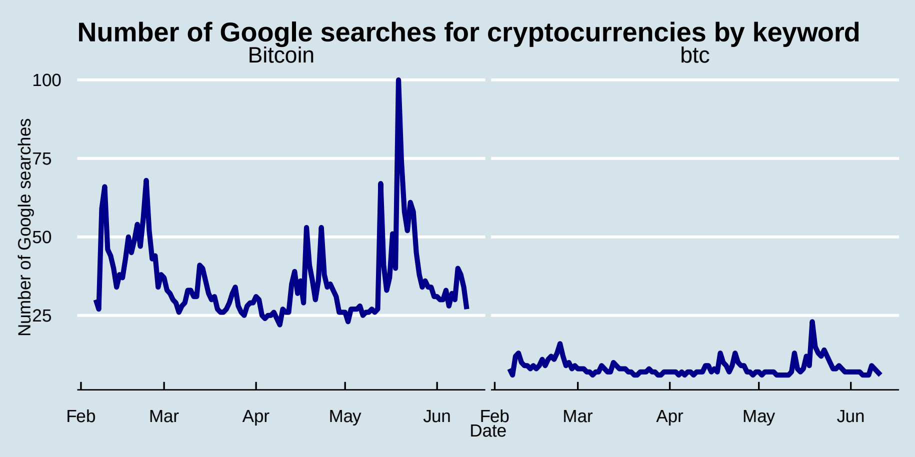


Figure: Number of Google Trends searches by keyword ("Bitcoin" on the left and "btc" on the right).

Preliminary results: Google trends

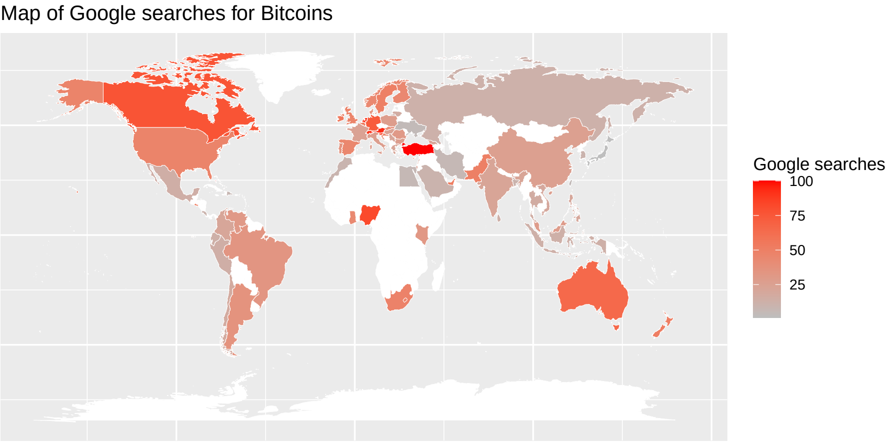
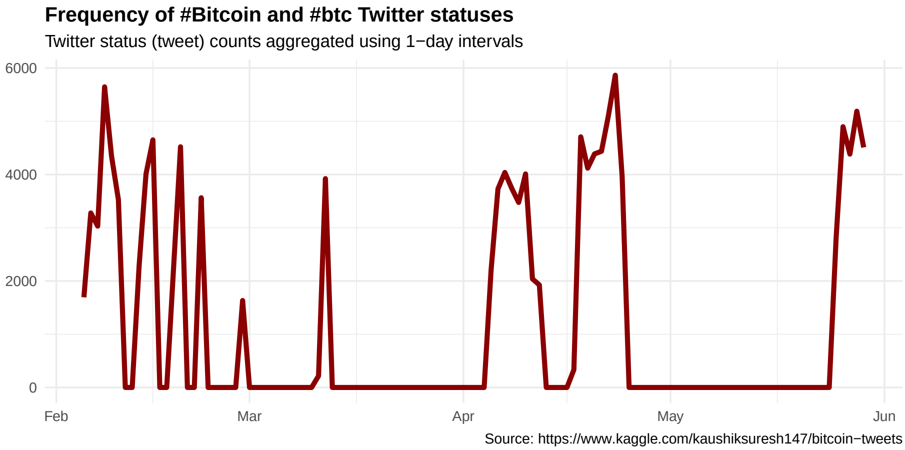


Figure: Map of the number of Google Trends searches by country.

Preliminary results: Twitter



Source: https://www.kaggle.com/kaushiksuresh147/bitcoin−tweets

Figure: Number of Tweets containing the hashtags “Bitcoin” and “btc”.

