Mutual funds in North and South America: Relationship between News and Mutual funds Returns

CERGE-EI

Raul Enrique Zavala Sanchez

July 2024

Abstract

This paper investigates the impact of news sentiment on mutual fund returns across several countries in the Americas, examining both developed and developing markets. By leveraging advanced Natural Language Processing (NLP) and machine learning techniques, the study integrates news sentiment into asset pricing models, thereby enhancing their precision. A Vector Error Correction Model (VECM) is used to analyze the complex interactions between news sentiment and mutual fund returns, addressing potential endogeneity and capturing the dynamic relationships over time. Though variations in the effects of news across different countries were expected, developed markets exhibit more consistent and synchronized reactions. It is observed that the effect of news sentiment on the first lag is generally lower than that of their respective stock market indices for both Canada and the USA. In contrast, mutual funds in developing countries such as Brazil, Argentina, and Chile show greater divergence in their reactions, both among themselves and relative to their market indices. Notably, Mexico exhibits characteristics of both developed and developing markets, with early responsiveness to news similar to the USA but with larger coefficients. This study offers valuable insights into regional differences in market behavior, emphasizing the need to account for these variations to improve the accuracy and effectiveness of financial analyses and investment strategies.

Keywords: News Sentiment, Mutual Fund Returns, Decision-Making, Asset Pricing Models, Natural Language Processing.

Abstrakt

Tento článek zkoumá dopad nálady v souvislosti se zprávami na výnosy podílových fondů v několika zemích Severní a Jižní Ameriky, a to jak na rozvinutých, tak na rozvojových trzích. S využitím pokročilých technik zpracování přirozeného jazyka (NLP) a strojového učení studie začleňuje sentiment zpráv do modelů oceňování aktiv, čímž zvyšuje jejich přesnost. K analýze komplexních interakcí mezi zpravodajským sentimentem a výnosy podílových fondů se používá model vektorové korekce chyb (VECM), který řeší potenciální endogenitu a zachycuje dynamické vztahy v čase. Ačkoli se očekávaly rozdíly v účincích zpráv v různých zemích, rozvinuté trhy vykazují konzistentnější a synchronizovanější reakce. Ukazuje se, že vliv zpravodajského sentimentu na první zpoždění je obecně nižší než vliv jejich příslušných burzovních indexů jak pro Kanadu, tak pro USA. Naproti tomu podílové fondy v rozvojových zemích, jako jsou Brazílie, Argentina a Chile, vykazují větší divergenci ve svých reakcích, a to jak mezi sebou, tak ve vztahu k indexům svých trhů. Pozoruhodné je, že Mexiko vykazuje charakteristiky jak rozvinutých, tak rozvojových trhů, přičemž jeho včasná reakce na zprávy je podobná jako v USA, ale s většími koeficienty. Tato studie nabízí cenné poznatky o regionálních rozdílech v chování trhů a zdůrazňuje potřebu zohlednit tyto rozdíly, aby se zlepšila přesnost a účinnost finančních analýz a investičních strategií.

Dedication

This thesis is dedicated to my grandparents, whose unwavering support has been a constant source of strength.

To my mother, whose endless encouragement has always lifted my spirits.

And to my father, whose memory continues to inspire me every day. You are forever in my thoughts.

Acknowledgements

I would like to express my deepest gratitude to my professors for their invaluable assistance, time, and advice throughout this process. Your guidance and support have been instrumental in completing this thesis. To my supervisor, Professor Slavik Ctirad, for his unwavering support,

Contents

Introduction

The interplay between news and market dynamics has garnered significant attention in both academic research and practical financial analyses. Understanding how news relates to stock market movements has been a crucial question for asset pricing and has been developed over many years. Since the advent of behavioral economics and decision theory, researchers have increasingly focused on how psychological factors and investor sentiment influence market behavior.

This shift has highlighted the importance of news not only as a source of information but also as a driver of market sentiment and investor behavior. Studies have shown that news can affect market prices by reducing information asymmetry and influencing investor perceptions and actions. As a result, the relationship between news and market movements has become a vital area of investigation, with implications for financial theory, market regulation, and investment strategies.

