ORIGINAL RESEARCH

Economic keywords in political communications and financial markets

Valerio Ficcadenti[1] · Roy Cerqueti[2,3]

Abstract

This paper investigates the interplay between the use of economically significant keywords in political speeches—indicative of their economic content—and observed trends in financial markets. To this end, we implement a statistical/bibliometric study on the speeches stated by the US Presidents and the Standard and Poor's 500 index (S&P 500). The corpus covers 1951-2017, including 376 discourses. The S&P 500 daily returns, prices, and volumes are considered in the same period. Our analysis utilises Kendall's τ correlation to assess the relationship between S&P 500 variables and the frequency of economic terms in the speeches. Additionally, four distance measures are applied to evaluate the similarities among these time series. Lastly, we compare Shannon entropies computed for each variable. The values of the index are observed the day before, the same day, and the day after the Presidents conducted their speeches. The joint evaluation of the results shows that the speeches' economic content and the S&P 500 volume have the most meaningful relationship. In particular, the volume shows some changes interrelated to the speeches before and after the days of the talks. The price has similar behaviour but with a lighter magnitude than volume. The daily returns are poorly responsive to the speeches' economic content, but entropy suggests that they present some distributional similarities.

Keywords Statistical analysis · Economic keywords · Text mining · Political speeches · Financial markets

✉ Roy Cerqueti
roy.cerqueti@uniroma1.it

Valerio Ficcadenti
ficcadv2@lsbu.ac.uk

[1] Business School, London South Bank University, 103 Borough Rd, SE1 0AA London, UK

[2] Department of Social and Economic Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185 Rome, Italy

[3] GRANEM. SFR CONFLUENCES, University of Angers, 49000 Angers, France

🌥 Springer

1 Introduction

Academic literature highlights the existence of cycles in economic thought linked with the ebbs and flows of economic activities (see, e.g. Kufenko and Geiger 2016). The referenced paper undertakes a bibliometric analysis of the scientific literature on business cycles, delving into the association between economic-related keywords and the evolution of economic activities.

The present paper contributes to this discussion through a statistical analysis of a comprehensive collection of official political speeches, focussing on economically significant keywords. Specifically, we examine whether and how the Standard and Poor's 500 (S&P 500 henceforth) interrelates to the economic terminology used by US Presidents in their public addresses. We operate under the premise that words used on official occasions carry distinct significance and weight.

Our study's foundation lies in recognising the US President as a key global influencer, whose carefully crafted speeches significantly impact the socio-economic milieu. For instance, Lee and Myers (2004) discusses the influence of political viewpoints on Enterprise Resource Planning (ERP) changes. The authors observe: *"In the political school of thought within strategic management, strategy formation and implementation are seen as being shaped by power and politics"*. Conversely, we highlight the effect of factual developments and contingencies on the choice of themes in US Presidents' speeches, suggesting a bidirectional interaction.

Our research utilises two datasets: the first comprises textual data, identical to that in Cinelli et al. (2021), Ficcadenti et al. (2019), Ficcadenti et al. (2020), encompassing US Presidential speeches from George Washington's tenure in April 1789 to Donald Trump's in February 2017. The second dataset includes daily prices, volumes, and returns of the S&P 500 index. The former dataset originates from the Miller Center,[1] a renowned Political Research Institution affiliated with the University of Virginia. The latter dataset is sourced from the freely accessible "Yahoo!Finance" website[2] (the rationale for selecting these sources is detailed in Sect. 3). The analysis period for the S&P 500 data spans from April 10, 1951, to March 1, 2017, aligning with the speech dataset dates.

Our exploration delves into the interplay between textual content (speech transcripts) and financial markets, offering a novel perspective on an this researched area; we signpost the reader to Gupta et al. (2020) and Fisher et al. (2016)—where extensive literature reviews are presented. Specific examples to contextualise our motives further are Kalyani et al. (2016), where sentiment classifications of Apple's financial news predict stock trends, and Renault (2020), which examines the correlation between investor mood and stock returns through sentiment analysis of messages on key financial topics. The use of named-entity recognition in finance is exemplified in Sagheer and Sukkar (2019), where texts from Arabic media outlets are analysed using the Al Khalil lexicon in the context of oil production. Furthermore, Qiu et al. (2013) conducted Chinese document modelling for annual report disclosure quality assessment. These studies predominantly leverage Machine Learning and AI tools, especially in the field of Natural Language Processing (NLP), which often rely on large tagged dictionaries. Research focusing explicitly on economic semantic tagged dictionaries includes Consoli et al. (2020), centred on sentiment analysis of economic texts, and Loughran and McDonald (2011) where 4000 words taken from 10Ks financial statement are classified in 7 sentiment categories. These studies underscore the widespread use of AI and Machine Learning, despite

---

[1] https://millercenter.org/

[2] https://it.finance.yahoo.com

their often limited explicability (Lipton, 2018). A key motivation for our research is the utilisation of a fully explicable approach to examine the connection between the economic content of US political speeches and the S&P 500, employing a predefined dictionary.[3] Additionally, the use of a politically focused corpus spanning [1951-2017] necessitates careful calibration of a pre-trained model due to changes in language over the years.

The scholarly debate in this realm has primarily focused on the interplay between Central Banks (through alternative data) and financial markets (refer to the discussion on the "evolution of information processing technology" in Kuroda 2017), with less attention to politicians' communications. For instance, Born et al. (2014) conducted a comprehensive textual analysis of communications from Central Bank governors across over 20 countries and their respective stock markets.

Studies on the impact of US politicians' speeches on financial markets present a scattered view; recent works like Marinč et al. (2021) have explored the role of the Presidents' (and candidates') Tweets from 2007 to 2020, and Ajjoub et al. (2020) focused on President Trump's Tweets. Thus, our second motivation is to examine the historical relationship between US Presidents' speeches and the US stock market.

To demonstrate the exploratory potential of our proposed method (see, Makri and Neely 2021; Saunders et al. 2009, where the exploratory research is defined and showcased), we treat the speeches as market news, employing a bag-of-words approach with an economic glossary sourced from Bishop (2009) and Wikipedia's economics glossary,[4] consistent with Cinelli et al. (2021). We then correlate the presence of economic terms in the speeches with S&P 500 variables one day before, on the same date, and one day after the speeches, allowing us to detect anticipatory or lagged effects of the discourse on the market, and vice versa. Our goal is not to quantify the effects of one variable on another or to establish the directionality of the relationship between the presence of economic words in US presidential speeches and S&P500 market dynamics. Rather, our aim is to identify the existence of a systematic relationship between these elements.

Methodologically, we examine the relationship between the economic content of the speeches and the S&P 500 index through comprehensive statistical analysis from two macroscopic perspectives. Firstly, we employ Kendall's τ correlation (introduced in Kendall 1938) to measure the ordinal association between the frequency of economic terms in the speeches and the financial variables of the index (for more on rank-rank correlation, see Melucci 2007; Gibbons 1993). In doing so, we assess the presence of a relation at a rank level among the occurrences of economic terms in the talks and the value of daily prices, returns and volumes. Secondly, we adopt a geometric and information-theoretic perspective, calculating max, min, Manhattan, and Euclidean distances between the time series of economic term frequencies and S&P 500 observations. These distances offer diverse insights into the relationship between the economic lexicon of the Presidents and the evolution of the S&P 500. Additionally, we compute Shannon entropy (see Shannon 1948) to gain insights into the distribution similarities of each series. This multifaceted approach enables a comprehensive understanding of the relationship between political economic messages and the S&P 500. The unique nature of our research subject further informs our methodology choice. The US presidential speeches in our dataset are distributed irregularly across a lengthy time span, encompassing various major global events from wars to financial crises (some changes on Presidencies are described here Eroukhmanoff 2018). Quantifying the impact of word frequencies on the stock market, or vice versa, would necessitate accounting for an undefined

---

[3] Our study does not require a sentiment-tagged dictionary as we aim to address the tones of the speeches.

[4] https://en.wikipedia.org/wiki/Glossary_of_economics

number of variables, potentially limiting our scope. Additionally, discerning the direction and consistency of variable impacts over time is challenging, given the evolving nature of presidential leadership and the institution itself (Rutledge & Larsen Price, 2014). The growth of the institutional Presidency and the resulting dynamics between the President and Congress, as discussed by (Dickinson & Lebo, 2007), highlight the complexities and changing credibility of the Presidency as an institution.

The paper is structured as follows: Sect. 2 lists comparable and complementary works related to our analytical methods. Section 3 details the datasets and their key statistical characteristics. In Sect. 4, we describe the methodological tools employed. Section 5 presents the results and related discussion. Finally, Sect. 6 offers concluding remarks.

2 Review of the literature

Textual data are attracting growing interest, particularly in fields where relevant information is disseminated through written documents, like pathology's diagnoses in medicine Huang et al. (2023), or in social science, like in Adamopoulos et al. (2018), where the authors used social media users' latent personality traits—extracted from unstructured textual data—to assess the impact of "word of mouth" on shaping their behaviours and preferences; and the notable case of Wei et al. (2002), in which advancements in document-category management for internet-sourced content were presented.