Preliminary results: Twitter

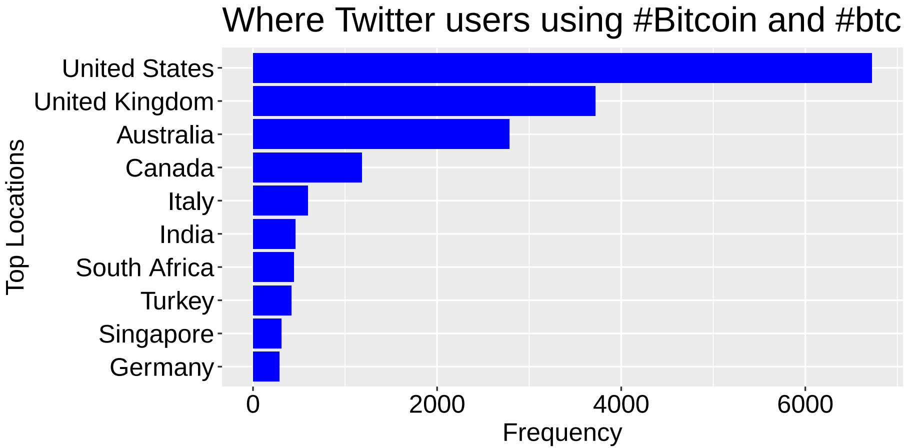
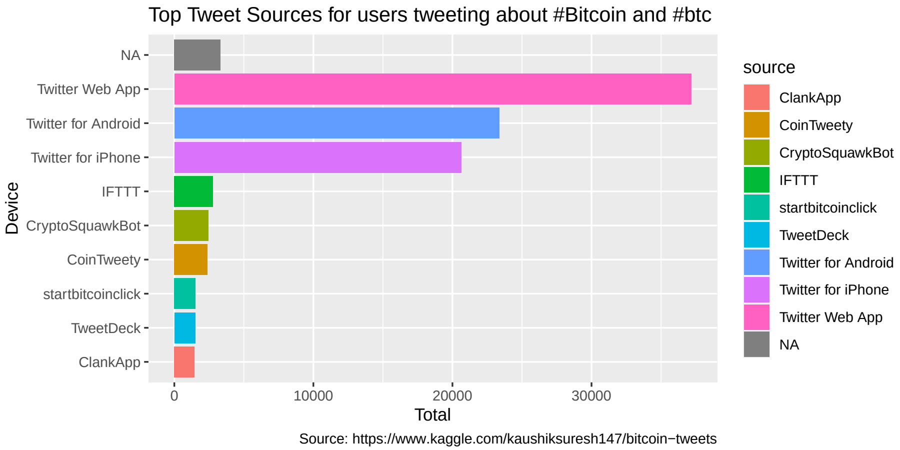


Figure: Top locations of tweets containing the hashtags "Bitcoin" and "btc".

Preliminary results: Twitter



Source: https://www.kaggle.com/kaushiksuresh147/bitcoin−tweets

Figure: Top tweet sources for users tweeting about “Bitcoin” and “btc”.

Preliminary results: Twitter

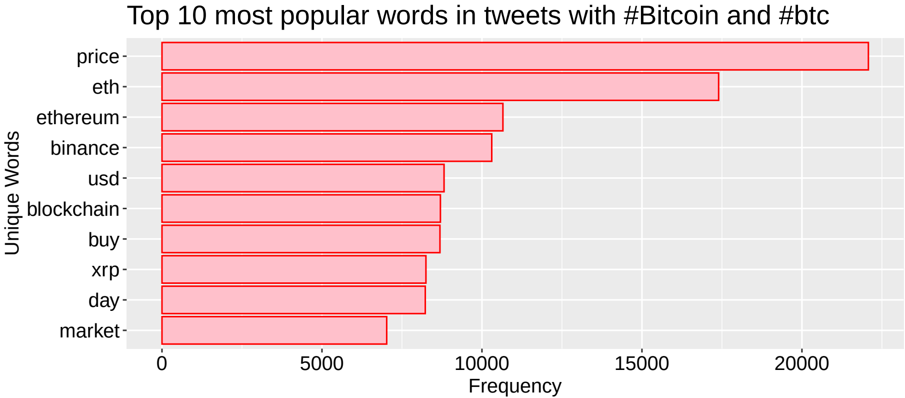


Figure: Top tweets containing “Bitcoin” and “btc”.