The analysis of news as a factor that predicts returns has historically been challenging, but recent advancements in technology, particularly machine learning and Natural Language Processing (NLP), have begun to overcome these obstacles. Implementing these new methods can significantly enhance the accuracy of asset pricing models by incorporating news as an additional factor. Machine learning and NLP techniques can extract critical information from news articles, such as the sentiment of the news and its relevance to particular assets.

This technological revolution has led to a surge in research on news sentiment in recent years, demonstrating its potential to provide deeper insights into market behavior and investor psychology. By leveraging these advanced tools, researchers can now analyze vast amounts of textual data more efficiently and effectively and can identify patterns and sentiments that were previously difficult to quantify. This has opened new avenues for understanding how news impacts financial markets, enabling the construction of more sophisticated and responsive asset pricing models.

However, extensive recent literature on news effects on stock market movements has largely overlooked comparative analysis of these effects across different regions and countries. Most studies have focused on single-country analyses and offer limited insights into how news impacts financial markets in varying economic and regulatory environments. Comparative analyses can shed light on whether the mechanisms through which news affects market behavior are universal, or if they vary significantly due to factors such as market maturity, investor behavior, information dissemination efficiency, or even culture.

This paper aims to contribute to and expand the existing literature by conducting a comparative analysis of mutual funds across several countries in the Americas. The focus on mutual funds is because mutual funds hold diversified portfolios, which implies that the reaction of a mutual fund to news sentiment can reflect a broader market response rather than the reaction of a single asset or sector. This provides a more comprehensive view of how news affects various components of the market. Additionally, mutual funds attract a diverse group of investors, ranging from retail investors to institutional investors, resulting in varied reactions to news. This diversity makes mutual funds a compelling channel through which to study collective market sentiment and behavior.

The literature review of this paper explores the evolution of news as a factor influencing market movements, tracing its development alongside the rise of behavioral economics and finance. This section delves into how the incorporation of news sentiment into asset pricing models has advanced and highlights the methodological innovations that have made this possible. The review examines the techniques used to measure news sentiment and the degree to which they are precise. It emphasizes the roles of machine learning and natural language processing (NLP) in enhancing the accuracy of sentiment analysis. Additionally, it discusses the existing literature on the heterogeneity of news effects across different countries, noting the limited comparative analyses conducted thus far.

The methodological framework of this study outlines its data sources, the process it uses to analyze news sentiment, and the rationale behind model selection for the analysis. Initially, the study employs a VAR (Vector Autoregressive) model to investigate the relationship between news sentiment and mutual fund returns. However, discovery of cointegration among the variables necessitated a shift to a VECM (Vector Error Correction Model), which is more effective in addressing potential endogeneity issues. The VECM model allows for the inclusion of lagged values, thereby capturing the dynamic interplay through which past news sentiment impacts current returns and vice versa. This methodology aligns with the work of Calomiris and Mamaysky (2019), who emphasize the bidirectional nature of the relationship between news sentiment and market reactions. Furthermore, the VECM framework enables the study to account for long-term equilibrium relationships while still modeling short-term dynamics, providing a comprehensive view of how news sentiment and market returns interact over time.

Finally, the paper presents the results and conclusions derived from the analysis. The results section offers a detailed examination of how mutual funds in North and South America respond to news sentiment across different countries. It highlights notable differences between developed and developing markets, showing that developed countries exhibit more significant reactions to news than do their developing counterparts. This divergence suggests that news sentiment has a more pronounced effect in developed markets, potentially due to more efficient dissemination and processing of information. The findings also indicate that mutual funds in developed countries tend to align more closely with market indices in their responses to the news, reflecting a more synchronized market reaction. In contrast, mutual funds in developing countries display varied and less consistent responses, suggesting a less homogeneous impact of news sentiment across different sectors and asset classes. The conclusions drawn from these results provide important insights into the varying effects of news sentiment across different regions and highlight the limitations of the current approach used in this paper, underscoring the need to account for regional differences in market reactions for more accurate analyses.