Getting closer to the core of our study, we mention scientific contributions that address the relationship of US presidencies and Presidents elections with financial markets to give further context for our study and to provide the reader with additional views on the matter. Santa-Clara and Valkanov (2003) investigated the relationship between political cycles and stock market performance, focusing on different US presidencies and election periods. Their findings revealed that presidential elections significantly affect stock market returns, suggesting a potential influence of speeches during these times, although they are less directly market-related than Federal Reserve announcements (Hayo et al., 2012). Santa-Clara and Valkanov (2003)'s methodology included a range of control variables like the log dividend-price ratio, the term spread, the default spread, and the relative interest rate, utilizing data from the CRSP portfolio and the DRI database. They employed linear regression to analyse the correlation between "political variables" and "financial variables", also incorporating time shifts, albeit on a monthly scale. Due to data distribution challenges and test power limitations, they used a Monte Carlo approach but faced interpretational difficulties, leading them not to present results post Bonferroni correction (see, Leamer and Leamer 1978). Kräussl et al. (2014) conducted a more detailed examination of the presidential cycle puzzle. They focused on the relationship around election dates but found no statistically significant evidence confirming a consistent presidential election cycle effect on stock returns. Their methodology, which involved using the S&P500's log-returns and dummy variables for presidential periods and Senate majorities, aimed to assess the explainability of stock market returns through political variables. They observed a business cycle pattern, suggesting various potential influences, including presidential actions on monetary policy, taxation, or spending. These actions could be communicated through legislative measures and presidential announcements, prompting our focus on the economic content of these speeches. Recent studies, like Brans and Scholtens (2020), have explored the impact of social media, particularly presidential tweets, on the stock market. They found that President Donald Trump's tweets did not significantly impact stock returns unless sentiment was included in the analysis. Their treatment of presidential tweets

aligns with our approach to presidential speeches, and they did not use control variables. Additionally, Kiessling et al. (2017) highlighted the influence of presidential signaling and rhetoric over a 20-year period, showing that presidential addresses on economic policy can affect market movements. Their findings, derived from a shorter list of economic words and a 20-year analysis period, complement our more extensive word list and longer study period. Abolghasemi and Dimitrov (2020) explored the causality between US presidential prediction markets and global financial markets, finding that US election-related market fluctuations can impact global financial indices. This suggests that presidential speeches, as vehicles for party positions and voter influence, can indirectly affect financial markets in various countries, underlining the US's unique and central position. Li and Born (2006) examined how presidential election uncertainty affects US stock returns, finding substantial evidence of political outcomes, including elections, influencing the business cycle and stock market. This supports the notion that presidential speeches can have wider financial market impacts beyond the US. Shaikh (2019) studied the 2012 and 2016 US presidential elections' effects on investor sentiment and stock market performance. Their empirical findings indicated a strong and significant relationship, including the influence of debates and speeches, on investor behavior and market dynamics, using indicators like the COBE VIX implied volatility index, which is S&P500 related. So, while some studies indicate that presidential speeches may not systematically influence financial markets, there is also evidence to suggest they can impact global investor sentiment, market volatility, and stock market returns, via, for exmple, their electoral campains messages. The relationship between presidential speeches and financial markets is complex and depends on various factors, such as the specific context and timing of the speeches.

In terms of methodological approach, our data treatment procedure is similar to those found in Tsai and Wang (2017), Alfaro et al. (2016), Maligkris (2017). However, the work most akin to our method is Goel and Chengalur-Smith (2010), which analysed organisational (public, private, educational, etc.) information security policies through content analysis of policy documents. Quoting the paper: *"Content analysis relies extensively on word frequencies to determine the themes (focus areas) and relative frequencies of the themes (or emphasis on focus areas)."* As stated in the Introduction and akin to Goel and Chengalur-Smith (2010), Maligkris (2017), we assess the economic content level in analysed speeches by referencing selected economic glossaries.

Our approach is comparable to Lavrenko et al. (2000), where the authors aligned news releases with current or immediately subsequent trends. Unlike Lavrenko et al. (2000), we investigate the co-influence of talks and speeches starting from a glossary, whereas the referenced paper aimed at creating language models for predicting stock market trends. Furthermore, our study differs from Maligkris (2017) in that we do not limit our analysis to speeches from presidential candidates and form Santa-Clara and Valkanov (2003) as it is more exploratory, focusing on trends and synchronicities between variables. We investigate the presence of an interaction between the economic content of U.S. presidential speeches and financial market movements, albeit recognizing similar challenges to those encountered by the authors in estimating the magnitude of this interaction.

Building on Santa-Clara and Valkanov (2003)'s findings, which demonstrated that the correlation between stock returns and political variables is not merely an indirect relation with business cycle factors, we posit that a synchronic behaviour between presidential economic addresses and market dynamics might exist. However, as they noted, the mechanism through which political variables impact stock returns remains an open question.

In the procedure we propose, we utilise rank-rank correlation. This technique is commonly employed in text analysis (for example, Baron et al. 2009) and is frequently used in sentiment

analysis and finance studies (see Tsai and Wang 2017, Pak and Paroubek 2011, Laih 2014). Our methodological decision is based on the collective aforementioned studies which provide a rich background for understanding the relationship between US presidential speeches, political cycles, investor sentiment, and the US stock market. They offer sufficient evidence to support our explorative approach to this investigation; in fact, benefiting from this literature and declaring that our objective is not to quantify the relationship between the entities at play or explain its direction, we deploy the analysis by means of tools described in Sect. 4.

The number of studies demonstrating a correlation between the stock market and political speeches using text mining techniques and rank-rank correlation is limited. A particularly relevant study is Preis et al. (2013), which, while not focusing on political messages, used Kendall's τ to measure the relationship between Google query volumes for finance-related search terms and the stock market.

In the realm of machine learning algorithms, distance measures are extensively used (see Nassirtoussi et al. 2014, for an extensive review of machine learning applications in stock market prediction through text mining methods). In news classification and impact studies, distances are a fundamental component. For example, Fung et al. (2002) employed Euclidean distance to segment stock time series and compare them with various classes of news grouped by keyword collections.

Many trading and prediction algorithms for news exploration involve distance measurements and are based on text analysis. In this respect, we refer to Shynkevich et al. (2015), Schumaker and Chen (2009) for providing further context.

Finally, Shannon entropy is commonly used in text mining, especially in text categorisation (see e.g. Largeron et al. 2011) and authorship attribution (see Rosso et al. 2009). In the financial sector, it is employed to analyse financial time series features or design portfolio selection strategies (see Bentes et al. 2008; Mendes et al. 2016; Tessmer et al. 1993). A notable review in finance is Zhou et al. (2013). Generally, entropy's use has a broad and longstanding history in information theory (see, e.g., Gray 2011). To our knowledge, this paper is the first to employ such a combination of tools to assess the similarity between the economic content of political speeches and the patterns of financial markets.

3 Data

We have chosen the Standard and Poor's 500 for our analysis as it is among the most representative index of the US stock market. The S&P 500 is often regarded as the quintessential stock index. It is recognized as the foremost representation of the "most significant financial markets" (the US exchanges) and is extensively referenced in authoritative financial literature (see, e.g., Golez and Koudijs 2018; Patton and Weller 2020) As the primary benchmark for large-cap US equities, the S&P 500 encapsulates approximately 13.5 trillion US dollars, either indexed or benchmarked against it. It comprises over 500 elite companies, representing around 80% of the available market capitalisation (S&P Dow Jones Indices, 2022). Consequently, the S&P 500 holds a pivotal position in the socio-economic-political landscape of the US, and the whole world. This prominence positions it at the heart of US Presidents' interests, providing a clear and direct motivation to explore the interplay between the economic content of US Presidents' speeches and the performance of such a stock index. Therefore, we have collated daily index closing prices and volumes from January 3rd, 1950, as these were the earliest available records from "Yahoo!Finance". The dataset extends up to March 1st, 2017, and includes an additional day following President Donald Trump's speech on Febru-

ary 28th, 2017 (Address to Joint Session of Congress), to encompass the latest presidential speech available in the dataset, in our analysis.

The corpus of Presidents' speeches analysed in this study comprises 951 addresses, as catalogued in Ficcadenti et al. (2019), Cinelli et al. (2021). This collection spans from George Washington's tenure in 1789 to Donald Trump's in February 2017[5]. We tailored this corpus to include speeches from January 3rd, 1950 onwards, aligning with the start of the S&P 500 series available on "Yahoo! Finance". The earliest speech in this adjusted corpus is Harry S. Truman's "Report to the American People on Korea" dated April 11th, 1951.

We compiled a list of economic terms from Bishop (2009), where the author elucidates "the most important economic terms and concepts." His work focuses on economics as it relates to jobs, prices, trade, and its impact on everyday life, drawing on articles from *The Economist* as a prime source. This list was expanded by incorporating terms from https://en.wikipedia.org/wiki/Glossary_of_economics. Following the methodology in Cinelli et al. (2021), we processed this compilation to derive a glossary of 383 economic terms. The presence and frequency of these terms in the speeches were analysed to gauge the economic emphasis of each address. We calculated the absolute frequency of economic terms in each speech and determined their relative frequency by dividing these counts by the total word count of the respective speeches.