Preliminary results: Twitter

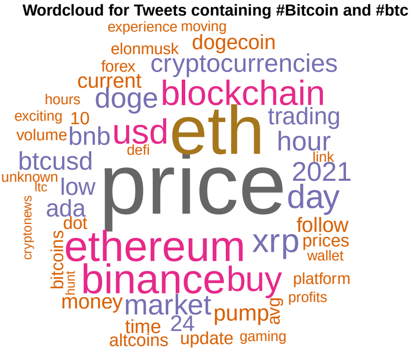


Figure: Wordcloud for tweets containing “Bitcoin” and “btc”.

Preliminary results: Twitter

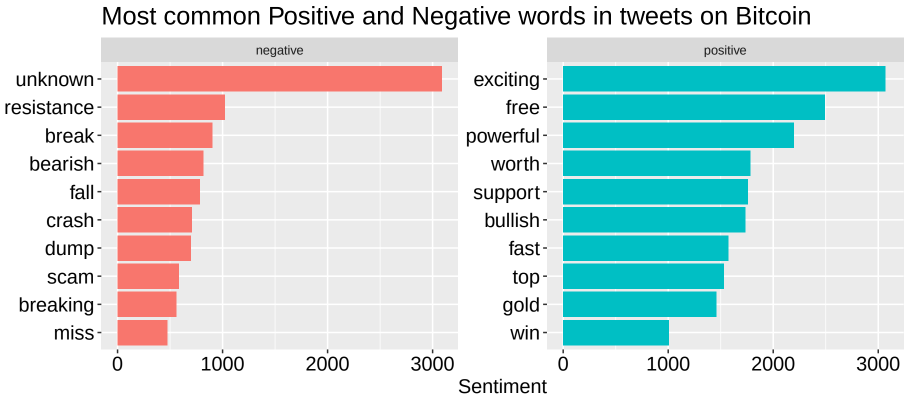


Figure: Most common positive and negative words in tweets containing the hashtags "Bitcoin" and "btc".

Preliminary results: Twitter

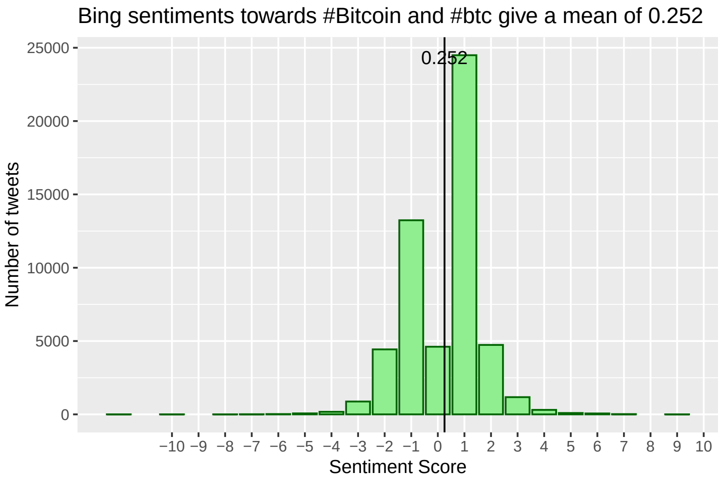
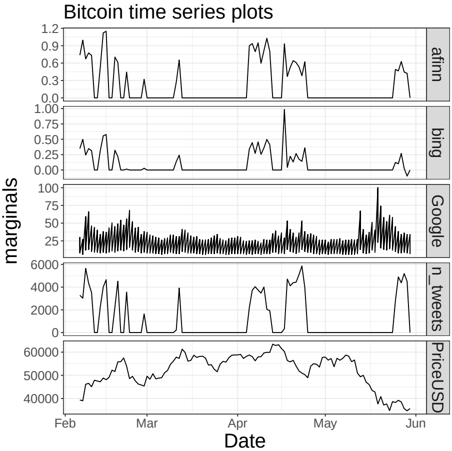


Figure: Histogram of Bing sentiment scores for tweets with hashtags "Bitcoin" and "btc".

Next steps

1. Time series analysis of crypto-asset prices and social media information
2. Data integration using graphical and dependence models
3. Calculate predictions based on the data integration model



References

Czado, C. (2019). *Analyzing dependent data with vine copulas*. Lecture Notes in Statistics, Springer, 222.

Maathuis, M., Drton, M., Lauritzen, S., & Wainwright, M. (Eds.). (2018). *Handbook of graphical models*. CRC Press.