Literature Review

News is widely recognized as a factor that can influence the returns of financial instruments. Costola et al. (2023) explain that news can either offer genuine information about the value of assets or affect investor psychology. In a similar manner, Marty et al. (2020) assert that news decreases information asymmetry and that market prices absorb the influx of news information.

Shiller et al. (1981) observed that financial markets often demonstrate significant volatility and tend to overreact to news information. This observation contrasts sharply with the Efficient Market Hypothesis (EMH), which posits that market prices reflect all available information efficiently. EMH is categorized into three forms: weak, semi-strong, and strong, each addressing the inclusion of non-public information in market prices. Despite this, empirical evidence continues to highlight instances where market behavior diverges from EMH predictions.

In the weak form of EMH, future prices cannot be predicted by analyzing past prices. The semi-strong form assumes that current stock prices adjust rapidly to new public information. The strong form posits that all public and private information is fully reflected in stock prices, making it impossible to consistently achieve excess returns. However, numerous studies, including those conducted by Hamilton (1995); Caporin et al. (2019); Carlini et al. (2020) have shown that markets react to news and especially to news in the areas of macroeconomic, corporate governance, earnings news, politics, and environmental (Costola et al., 2023).

A key metric for gauging the market's response to news lies in the movement of financial instruments, such as stocks and futures, characterized by their returns and volatility. It follows logically that positive news tends to correlate with increased returns, while negative news is associated with lower returns or losses. Additionally, an influx of positive or negative news is often linked to heightened market volatility.

Market Prices and Returns

In the book Asset Pricing, Cochrane (2009) said, "We observed the prices or returns of many assets, we can use the theory positively, to try to understand why prices or returns are what they are," referring to what Asset Pricing Theory is used for. Asset Pricing Theory deals with two main concepts: prices and returns of assets.

These two concepts are massively relevant to the world of finance and, ultimately, to the economy because they deal with investors' decision-making and, ultimately, the allocation of resources in the economy. In order to analyze the returns and prices, many asset pricing models have been developed. One of the first asset pricing models is the Capital Asset Pricing Model (CAPM), stated as follows:

$$R_i = R_f + \beta_i(R_m - R_f) + \alpha_i \tag{1}$$

Perold (2004) claims that the CAPM was the first coherent framework to answer how risk can affect the return of an asset; in the equation of the CAPM, it is shown how the return of an asset R_i can be explained by the risk of the market β_i or in other words systemic risk.

However, the CAPM has been criticized for its assumptions and limitations. For example, it assumes that all investors have the same expectations and access to information, which may not be true in practice; the model also assumes that the portfolios are well diversified, which is a tough goal to achieve in practice. Additionally, the CAPM does not account for other factors affecting returns.

Alternative asset pricing models have been developed to address some of these limitations; one of these models is the Fama-French. Although these models are widely used in asset pricing, the returns in these models are primarily explained by objective information, and this aspect is shared by most of the models; furthermore, all the models face some limitations and criticism, such as CAPM assuming the information is equally available or the ATP model (a model based on macroeconomic factors), which has been criticized of being unprecise.

Studies on how news affects the returns and prices of financial assets began to grow with the rise of behavioral finance. Criticism that EMH received and irrationality findings in the financial markets open the door to studies on how people interpret information (Uhl, 2014).

For instance, Tversky and Kahneman (1981) found that individuals tend to overreact to new information when making probabilistic judgments. Of course, many studies have shown that EMH cannot be rejected (Uhl, 2014), but at least this is consistent with the mentioned by Costola et al. (2023), who noted that news could influence the psychology of investors.

Given that individuals tend to overreact to information and rely on biases and heuristics when making decisions, as highlighted by the development of behavioral economics, many studies have examined the relationship between investor sentiment and asset pricing models. The studies made by Neal and Wheatley (1998); Brown and Cliff (2005) are among those who have explored how investor sentiment influences market dynamics and asset prices.