A minor bias exists in our approach to defining relative frequency, as the proportion of economic words in a speech might not have one as an upper bound. However, the effect of such a bias is negligible and does not detract from the validity of our analysis, similar to the approach in Goel and Chengalur-Smith (2010), where authors used this method for evaluating the significance of Breadth in security policy-related documents.

Figure 1 illustrates the relative frequencies of economic terms in the speeches over time.

This procedure aligns with the methodologies of Baker et al. (2016), Tsai and Wang (2017), and others, discussed in Sect. 2. For example, Baker et al. identified a set of terms related to economic insecurity to create an index of economic policy uncertainty, while Tsai and Wang utilised word lists to analyse financial reports of firms. Similarly, Wei et al. (2015) selected keywords to identify relevant studies in climate change modelling.

For comparative analysis, we normalised all data, including the absolute frequencies of economic terms in speeches and the S&P 500's daily closing prices, volumes, and returns. Let $\{x_1, \ldots, x_N\}$ represent a time series with N realisations, where $t = 1, \ldots, N$ corresponds to a trading day for financial variables and a speech date for US Presidents' talks. The normalised value of x_t is given by

$$x'_t = \frac{x_t - min(x)}{max(x) - min(x)}, \qquad t = 1, \ldots, N, \tag{1}$$

where

$$max(x) = \max_{s=1,\ldots,N} x_s$$

and

$$min(x) = \min_{s=1,\ldots,N} x_s$$

---

[5] Presidents George Washington and John Adams initiated the tradition of Annual Messages to Congress on the State of the Union, delivered in person. Subsequent communications were often written, but the trend of live addresses was revived with President Franklin D. Roosevelt's Fireside Chats. This evolution is detailed in Peters and Wooley (2019), highlighting the predominance of spoken State of the Union Addresses in our study period. We sourced speech transcripts from the Miller Centre website (https://millercenter.org/), including both originally written and orally delivered communications.
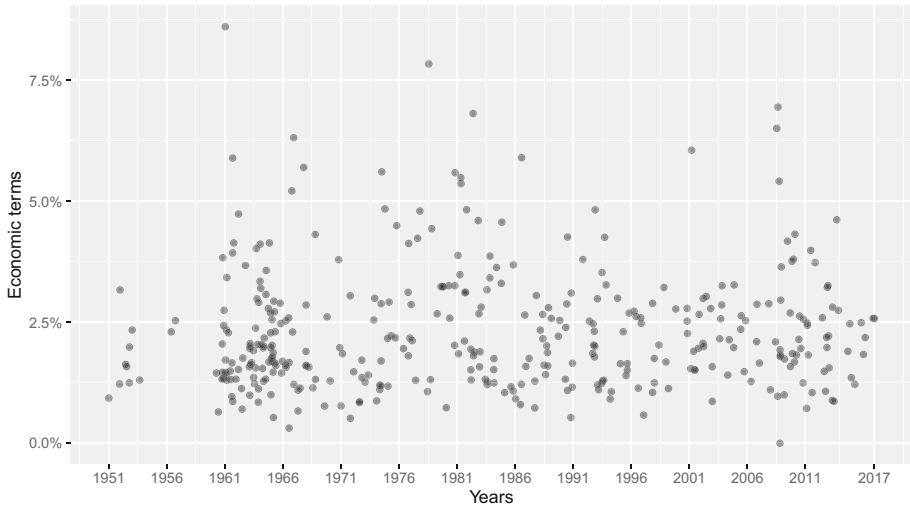
Fig. 1 Percentage of economic terms in each speech over the years, calculated as the proportion of economic terms relative to the total word count in the speech

are the maximum and minimum values across the series. This normalization allows for a direct comparison between variables with differing scales, making it ideal for our context. Note that the relative frequencies of economic terms fall in [0, 1] by definition and thus do not require normalization.

To compare the speeches with the S&P 500 variables, these datasets have to be aligned. As mentioned, all the talks delivered before January 3rd, 1950 (the date on which the first S&P 500's closing price is observable in the considered dataset) are not considered. Since the first available speech after January 3rd, 1950, dates back to April 11th, 1951, the financial information is considered from April 10th, 1951. We start "one day before" the speech delivery date to allow the analysis of the financial data recorded before the speech. So, the resulting considered messages are 376.

To compare speeches' economic content and the S&P500, we need to ensure that information is recorded for both entities. The speeches are often delivered during weekends or days when the market has not opened, so the number of days we have info for both entities is 327. As we will see in Sect. 4, we perform analyses comparing the speeches' economic content against the S&P 500 realisations registered "one day before", "contemporaneously" and "one day after" the speeches' dates. Thus, it happens that data availability changes. To better explain how the differences in the number of speeches available for the analysis occur, we use the example driven by the following two transcripts:

- Saturday, 13/03/1965 – Press Conference at the White House, Lyndon B. Johnson
- Monday, 15/03/1965 – Speech Before Congress on Voting Rights, Lyndon B. Johnson

In the "contemporaneous" analysis, the speech on 13/03/1965 is excluded due to the absence of corresponding financial data (it being a Saturday). However, when assessing its economic content against the S&P 500 data from the day before, the financial data from 12/03/1965 is considered. For the 15/03/1965 speech, if the analysis includes market data from the day before, only the information from 12/03/1965 is relevant, as the 13th and 14th were non-trading days. Thus, both speeches are included in this case, and the same applies when comparing with financial data from the day after the speech.

Table 1 The second column indicates the number of speeches delivered on each day of the week from April 30, 1789, to February 28, 2017. The third column reflects the number from April 11, 1951, to February 28, 2017

Day of the week	No. of speeches in the entire dataset	No. of speeches in the adjusted dataset
Monday	196	59
Tuesday	240	78
Wednesday	151	83
Thursday	151	73
Friday	98	42
Saturday	75	19
Sunday	40	22

There are 41 speeches delivered on weekends (see Table 1). The difference between 376 and 327 (49, see Tables 5 and 4) includes speeches on non-trading days. For instance, President Barack Obama's Second Inaugural Address on Monday, 21/01/2013 (Martin Luther King Jr. Day), is an example where the market was closed, and it was not a weekend.

4 Methodology

To investigate the interplay between the economic terminology in Presidential speeches and the S&P 500 index, we have considered several scenarios. These are outlined below for ease of reference, examining various pairings:

(a) Normalised S&P 500 returns observed on the day before, the day of, and the day after the President's speeches, each paired with the relative frequencies of the selected economic terms.

(b) Normalised S&P 500 returns observed on the day before, the day of, and the day after the President's speeches, each paired with the normalised absolute frequencies of the economic terms.

(c) Normalised S&P 500 closing prices observed on the day before, the day of, and the day after the President's speeches, each paired with the relative frequencies of the economic terms.

(d) Normalised S&P 500 closing prices observed on the day before, the day of, and the day after the President's speeches, each paired with the normalised absolute frequencies of the economic terms.

(e) Normalised S&P 500 volumes observed on the day before, the day of, and the day after the President's speeches, each paired with the relative frequencies of the economic terms.

(f) Normalised S&P 500 volumes observed on the day before, the day of, and the day after the President's speeches, each paired with the normalised absolute frequencies of the economic terms.

The enumerated cases (a) through (f) encapsulate six investigated relationships, combining two variables from the economic glossary (relative and absolute frequencies) with three market variables (closing prices, returns, and volumes).

The analysis of the days preceding and following the speeches aims to discern if the market reacts in anticipation of the speeches or, alternatively, if it seems to pre-empt the economic themes addressed in the speeches.

We employed Kendall's τ rank-rank correlation for our correlation analysis. This indicator is calculated for pairs of series observed jointly and of equal length; denoted as

Table 2 A synopsis of the variables considered in this study. The market variables analysed include the normalised S&P 500's daily returns, normalised closing prices, and volumes. From the speeches, we examine both the normalised absolute and relative frequencies of the economic glossary, as described in Sect. 3

Market variables ↓	Economic glossary variables → Instances ↓	Rel. freq	Norm. abs. freq
Norm. S&P 500 returns	One day before	a	b
	Contemporaneous		
	One day after		
Norm. S&P 500 closing prices	One day before	c	d
	Contemporaneous		
	One day after		
Norm. S&P 500 volumes	One day before	e	f
	Contemporaneous		
	One day after		

$\{(k_i, h_i)\}_{i=1,\ldots,N}$. The computation procedure initiates by ranking the ks and the hs in either ascending or descending order. These ranks are then coupled based on their original joint observations. The Kendall rank correlation coefficient τ is defined as:

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{N(N-1)/2} \tag{2}$$

A pair of observations (k_i, h_i) and (k_j, h_j) is concordant if the ranks of both k and h agree post-sorting. The number of such concordant pairs is given by:

$$\begin{cases} k_i < k_j \\ h_i < h_j \end{cases} \quad \text{or, alternatively,} \quad \begin{cases} k_i > k_j \\ h_i > h_j \end{cases} \tag{3}$$

The number of pairs not satisfying equation (3) are classified as discordant.