These studies often focus on the beliefs of investors about the market's current and future state. As Neal and Wheatley (1998) suggest, "Buy when investors are bearish and sell when investors are bullish." The information that shapes or represents these beliefs is frequently derived from news sources. Therefore, it is logical to suppose that news can significantly affect investors' perceptions of the market, ultimately affecting returns and prices.

A clear example of how news can affect returns is the case of GAMESTOP shares. The share prices skyrocketed with a return of around 400% in just a week, which was only explained by the available information in a Reddit forum (Li, 2021). This case shows how news and information, in general, can affect the price of assets and, therefore, the return.

Models that consider news as an explanatory variable of returns are relatively new because subjective information is difficult to treat as input for models. Determining whether a newspaper article is good or bad news for a certain asset is challenging. Furthermore, even if news is classified as good, it is difficult to quantify whether it is better or worse than other good news. Thus, assigning values that will objectively explain the returns of a particular asset is complex and remains a significant challenge.

As early as 2006, Baker and Wurgler noted that it was no longer debatable that investor sentiment, which can be influenced by news, affects stock returns. The real challenge was how to measure and quantify the effects of these sentiments. Fortunately, technology has improved enormously; nowadays, computer power has increased, and technologies like AI and machine learning are on the rise, improving the ability to treat information like news and use it as input for improving asset pricing models.

News Sentiment

Research on the effects of news on financial markets has been increasing. For example, as early as 2009, Fang and Peress started exploring the relationship between media coverage and expected stock returns, finding that firms with low media coverage tend to have higher earnings. Dougal et al. (2012) found that journalist biases, whether bullish or bearish, could affect market reactions.

Another study by Chi et al. (2023) explored using Google Trends to enhance hedging performance. According to the authors, Google Trends can outperform other investment sentiment indices, suggesting that investment decisions can be improved by using publicly available information on the internet, similar to the role of news articles in providing relevant information to investors. Furthermore, Kogan et al. (2023) demonstrated that fraudulent news impacts retail trading and prices, although these effects were later discounted by trading volumes, even counting the effects of legitimate news.

Aforementioned, one of the main challenges in applying the concept of news to financial markets is how to incorporate news as a factor and quantify its effect. With the development of AI and machine learning, a promising method is using news sentiment as a proxy for news. News sentiment refers to the tone or emotion conveyed in news articles about a particular topic. Positive news sentiment is associated with a positive tone, such as optimism or confidence, while negative news sentiment is associated with a negative tone, such as fear or uncertainty.

The study of how an article can be classified as positive or negative has primarily relied on the bag-of-words method. This approach involves counting words that are either positive or negative and assigning values accordingly. Using specialized dictionaries containing words specific to finance and economics has been crucial in this process (Garcia et al., 2023). Then these dictionaries have been used to evaluate and make studies on news articles getting the news sentiment.

Xing et al. (2018) argue that the bag-of-words technique may not be sufficient for a deep and comprehensive understanding of a text, as it restricts the focus to a small segment of highly structured text. In contrast, Wan et al. (2021) highlight that advancements in natural language processing (NLP) techniques have enabled the algorithmic analysis of large-scale unstructured data, such as text data in financial news articles or research reports, which were previously challenging to process numerically.

Supporting this critique, Gentzkow et al. (2019) noted the limited effectiveness of traditional methods compared to the new machine learning techniques. While not entirely dismissing machine learning approaches, Garcia et al. (2023) proposed the creation of specialized dictionaries based on NLP to enhance the analysis.

Early NLP techniques for financial forecasting primarily employed the bag-of-words approach. This method, as previously stated, involves identifying a set of words and their frequency of occurrence. A significant issue of this approach is that it does not consider the position of the words within the text, leading to potential misinterpretations (Xing et al., 2018). For instance, phrases like "Amazon surpasses Microsoft in profits" and "Microsoft surpasses Amazon in profits" could yield the same market reaction according to the model analysis, despite their different implications. Additionally, equivalent phrases such as "the stock market is bullish" and "the NYSE is at maximums" might not be recognized as conveying the same sentiment. These limitations highlight the need for more sophisticated NLP techniques to capture the nuances in financial texts accurately.