We decided to make the analysis through Kendall's rank-rank correlation because we want to measure the ordinal association between economic content appearance in the speeches (frequency of locutions presence) and the considered financial variables. Namely, we look for the presence of a relation at a rank level among the occurrences of economic terms in the talks and the S&P 500's daily prices, returns and volumes. In so doing, we avoid the distorting effect on correlation due to the very different sizes of the quantitative terms of interest, hence, measuring the association between the considered variables. So, Kendall's τ is suitable for our case because it allows for an evaluation of the strength of association between the considered variables, which are radically different. Indeed, such a methodological device helps to visualise the synchronicity of the occurrences of economic content in speeches and the S&P 500 levels. In other words, Kendall rank-rank correlation allows for an evaluation of positive or negative strength of co-influence of the elements under analysis. Therefore, the correlation of rank-rank nature allows disregarding assumptions on the data's empirical distributions (Mata & Fuerst, 1997).

The association between frequencies of a set of words and another variable of interest can explain if stressing a theme in a speech is synchronised with such a variable. From this perspective, ranks are more informative than sizes. Indeed, ranks are obtained by comparing frequencies of words that occurred along the considered history (because they result from the ranking exercise), while sizes are absolute terms. Dealing with ranks allows exploiting

the information brought by the speeches in the period and the implementation of imitative behaviours among Presidents to gain insights into the performance of the S&P 500—the main US financial index. This offers immediate and intuitive information on the role of Presidents speeches' economic content in the interaction with the financial markets. This argument explains why rank-rank correlation is often employed in text analysis (for example, see Baron et al. 2009; Teevan et al. 2018). For an extensive review of Kendall's τ and other similar measures, see Kruskal (1958).

We have also explored the datasets using various distance measures between the normalised frequencies of economic glossary terms and the S&P 500 data. The analysed cases are (a),(b),(c),(d),(e) and (f) of the list above. These measures are:

$$d_{mx}(f, S\&P500^{(k)}) = max_t |f_t - S\&P500_{t+k}| \tag{4}$$

$$d_{mn}(f, S\&P500^{(k)}) = min_t |f_t - S\&P500_{t+k}| \tag{5}$$

$$d_{am}(f, S\&P500^{(k)}) = \frac{1}{T^{(k)}} \sum_{t=1}^{T^{(k)}} |f_t - S\&P500_{t+k}| \tag{6}$$

$$d_{ec}(f, S\&P500^{(k)}) = \sqrt{\sum_{t=1}^{T^{(k)}} (f_t - S\&P500_{t+k})^2} \tag{7}$$

where $k = -1, 0, 1$; $S\&P500^{(k)}$ is representing the normalised S&P 500 data (volume, closing price, or return) observed one day before, on the day, and one day after for $k = -1, 0, 1$, respectively. $T^{(k)}$ denotes the total number of observations, which varies with the time selection; f_t represents the summed (absolute normalised or relative) frequency of the economic terms in the speech at time t.

Furthermore, we computed and compared the Shannon entropy for each data series to quantify and discuss the information contained therein. At this aim, the variation range of each series is divided into N intervals of equal size. Entropy is defined as

$$H = -\sum_{j=1}^{N} p_j \log_2 p_j \tag{8}$$

where p_j is the probability of an observation falling within class j. In our analysis, N is set to 320, based on empirical considerations and trials. For further details, see Shannon (2001).

5 Results and discussion

In this section, we present and discuss the results derived from the analysis of three S&P 500 variables: normalised returns, closing price, and volumes and two variables from an economic glossary: relative and normalised absolute frequencies. Our analysis considers three distinct temporal instances: "one day before", "contemporaneous", and "one day after" the event of interest. A synopsis of the variables and instances considered in this study are presented in Table 2.

Tables 3, 4, and 5 offer detailed statistical summaries. These Tables compare the relative and absolute normalised frequencies of economic terms with the normalised variables of the S&P 500 for each of the specified time frames: contemporaneous, one day prior, and one day subsequent to the speeches.

Table 3 Statistical summary of the variables used for evaluating the relationship between the Presidents' speeches and the S&P 500 observations on the same days the talks are delivered. * The inverse of the variation coefficient is chosen. We mention Kalkur T. and Rao (2017), where the authors make a methodological proposal for estimating both the coefficient of variation and its inverse

	Economic Rel. Freq	Economic Abs. Freq. Norm	Prices Norm	Volume Norm	Ret. Norm
N. Obs	327.000	327.000	327.000	327.000	327.000
Max	0.086	1.000	0.986	0.685	0.874
Min	0.000	0.000	0.000	0.000	0.516
Median m	0.021	0.063	0.053	0.007	0.678
Mean μ	0.024	0.102	0.190	0.084	0.680
RMS	0.027	0.146	0.296	0.174	0.681
St. Dev. σ	0.013	0.105	0.227	0.152	0.032
Var	0.000	0.011	0.051	0.023	0.001
Sd. Err	0.001	0.006	0.013	0.008	0.002
Skewness	1.437	2.848	1.247	1.925	0.549
Kurtosis	5.791	19.239	3.517	5.723	9.729
μ/σ*	1.831	0.969	0.840	0.555	21.150
$3(\mu - m)/\sigma$	0.788	1.111	1.811	1.533	0.214

Table 4 Statistical summary of the variables used for evaluating the relationship between the Presidents' speeches and the S&P 500 on the day before the talks' delivery dates

	Economic Rel. Freq	Economic Abs. Freq. Norm	Prices Norm	Volume Norm	Ret. Norm
N. Obs	376.000	376.000	376.000	376.000	376.000
Max	0.086	1.000	0.989	0.651	0.799
Min	0.000	0.000	0.000	0.000	0.519
Median m	0.020	0.058	0.048	0.005	0.677
Mean μ	0.023	0.096	0.187	0.078	0.678
RMS	0.026	0.141	0.294	0.161	0.678
St. Dev. σ	0.013	0.103	0.227	0.141	0.029
Var	0.000	0.011	0.052	0.020	0.001
Sd. Err	0.001	0.005	0.012	0.007	0.002
Skewness	1.507	2.957	1.287	1.931	−0.273
Kurtosis	6.161	19.640	3.609	5.686	8.939
μ/σ	1.822	0.931	0.823	0.551	23.172
$3(\mu - m)/\sigma$	0.733	1.104	1.841	1.546	0.017

Furthermore, we have computed Kendall's τ coefficient to evaluate the rank-rank correlation for pairs (a), (b), (c), (d), (e), and (f). These calculations' results are documented in Table 6.

To augment our analysis, we have calculated distances in accordance with the formulas (4), (5), (6), and (7). The outcomes of these computations are detailed in Table 7. Lastly, the Shannon entropies, as defined in formula (8), are elucidated in Table 8, providing a comprehensive view of the data's informational content.

Figure 2 presents the histograms of the economic terms' relative frequencies. They are asymmetric with positive skewnesses, which suggests right-tailed distributions (see also Tables 3, 4 and 5). Furthermore, the values of the kurtoses indicate leptokurtic behaviours.

Table 5 Statistical summary of the variables used for evaluating the relationship between the Presidents' speeches and the S&P 500 observations on the day after the talks' delivery dates

	Economic Rel. Freq	Economic Abs. Freq. Norm	Prices Norm	Volume Norm	Ret. Norm
N. Obs	376.000	376.000	376.000	376.000	376.000
Max	0.086	1.000	1.000	0.708	0.814
Min	0.000	0.000	0.000	0.000	0.517
Median m	0.020	0.058	0.047	0.006	0.679
Mean μ	0.023	0.096	0.187	0.082	0.677
RMS	0.026	0.141	0.295	0.170	0.678
St. Dev. σ	0.013	0.103	0.228	0.149	0.033
Var	0.000	0.011	0.052	0.022	0.001
Sd. Err	0.001	0.005	0.012	0.008	0.002
Skewness	1.507	2.957	1.286	1.976	−0.275
Kurtosis	6.161	19.640	3.607	6.007	8.341
μ/σ	1.822	0.931	0.823	0.548	20.692
$3(\mu - m)/\sigma$	0.733	1.104	1.849	1.523	−0.134

There is an evident presence of outliers: for example, some speeches exhibit more than 7.5% of terms related to economics. Furthermore, Figs. 2 and 3 display the frequencies of economic terms, adjusted to align with the S&P500 data according to the specified instances. Specifically, we analyse the economic content of speeches in relation to financial data recorded one day before, on the same day, and one day after the speeches. The differences in these histograms primarily result from the exclusion of speeches for which corresponding financial data were unavailable, thus precluding proper alignment (refer to Sect. 4 for more details). These histograms, in conjunction with Tables 1, 3, 4, and 5, provide insights into the empirical distribution of the observed variables and their variations for each instance.