A significant challenge for NLP techniques has been accurately understanding the context of words within a text. To address this, the introduction of neural networks marked a substantial advancement in the computational understanding of text semantics. Hochreiter and Schmidhuber (1997) introduced the Long Short-Term Memory (LSTM) neural network, which enabled better handling of context over long sequences of text. Further developments by Bengio et al. (2003) proposed neural probabilistic language models, which improved the ability to capture the semantic meaning of words and their contextual relationships. These innovations have paved the way for more sophisticated and accurate text analysis in financial forecasting and other applications.

Ultimately, sentiment analysis needs to account for various complex aspects such as subjectivity, sarcasm, personality, and entity recognition. These tasks require a sophisticated understanding of language nuances, which NLP techniques have demonstrated a superior capability to handle compared to traditional methods (Xing et al., 2018). By leveraging advanced NLP models, researchers and analysts can more accurately capture the sentiment behind financial news and its potential impact on the market.

As previously mentioned, News can provide valuable information about asset values or influence investor psychology. Therefore, determining the sentiment of news—whether it is positive or negative—and quantifying this sentiment is crucial. This process allows news to be used as an explanatory variable for financial returns, enhancing our understanding of how news impacts market movements. In this context, NLP techniques become invaluable for performing sentiment analysis. Advanced NLP models like BERT (Bidirectional Encoder Representations from Transformers) and VADER (Valence Aware Dictionary and Sentiment Reasoner) can be used to derive sentiment scores from text. These sentiment scores can then be incorporated into return models. An example of this approach is demonstrated in the model developed by Costola et al. (2023), which effectively uses sentiment analysis to explain the financial returns of the S&P 500 during the COVID-19 pandemic.

In their paper, Costola et al. (2023) utilize NLP to construct news sentiment. One of their most important findings is that sentiment analysis is highly associated with volatility. They found that sentiment is statistically significant and related to returns. Interestingly, their results indicate that negative news has a more substantial impact than positive news. Specifically, a decrease in negative news was more beneficial to market movements than an equivalent decrease in positive news was harmful. However, the sentiment did not show any statistically significant relation to trading volume.

Similarly, the study by Allen et al. (2019) based on the stock prices of Dow Jones Industrial Average constituents, demonstrated that daily news sentiment has a significant effect on the returns of Dow Jones companies, even when accounting for other market factors using an expanded version of the Fama-French model. Additionally, the significance of lagged values suggests that information is not immediately reflected in stock prices, highlighting the delayed impact of news on market movements.

The significance of lagged values can persist for extended periods, not only for daily frequencies. For instance, Eachempati and Srivastava (2021) examined the effects of lagged news sentiment on asset pricing in their study. They considered lagged values of up to three months and found that news sentiment significantly influences asset prices with a two-month lag. This indicates that the impact of news sentiment on asset prices may not be immediate and can extend over a period of time, highlighting the importance of incorporating lagged sentiment variables in econometric models to capture the delayed effects of information on financial markets. The significance of lagged values is also confirmed by Hsu et al. (2021) in their research on how news sentiment can affect stock market volatility.

Much of the research related to news sentiment and market movements published in recent years has frequently confirmed that news sentiment can forecast market movements. For example, Shah et al. (2018) reported achieving a 70.59% accuracy rate in predicting short-term stock price trends in the pharmaceutical market based on news sentiment. Similarly, Chowdhury et al. (2014) reported a 67% correlation between news sentiment and market movements. Gupta and Banerjee (2019) have also demonstrated that OPEC news sentiment is related to the oil and gas index and, therefore, the returns of oil stocks, as in Figure 1.