Figure 3 shows the histograms of normalised absolute frequencies. The asymmetry indexes in Tables 3, 4 and 5 manifest the presence of highly skewed distributions, and kurtoses almost tripled with respect to the ones of the previous case. The outcome is motivated by the higher concentration of observations on the left side of the distributions and the presence of outliers.

Figures 4 and 5 illustrate the empirical distributions of the normalised daily closing prices and volumes. These should be considered alongside Tables 3, 4, and 5. The data suggest that on the days preceding and following presidential speeches, prices tend to hover near the series' minimum (i.e., the minimum before normalisation), except on the days of the speeches themselves, where prices close slightly further from their minimum. A similar trend is observed in trading volumes, with an increased number of transactions on the days of the speeches. In contrast, returns (as depicted in Fig. 6) tend to cluster around the average on the day before and on the day of the speeches.

A visual examination of Figs. 4 and 5 reveals asymmetries, characterised by less pronounced right tails compared to previous cases. The kurtosis is more pronounced in the case of normalised volumes, with the tallest bin in the histograms containing the majority of values, predominantly clustering around zero. These occurrences influence all statistical position indicators and result in lower series variances, which are smaller than those calculated for normalised prices. The presence of outliers is also significant in these cases.

Figure 6 shows different behaviours for the considered variables. This outcome supports the idea of more symmetric distributions of normalised returns, with skewness values very close to zero and means and medians almost coinciding. The variances are tiny inasmuch

Table 6 The Kendall's τ estimations. Each couple of columns contains results coming from a different sampling time of S&P 500 observations

Instances → Statistical Indicator → Couple analysed ↓	One day back of S&P 500 obs.		Contemporaneous date		One day ahead of S&P 500 obs.	
	τ	p-value	τ	p-value	τ	p-value
Norm. Ret. and Rel. Freq. - (a)	0.008	0.412	0.014	0.338	−0.023	0.255
Norm. Ret. and Abs. Freq. Norm. - (b)	0.008	0.416	0.011	0.370	−0.017	0.312
Norm. Closing Price and Rel. Freq. - (c)	0.045	0.111	0.047	0.086	0.045	0.096
Norm. Closing Price and Abs. Freq. Norm. - (d)	0.085	0.011	0.088	0.005	0.088	0.006
Norm. Vol. and Rel. Freq. - (e)	0.061	0.051	0.062	0.037	0.063	0.034
Norm. Vol. and Abs. Freq. Norm. - (f)	0.097	0.004	0.096	0.003	0.100	0.002

The first block shows the analysis results of S&P 500 variables observed the day before the Presidents' speeches. The second and the third couples of columns report the cases of the same day and the subsequent day of the Presidents speeches, respectively

Table 7 The different distance measures described in Sect. 4 are computed and reported, a reference them is found in the first column

Measure used	Instance	Data used	One day back of S&P 500 obs	Contemporaneous dates	One day ahead of S&P 500 obs
Eq. (4)	(d)	Abs. Freq. Norm. and Norm. Closing Price	0.981	0.981	0.981
Eq. (4)	(f)	Abs. Freq. Norm. and Norm. Vol	1.000	1.000	1.000
Eq. (4)	(b)	Abs. Freq. Norm. and Norm. Ret	0.755	0.771	0.775
Eq. (4)	(c)	Rel. Freq. and Norm. Closing Price	0.963	0.961	0.974
Eq. (4)	(e)	Rel. Freq. and Norm. Vol	0.607	0.662	0.691
Eq. (4)	(a)	Rel. Freq. and Norm. Ret	0.788	0.805	0.784
Eq. (5)	(d)	Abs. Freq. Norm. and Norm. Closing Price	0.001	0.000	0.000
Eq. (5)	(f)	Abs. Freq. Norm. and Norm. Vol	0.001	0.001	0.001
Eq. (5)	(b)	Abs. Freq. Norm. and Norm. Ret	0.178	0.190	0.129
Eq. (5)	(c)	Rel. Freq. and Norm. Closing Price	0.000	0.000	0.000
Eq. (5)	(e)	Rel. Freq. and Norm. Vol	0.000	0.000	0.000
Eq. (5)	(a)	Rel. Freq. and Norm. Ret	0.449	0.498	0.481
Eq. (6)	(d)	Abs. Freq. Norm. and Norm. Closing Price	0.171	0.173	0.172
Eq. (6)	(f)	Abs. Freq. Norm. and Norm. Vol	0.115	0.123	0.118
Eq. (6)	(b)	Abs. Freq. Norm. and Norm. Ret	0.584	0.580	0.583
Eq. (6)	(c)	Rel. Freq. and Norm. Closing Price	0.169	0.171	0.169
Eq. (6)	(e)	Rel. Freq. and Norm. Vol	0.081	0.087	0.085
Eq. (6)	(a)	Rel. Freq. and Norm. Ret	0.654	0.656	0.654
Eq. (7)	(d)	Abs. Freq. Norm. and Norm. Closing Price	5.034	4.668	5.041
Eq. (7)	(f)	Abs. Freq. Norm. and Norm. Vol	3.265	3.230	3.392
Eq. (7)	(b)	Abs. Freq. Norm. and Norm. Ret	11.475	10.641	11.467
Eq. (7)	(c)	Rel. Freq. and Norm. Closing Price	5.438	5.086	5.446
Eq. (7)	(e)	Rel. Freq. and Norm. Vol	2.919	2.948	3.099
Eq. (7)	(a)	Rel. Freq. and Norm. Ret	12.706	11.885	12.701

The last three columns refer to selections of the S&P 500 variables. The fifth column is about the case of S&P 500 data taken on the same dates as the Presidents' speeches, while the fourth and sixth columns are the cases of the day before and the day after the speeches, respectively. The column titled"Instance" report the leterrers used to identify the analysis as described in Sect. 4

Table 8 Shannon entropy for each series. The columns distinguish the different data selections made in accord with the selected dates. The third column points to the S&P 500 observations registered on the same days Presidents have stated their talks, while the second and the last ones are about the series of S&P 500 observations registered the days before and after the Presidents' talks

	One day before the speeches' dates	Contemporaneous dates	One day after the speeches' dates
Terms rel. freq	6.872	6.825	6.872
Terms abs. freq. norm	5.899	5.951	5.899
S&P 500 Closing price norm	6.031	5.973	5.962
S&P 500 Vol. norm	4.380	4.436	4.375
S&P 500 Ret. norm	6.308	6.202	6.488
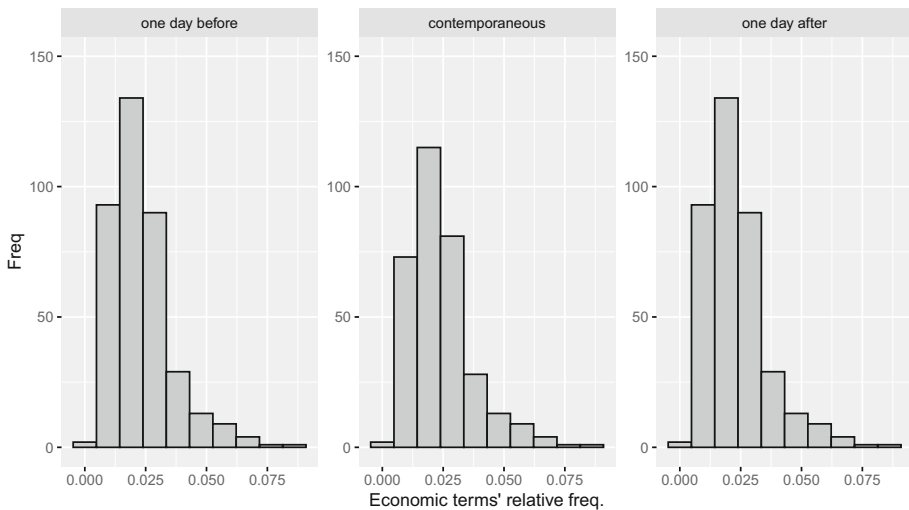
Fig. 2 Histograms illustrating the relative frequencies of economic terms. These histograms are divided into sub-figures based on the dates selected. In the "contemporaneous" case, speeches without corresponding S&P 500 observations are omitted. This includes speeches on days like Sunday, when the financial market is closed, thereby lacking S&P 500 data. Such instances are excluded from the analysis to ensure accurate comparison with market activities

as the distributions have a high concentration of values around their centres. From a visual inspection, it is possible to note that the outliers are present in all the histograms, but in the cases of the day after the Presidents' talks dates, we have a slightly fatter right tail.

Table 6 demonstrates that the τ estimations lack statistical significance when involving the S&P 500 normalised returns, as evidenced in pairs (a) and (b). This is also observed in pairs comprising normalised closing prices and the relative frequencies of economic terms, as in the (c) scenarios.

Conversely, other cases exhibit markedly high levels of statistical significance. In instances (d), there exists a positive correlation between the prevalence of economic expressions and the normalised closing prices. Notably, the weakest positive rank correlation is observed on the day preceding the speeches, while the correlations recorded on the speech day and the following day are marginally higher. In (f), the p-values are substantially lower compared to those in (d) and (e), indicating exceptionally significant τ estimations. The analysis for
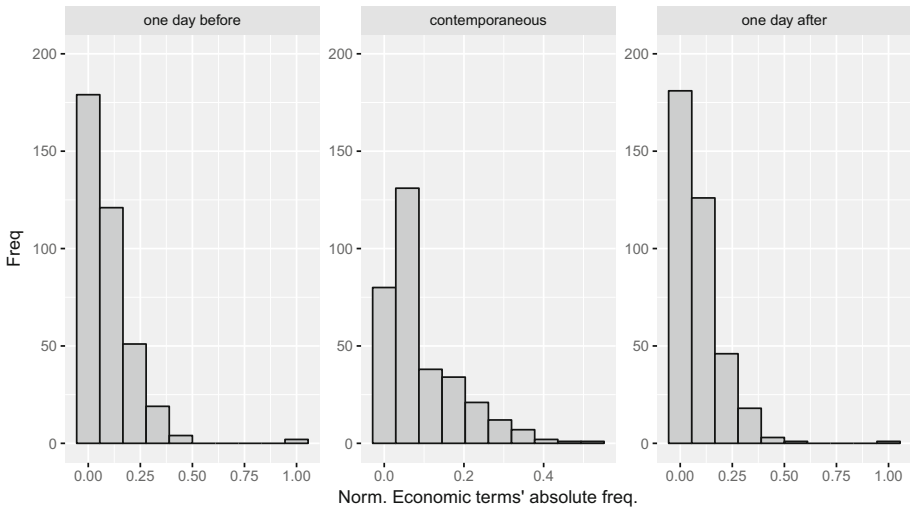
Fig. 3 Histograms depicting the normalised absolute frequencies of economic terms. Variations in the histograms arise from the different dates selected. In the "contemporaneous" scenario, speeches lacking corresponding S&P 500 observations are excluded. For instance, speeches delivered on Sundays are omitted, as there are no S&P 500 observations on these days due to market closure. This exclusion ensures that only speeches with relevant financial market data are considered in the analysis
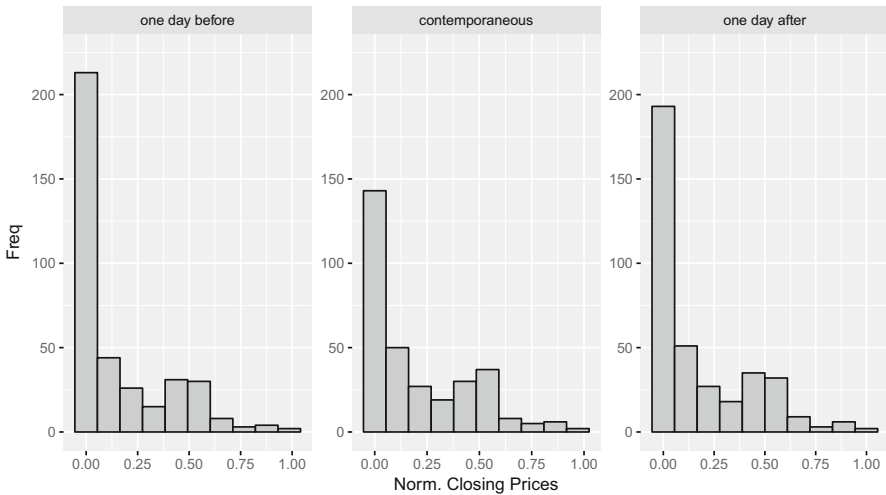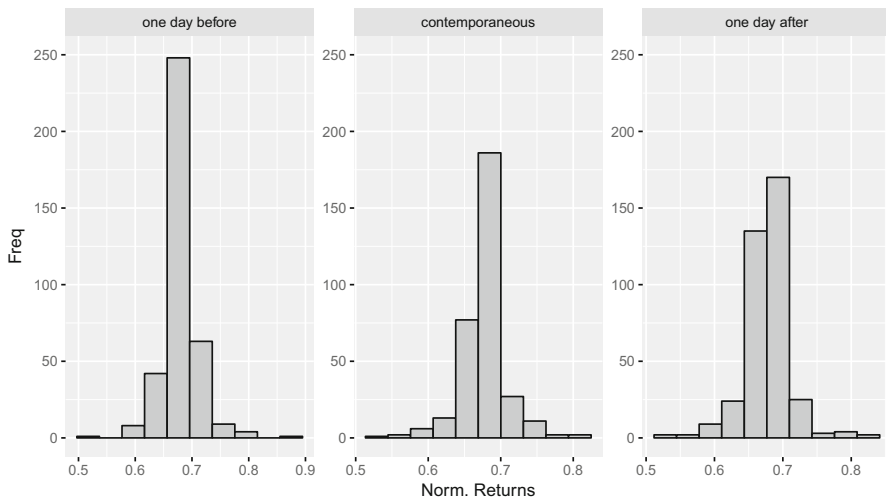

Fig. 4 Histograms of the S&P 500 normalised daily closing prices

(e) on the day prior to the speeches is rendered inconclusive due to a p-value of 0.051. However, the latter two columns of Table 6 for this case show statistically significant and positive correlations. The most pronounced positive τ value, reaching 10%, is observed in the interplay between the normalised frequency of economic terms and the transaction volume on the day following the speeches. For (f), the lowest correlation is 9.6%; a minor increment occurs between the volume of the day before the speech and the economic words' presence.
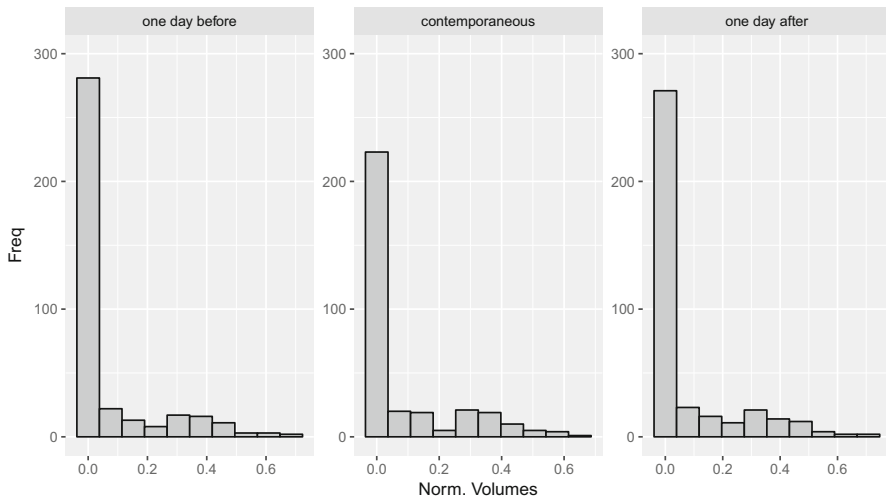
Fig. 5 Histograms of the S&P 500 normalised daily volumes

Fig. 6 Histograms of the S&P 500 normalised daily returns

The positive rank-rank correlations between the inclusion of economic locutions in the speeches and the S&P 500 index range from a minimum of 8.5% in (d)—bserved when normalised closing prices are logged a day before the speeches—o a maximum of 10% in (f), specifically when S&P 500 volume data are recorded on the days following the speeches. These correlations corroborate the intuitive notion that the considerable influence of both US Presidents and the stock market, coupled with their interrelations, inevitably intersects their spheres of influence. Consequently, the estimated correlations reveal that the economic discourse of US Presidents both influences and is influenced by stock market dynamics. In this context, the positive correlations between trade volume and the mention of economic terms—as in cases (e) and (f)—potentially indicate the traders' attentiveness to the economic content of US Presidents' speeches. In essence, the volume of trades escalates in proximity to

these speeches, suggesting that a higher incidence of economic terminology in the speeches correlates with an increased volume of transactions.

From this analysis, we can partially elucidate the variations in price observed within the specified time frames—from one day preceding to one day following the speeches—through the lens of the economic substance in Presidents' addresses. Notably, the correlations evolve over time, exhibiting an increase on the days of the speeches ("at the date" comparison) in scenarios (d) and partially (c). On these occasions, the influence of Presidents' speeches on pricing can be attributed to factors such as rumours (since speeches are typically delivered post-market closure), the content of the messages, or merely due to anticipation of the President speaking recently anticipated by, for example, Tweets.

The price appears to mirror the traders' tentative expectations regarding the economic substance of the speeches on the preceding day. For instance, when a speech is scheduled on the political agenda, market participants formulate trading strategies based on their assumptions about the forthcoming information from the President. Subsequently, price adjustments also materialise as a result of modifications in traders' expectations following the speech. Specifically, both price and volume undergo changes as they reflect shifts in traders' perspectives, with strategies being revised to incorporate the newly acquired information (evident in the correlation increases for observed S&P 500 variables on the day after the speeches). The measured correlation on the volume, especially in cases (f), further substantiates this observation.