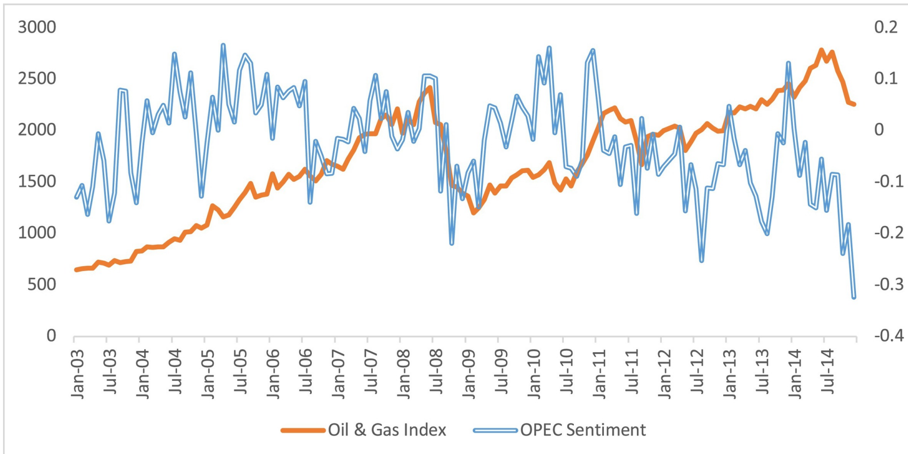


Figure 1: OPEC News sentiment. From Gupta and Banerjee (2019)

While these findings are promising, not all research aligns perfectly with the predictive power of news sentiment. For instance, Smales (2016) found that news sentiment is less accurate than the VIX (Volatility Index) at forecasting future returns. Using a Vector Auto-regression (VAR) model, Smales demonstrated that the VIX is strongly related to implied volatility and stock returns, whereas news sentiment is significantly related to stock returns but not to implied volatility. However, Smales also found that combining both VIX and news sentiment leads to a low-frequency trading strategy that results in consistently profitable trades. This suggests that while news sentiment alone may not be as robust a predictor as the VIX, its combination with other market indicators can enhance trading strategies and improve forecast accuracy.

Nonetheless, plenty of recent studies explore and demonstrate how news sentiment can explain and forecast returns and market movements. The challenge often lies in identifying the most suitable model and NLP technique to establish the relationship between news sentiment and market variables like returns, prices, and volatility.

For example, Fernandez et al. (2021) compared different techniques, including a dictionary with word polarities, a Support Vector Machine (SVM) classifier, a neural network, and a voting classifier that considers each technique. Although the voting classifier outperformed each individual technique, they found that each technique, with its pros and cons, has shown a good level of accuracy, with the dictionary approach having the lowest accuracy (around 61%) because it cannot adapt to texts. In contrast, SVM and neural networks achieved higher accuracy levels of around 73%.

These findings suggest that news sentiment could be used as an important factor to consider when evaluating the performance of any asset. While traditional asset pricing models may not fully account for the impact of news sentiment on returns, incorporating sentiment analysis into investment strategies may help investors identify funds that are well-positioned to benefit from positive news sentiment and avoid those that are more vulnerable to negative sentiment.

Heterogeneity in News Impact Across Stock Markets

Research on news sentiment and its relation to market movements has expanded significantly in recent years, driven by advancements in NLP models. These models have improved in accuracy and are capable of handling large datasets, identifying text sentiment, and managing complex aspects like word context, semantics, and even sarcasm.

However, most of this research has focused on major stock markets, particularly in the USA. For example, Costola et al. (2023) base their study on the S&P 500 index, which comprises the largest 500 companies in the USA, while Allen et al. (2019) focus on the Dow Jones Industrial Average, an index of 30 prominent US companies.

The body of literature on the NASDAQ and NYSE has grown extensively. Conversely, research on the impacts of news sentiment on stock markets in other countries remains limited. This indicates a gap in the literature that needs to be addressed to understand the heterogeneity in news impact across different stock markets globally.

Outside the studies conducted in developed countries—primarily in the USA—the papers on news sentiment studies I have found are in India (Eachempati and Srivastava, 2021), China (Xu et al., 2022; Zong et al., 2022), and a few other countries, but they remain very limited. This disparity is understandable given the relative ease of collecting data from developed countries like the USA and the challenges associated with data collection in developing countries.