Let us now examine the distances delineated in Table 7. Distances calculated using equations (4) and (5) facilitate the identification of the extremities in the deviation ranges between series. Observing the results from equation (4), one notices that the distance estimates are considerably large and exhibit homogeneity across both the columns and the pairs of comparisons. Contrastingly, equation (5) reveals notable disparities between the estimates, including some instances of null values. It is particularly noteworthy that pairs (a) and (b), which involve returns, display the highest minimum levels across the columns.

Distances derived from formulas (6) and (7) highlight two distinct facets. The former represents a fair mean of the distances between points, while the latter mitigates the impact of larger distances by squaring numbers in the (0,1) range. Thus, we obtain a balanced measure alongside an underestimation of the differences between the series in question.

Regarding distances computed via equation (6), the outcomes for pairs (a), (c), and (d) exhibit minimal variations across the columns, in contrast to the more pronounced changes seen in other pairs. Excluding case (b), a consistent pattern emerges: the distances are notably greater in contemporaneous instances compared to those where S&P 500 data is recorded one day before or after the speeches. Additionally, the distances recorded for the latter exceed those for the former. The magnitude of these differences is more pronounced when volumes are included in the analysis, as seen in pairs (e) and (f). This observation suggests that the selected variables demonstrate varied sensitivities towards each other and differ according to the temporal dimensions chosen. Employing the mean (6), the smallest distance across the columns in Table 7 is observed for pairs analysing the relative frequencies of economic terms and normalised volumes. These results corroborate the findings from Kendall's τ (refer to Table 6) and further emphasize the correlation between S&P 500 volume and the presence of economic terms in the speeches.

Distances calculated using formula (7) yield intriguing findings for pairs (a), (b), (c), (d), and (f). Minimum values are observed in "contemporaneous" observation dates ("at the date" case), with the maximum occurring when considering S&P 500 data the day after the speeches, and a slight deviation when the data are recorded a day before. This pattern can be attributed to traders adjusting their positions in anticipation of the speeches, subsequently

influencing both price and volume. The minimum Euclidean distance for pairs (e) reaffirms that S&P 500 volumes are closely aligned with the frequency of economic terms, thereby validating conclusions drawn from the τ coefficient estimations. Notably, this pair exhibits a unique trajectory across different time windows.

Conversely, when employing both formulas (6) and (7), it becomes apparent that the series most distanced from the frequencies of economic locutions (normalised absolute and relative) are the normalised daily returns series. This finding aligns with the limited statistical significance observed in the τ estimations involving returns. The prominent cases of distances measured with equation (5) further reinforce this conclusion, with pairs involving normalised returns displaying the most significant deviations from zero.

Table 8 presents the outcomes derived from Shannon's entropy, calculated using formula (8). It is observed that the series representing the relative frequencies of economic terms manifests the highest entropy, whereas the S&P 500 volume series exhibits the lowest.

The relative frequency series of locutions closely aligns with the returns in terms of entropy. Likewise, the entropy of the locutions' normalised absolute frequencies bears a close resemblance to that of the S&P 500 normalised closing prices. These observations suggest that the degrees of disorder in these series are similar on a pairwise basis, and this parallel holds true in scenarios one day before, on the day, and one day after the events in question. When juxtaposing this insight with the results derived from Kendall's τ and the distances, it is feasible to propose that the frequencies of economic locutions might serve as a proxy for the distributions and statistical characteristics of returns and closing prices, albeit with non-simultaneous behaviours. However, the entropy analysis does not enable assertions regarding potential deviations or convergences in the paths of returns, prices, and frequencies of economic terms. The alignment in entropy is more pronounced in the case of normalised absolute frequencies as opposed to relative frequencies, particularly with respect to normalised closing prices. The most significant divergences are noted between the frequencies of locutions and the volumes. Reflecting upon the earlier discussions concerning Kendall and distances, it is evident that while the volume series exhibit positive rank-rank correlations with frequencies, their empirical distributions, and those of the frequencies, differ in shape. Consequently, while the frequencies of economic terms can be utilised as proxies for analysing the evolutionary properties of volumes, they are less suited for examining the statistical attributes of the corresponding empirical distributions.

6 Conclusive remarks

The presence of economic terminology in US Presidents' speeches exhibits a discernible association with the Standard and Poor's 500 index, with the degree of this interconnection varying depending on the specific S&P 500 variables considered. As per the evidence presented, the S&P 500 volume is the most sensible to the presence of economic expressions in these addresses. While initial indications of this are perceptible in Fig. 5, it is the Kendall τ coefficient that more convincingly suggests a positive and significant correlation. The distance measurements further corroborate a high level of synchronisation between these series. Moreover, the Shannon entropy values reveal comparable patterns between the volume's fluctuations and the frequencies of economic terms, thereby affirming the relationship from an alternate standpoint.

Concerning the S&P 500 closing prices, their correlation with the relative frequencies of economic locutions is not statistically significant as per the τ coefficient. Conversely,

when considering normalised absolute frequencies, the Kendall τ demonstrates positive—and significant—correlations. The distance measurements in these scenarios are higher compared to the latter cases, and the entropy values for the closing price series more closely align with those associated with the economic word frequency series. Mild indicators of variation in the empirical distribution shapes, as suggested by Fig. 4, intuitively support this observation. Thus, the insights regarding the interplay between the closing prices and the economic content of Presidential speeches align with those observed for the volume, albeit with a lesser intensity in the S&P 500 prices. The distributions disorder (measured by Shannon's entropy) proves the presence of shared characteristics between the closing price series and the frequency of economic terms. This relation is remarkably lower in the context of trading volumes.

In contrast, the S&P 500 daily returns do not exhibit significant co-movements with the economic locutions within the speeches, particularly when analysing Kendall's τ. Various distance measurements confirm that the returns series consistently diverge from the series of glossary terms' frequencies, which may partially explain the lack of statistical significance in the τ estimations. However, when assessing Shannon's entropy, the return series' entropies are akin to those of the economic terms' frequencies. This finding implies that, although an instantaneous relationship may be absent, a connection at the level of empirical distributions could exist.

In light of these findings, several practical implications emerge:

- Presidents may emphatically emphasise economic themes in their speeches to facilitate market trading activity.
- A President's speech scheduled in the near future generates a movement in the market— mainly when the speech is assumed/expected to have economic content. Therefore, announcing the economic content of a speech beforehand is a way to interact with the financial market.
- There exists a limited (positive) relationship between returns, prices, and the economic content of speeches. In terms of correlation, this relationship is insignificant for returns but significant for prices. Consequently, the economic content of Presidential speeches is not a reliable policy instrument for influencing daily index levels and changes.

These results are relevant, and succinctly demonstrate that the economic content of Presidential speeches has a more profound impact on, and is more influenced by, the S&P 500's daily volumes than by its prices and returns. Therefore, the findings facilitate the identification of an immediate communication-based strategy for interacting with some market's features.

Furthermore, this study underscores the necessity of contemplating whether our findings and proposed methodology could be extrapolated to other countries. We posit that our methodology could be applicable in different national contexts, although evidence to confirm its effectiveness in these settings is currently lacking. We anticipate that specific effects may vary due to diverse cultural, economic, and political environments, yet the overarching approach would remain pertinent.

There is scope for a research agenda that explores the speeches of leaders from countries with substantial global or regional financial influence, such as the United Kingdom, Germany, or China. Such an expansion would enable comparative analysis and facilitate the establishment of theoretical foundations underpinning the interactions between political rhetoric focussed on economics and financial markets.

References

Abolghasemi, Y., & Dimitrov, S. (2020). Determining the causality between U.S. presidential prediction markets and global financial markets. *International Journal of Finance & Economics, 26*, 4534–4556.

Adamopoulos, P., Ghose, A., & Todri, V. (2018). The impact of user personality traits on word of mouth: text-mining social media platforms. *Information Systems Research, 29*(3), 612–640.

Ajjoub, C., Walker, T., & Zhao, Y. (2020). Social media posts and stock returns: The Trump factor. *International Journal of Managerial Finance, 17*(2), 185–213.

Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. M., & Ortega, F. (2016). A multi-stage method for content classification and opinion mining on weblog comments. *Annals of Operations Research, 236*(1), 197–213.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics, 131*(4), 1593–1636.

Baron, A., Rayson, P., & Archer, D. (2009). Word frequency and key word statistics in corpus linguistics. *Anglistik, 20*(1), 41–67.

Bentes, S. R., Menezes, R., & Mendes, D. A. (2008). Long memory and volatility clustering: Is the empirical evidence consistent across stock markets? *Physica A: Statistical Mechanics and its Applications, 387*(15), 3826–3830.

Bishop, M. (2009). *Essential economics: an A to Z guide,* (Vol. 22). Wiley.

Born, B., Ehrmann, M., & Fratzscher, M. (2014). Central Bank Communication on Financial Stability. *The Economic Journal, 124*(577), 701–734.

Brans, H., & Scholtens, B. (2020). Under his thumb the effect of president Donald Trump's Twitter messages on the US stock market. *PLOS ONE, 15*(3), 1–11.