This paper focuses on the Americas, and particularly for the Latin American region, literature on news sentiment is even more limited than in the previously mentioned countries. To the best of my knowledge, to date, I have found only a few papers focused on the impact of news sentiment on stock returns in Latin American markets.

Most of these papers are from Mexico, such as the study by Fernandez et al. (2021) that examines the impact of Twitter messages on stock prices, and Brazil, like the study by Januário et al. (2021) which used three different sentiment analysis strategies. It makes sense that these two countries have the most related research in the region, as they have the largest national stock markets in the region by volume. However, research on other important stock markets, such as the Santiago Stock Exchange in Chile or the Buenos Aires Stock Exchange in Argentina, is highly limited. Additionally, there is a lack of studies on other countries in the region, such as Peru, Colombia, and Panama.

Moreover, comparing the impact of news sentiment across different stock markets is also under-researched. With the development of new technologies, much of the research has focused on developing better NLP techniques and models to improve forecasting accuracy. However, this focus may have overlooked the potential variations in how news sentiment affects different countries or regions. It is essential to consider that news might not be received or interpreted uniformly across diverse markets, which could lead to different impacts on market movements.

For example, the study by Fraiberger et al. (2021) found that news sentiment robustly predicts future daily returns of equity. Interestingly, they also discovered a difference between the effects of local and global news. Their findings indicate that while both types of news significantly affect equity returns, the effects of local news were more transitory and smaller compared to global news. Global news effects were not easily reversed in the short run and had a larger impact than local news. They attributed this difference to the influence of global news on foreign investors' perceptions and decisions, which subsequently affects Foreign Direct Investment in the country.

It is worth mentioning that Fraiberger et al. used the bag-of-words method for sentiment analysis. As noted earlier, this method generally has lower accuracy compared to more advanced techniques. Despite this drawback, Fraiberger et al. were able to accurately predict daily returns and demonstrate that the impact of news varies depending on whether the news is local or international.

The distinction between local and international news suggests that the impact of news on stock markets may vary by country or region. This implies that news may not uniformly affect stock markets and that differences in news impact could exist even within the same topic or nature. For example, Kamal and Wohar (2023) found that news about COVID-19 cases had a different impact on the UK stock market compared to news about COVID-19 deaths.

The impact of news may vary significantly across countries, and this diversity in outcomes could be attributed to various factors. Part of the heterogeneity that could occurred across countries may stem from the existence of news leakage. Essentially, this implies that news is known in advance, challenging the notion that news is truly 'news' when it becomes public.

Bernile et al. (2016) suggest that, for instance, in the United States, news agencies are granted pre-release access to government information, subject to embargo agreements that restrict publication until specific conditions are met. This highlights the notion that news may not be as novel or instantaneous as commonly perceived. The study found that during embargo periods, there were disturbances observed in the S&P 500, the U.S. stock market index.

Hence, this demonstrates that there are possibilities that news leakage may affect the stock market. In the context of explaining potential differences across countries in my research, news leakage seems to play a crucial role and may serve as a significant explanatory factor. There is relevant research, such as the study conducted by Akey et al. (2022), which reveals a 15% increase in earnings for individuals engaged in insider trading—those with non-public information, in other words, information known in advance. The research also highlights instances of hacking events wherein hackers engaged in insider trading activities that influenced the stock market prices.

Considering that news may affect the returns differently depending on the country. Li et al. (2018) showed how institutions trade ahead of false news in emerging countries, suggesting institutions in emerging countries know what will be published in advance. The same is suggested by Sun et al. (2021) paper on the role that news and insider trading play in emerging countries, where they found that there is a relation between the coverage and instantaneously of news in the insider trading activity. While a firm's governance can influence the media's predictive power.

Given these insights, it occurs to me that in developed countries, where regulatory frameworks regarding news and insider trading are stricter, the effects of news may be more pronounced because news is novel. Conversely, in developing countries, the effects will be diminished depending on how strict their norms are.