Cinelli, M., Ficcadenti, V., & Riccioni, J. (2021). The interconnectedness of the economic content in the speeches of the US Presidents. *Annals of Operations Research, 299*, 593–615. https://doi.org/10.1007/s10479-019-03372-2

Consoli, S., Barbaglia, L., & Manzan, S. (2020). Fine-grained, aspect-based semantic sentiment analysis within the economic and financial domains. In 2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI), pp. 52–61. https://doi.org/10.1109/CogMI50398.2020.00017.

Dickinson, M. J., & Lebo, M. J. (2007). Reexamining the growth of the institutional presidency, 1940–2000. *The Journal of Politics, 69*(1), 206–219.

Eroukhmanoff, C. (2018). It's not a Muslim ban! Indirect speech acts and the securitisation of Islam in the United States post-9/11. *Global Discourse, 8*(1), 5–25.

Ficcadenti, V., Cerqueti, R., & Ausloos, M. (2019). A joint text mining-rank size investigation of the rhetoric structures of the US Presidents' speeches. *Expert Systems with Applications, 123*, 127–142.

Ficcadenti, V., Cerqueti, R., Ausloos, M., & Dhesi, G. (2020). Words ranking and hirsch index for identifying the core of the hapaxes in political texts. *Journal of Informetrics, 14*(3), 101054.

Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management, 23*(3), 157–214.

Fung, G. P. C., Yu, J. X., & Lam, W. (2002). News sensitive stock trend prediction. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 481–493. Springer.

Gibbons, J. D. (1993). Nonparametric measures of association. Number 91. Sage.

Goel, S., & Chengalur-Smith, I. N. (2010). Metrics for characterizing the form of security policies. *The Journal of Strategic Information Systems, 19*(4), 281–295.

Golez, B., & Koudijs, P. (2018). Four centuries of return predictability. *Journal of Financial Economics, 127*(2), 248–263.

Gray, R. M. (2011). *Entropy and information theory*. Springer Science & Business Media.

Gupta, A., Dengre, V., Kheruwala, H. A., & Shah, M. (2020). Comprehensive review of text-mining applications in finance. *Financial Innovation, 6*(39), 1–25.

Hayo, B., Kutan, A., & Neuenkirch, M. (2012). Federal Reserve communications and emerging equity markets. *Southern Economic Journal, 78*, 1041–1056.

Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T. J., & Zou, J. (2023). A visual–language foundation model for pathology image analysis using medical twitter. *Nature Medicine*, 1–10.

Kalkur, T., & Rao, A. (2017). Bayes estimator for coefficient of variation and inverse coefficient of variation for the normal distribution. *International Journal of Statistics and Systems, 12*(4), 721–732.

Kalyani, J., Bharathi, H. N., & Jyothi, R. (2016). Stock trend prediction using news sentiment analysis. *International Journal of Computer Science & Information Technology, 8*(3), 67–76.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*(1/2), 81–93.

Kiessling, T., Martin, T. M., & Yasar, B. (2017). The power of signaling: presidential leadership and rhetoric over 20 years. *Leadership & Organization Development Journal, 38*(5), 662–678.

Kräussl, R., Lucas, A., Rijsbergen, D. R., van der Sluis, P. J., & Vrugt, E. B. (2014). Washington meets Wall Street: A closer examination of the presidential cycle puzzle. *Journal of International Money and Finance, 43*, 50–69.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association, 53*(284), 814–861.

Kufenko, V., & Geiger, N. (2016). Business cycles in the economy and in economics: An econometric analysis. *Scientometrics, 107*(1), 43–69.

Kuroda, H. (2017). Theory on Financial Markets and Central Banks. In *Spring Annual Meeting of the Japan Society of Monetary Economics*. Tokyo.

Laih, Y. W. (2014). Measuring rank correlation coefficients between financial time series: A garch-copula based sequence alignment algorithm. *European Journal of Operational Research, 232*(2), 375–382.

Largeron, C., Moulin, C., & Géry, M. (2011). Entropy based feature selection for text categorization. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, pp. 924–928. ACM.

Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D., & Allan, J. (2000). Language models for financial news recommendation. In *Proceedings of the ninth international conference on Information and knowledge management*, pp. 389–396. ACM.

Leamer, E. E., & Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data* (Vol. 53). Wiley.

Lee, J. C., & Myers, M. D. (2004). Dominant actors, political agendas, and strategic shifts over time: a critical ethnography of an enterprise systems implementation. *The Journal of Strategic Information Systems, 13*(4), 355–374.

Li, J., & Born, J. A. (2006). Presidential election uncertainty and common stock returns in the United States. *The Journal of Financial Research, 29*, 609–622.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue, 16*(3), 31–57.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35–65.

Makri, C., & Neely, A. (2021). Grounded theory: A guide for exploratory studies in management research. *International Journal of Qualitative Methods, 20*, 16094069211013654.

Maligkris, A. (2017). Political speeches and stock market outcomes. In *30th Australasian Finance and Banking Conference*.

Marinč, M., Massoud, N., Ichev, R., & Valentinčič, A. (2021). Presidential candidates linguistic tone: The impact on the financial markets. *Economics Letters, 204*, 109876.

Mata, F. J., & Fuerst, W. L. (1997). Information systems management issues in central America: A multinational and comparative study. *The Journal of Strategic Information Systems, 6*(3), 173–202.

Melucci, M. (2007). On rank correlation in information retrieval evaluation. In *ACM SIGIR Forum*, vol. 41, pp. 18–33. ACM.

Mendes, R., Paiva, A., Peruchi, R. S., Balestrassi, P. P., Leme, R. C., & Silva, M. (2016). Multiobjective portfolio optimization of Arma-Garch time series based on experimental designs. *Computers & Operations Research, 66*, 434–444.

Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications, 41*(16), 7653–7670.

Pak, A. & Paroubek, P. (2011). Twitter for sentiment analysis: When language resources are not available. In *Database and Expert Systems Applications (DEXA), 2011 22nd International Workshop on*, pp. 111–115. IEEE.

Patton, A. J., & Weller, B. M. (2020). What you see is not what you get: The costs of trading market anomalies. *Journal of Financial Economics, 137*(2), 515–549.

Peters, G. & Wooley, J. T. (2019). The state of the union, background and reference table. The American Presidency Project. Online; Accessed on 18/07/2022.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports, 3*, 1684.

Qiu, X. Y., Jiang, S., and Deng, K. (2013). Automatic assessment of information disclosure quality in Chinese annual reports. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pp. 288–298. Springer.

Renault, T. (2020). Sentiment analysis and machine learning in finance: A comparison of methods and models on one million messages. *Digital Finance, 2*(1), 1–13.

Rosso, O. A., Craig, H., & Moscato, P. (2009). Shakespeare and other English renaissance authors as characterized by information theory complexity quantifiers. *Physica A: Statistical Mechanics and its Applications, 388*(6), 916–926.

Rutledge, P. E., & Larsen Price, H. A. (2014). The president as agenda setter-in-chief: The dynamics of congressional and presidential agenda setting. *Policy Studies Journal, 42*(3), 443–464.

Sagheer, D., & Sukkar, F. (2019). Text template mining using named entity recognition. *International Journal of Computer Applications, 1982*(46), 34–40.

Santa-Clara, P., & Valkanov, R. (2003). The presidential puzzle: Political cycles and the stock market. *The Journal of Finance, 58*(5), 1841–1872.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students, Pearson education*.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems (TOIS), 27*(2), 12.

Shaikh, I. (2019). The U.S. presidential election 2012/2016 and investors' sentiment: The case of CBOE market volatility index. *SAGE Open, 9*(3), 2158244019864175.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*(1), 3–55.

Shynkevich, Y., McGinnity, T. M., Coleman, S., & Belatreche, A. (2015). Predicting stock price movements based on different categories of news articles. In *Computational Intelligence, 2015 IEEE Symposium Series on*, pp. 703–710. IEEE.

S&P Dow Jones Indices. (2022). S&p500®. https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview. Online; Accessed on 18/07/2022.

Teevan, J., Dumais, S. T., & Horvitz, E. (2018). Personalizing search via automated analysis of interests and activities. In *ACM SIGIR Forum*, vol. 51, pp. 10–17. ACM.

Tessmer, A. C., Shaw, M. J., & Gentry, J. A. (1993). Inductive learning for international financial analysis: A layered approach. *Journal of Management Information Systems, 9*(4), 17–36.

Tsai, M. F., & Wang, C. J. (2017). On the risk prediction and analysis of soft information in finance reports. *European Journal of Operational Research, 257*(1), 243–250.

Wei, C., Hu, P., & Dong, Y. (2002). Managing document categories in e-commerce environments: an evolution-based approach. *European Journal of Information Systems, 11*(3), 208–222.

Wei, Y. M., Mi, Z. F., & Huang, Z. (2015). Climate policy modeling: An online SCI-E and SSCI based literature review. *Omega, 57*, 70–84.

Zhou, R., Cai, R., & Tong, G. (2013). Applications of entropy in finance: A review. *Entropy, 15*(11), 4909–4931.