Apart from the issue of potential news leakage, the release of false news can also significantly impact the market. Arcuri et al. (2023) find that fake news can affect the stock market because there will always be investors who cannot differentiate between legitimate news and fake news. Their findings indicate that negative fake news has a short-term negative effect on stock prices, although they could not identify a clear impact of positive or neutral fake news.

Similarly, Karppi and Crawford (2016) mention that fake news about two explosions at the White House caused the S&P 500 to lose around $130 million in market capitalization. Additionally, Kogan et al. (2023) highlighted how fraudulent news impacts retail trading and prices, suggesting broader societal implications. Their study on news manipulation in the US market also raises concerns about similar issues in other countries, where cultural factors and varying proportions of fake to legitimate news could alter the effects of misinformation.

Bias in news reporting is another important consideration that does not necessarily mean fake news or news leakage, as the same story can be portrayed with different tones depending on the source. For instance, news from The Wall Street Journal, which has a conservative slant, can differ significantly from that in The New York Times, which is more liberal. Goldman et al. (2024) found that this politically driven variation in corporate financial news leads to increased abnormal trading volume for firms with strong political affiliations. Their findings suggest that polarized reporting on corporate financial news creates a new source of investor disagreement, driving trading activity among investors.

There are many examples of how news can be biased. For instance, Baloria and Heese (2018) suggest that Fox News displays political bias against Democratic firms, with findings indicating that Democratic-affiliated firms in markets with Fox News channels tend to downplay negative news. Knill et al. (2022) noted that companies with Republican-leaning managers located in areas where Fox was introduced increased their overall investment expenditures and financial leverage.

Likewise, Goldman et al. discusses the findings from Luo, Manconi, and Massa (2020), where they discovered that after News Corporation's acquisition of Dow Jones, stock prices of Republican-affiliated firms became less responsive to Dow Jones Newswires sentiment. This finding, despite not indicating increased bias, suggests a potential influence akin to 'fake news'.

This highlights the nuanced effects of various news on financial markets and how news can vary across different regions of the world. The potential for news manipulation exists, whether through news leakage, which diminishes its novelty, the dissemination of fake news—shown earlier to affect returns and trading volumes, particularly negative fake news—or biases inherent in certain media outlets that favor specific narratives. Furthermore, there are numerous other factors that can influence how news impacts the stock market, including cultural dispositions towards different types of news and many more unforeseen aspects.

As mentioned earlier, the literature on news sentiment and its impact on stock markets outside the USA is quite limited. Consequently, comparative studies examining how news affects different countries are even rarer than country-specific analyses. However, some research provides insight into this area. Notably, the papers by Calomiris and Mamaysky (2019) and Fraiberger et al. (2021) offer valuable perspectives on the development of this field of study.

Calomiris and Mamaysky (2019) claim that their study is the first to analyze country-level stock returns and risks in relation to news. They found notable differences between developed and emerging economies in how news impacts these countries. Their analysis of 51 countries revealed that their news measures, which encompass sentiment, frequency, and entropy, have better predictive power for emerging economies compared to developed ones.

They categorized news into markets (Mkt), governments (Govt), commodities (Comms), corporate governance and structure (Corp), and macroeconomic topics (Macro). For both developed and emerging economies, incorporating text measures significantly improves out-of-sample forecasts relative to a baseline model that excludes these measures.

An important aspect of Calomiris and Mamaysky (2019) work is their use of a dictionary or bag-of-words approach, which, as mentioned earlier, is less accurate and has lower predictive power than other methods. They analyzed available news in English from various countries using a monthly frequency. With the development of new NLP techniques capable of handling news in their native languages, their accuracy could be improved. Additionally, exploring the use of daily frequencies could provide deeper insights.

Fraiberger et al. (2021) build upon the work of Calomiris and Mamaysky (2019), their main focus is comparing local and global (international) news. As mentioned above, their results show that local news has short-run effects compared to global news, indicating that the source can influence the duration of these effects. They analyzed 25 countries, both advanced and emerging economies, similar to Calomiris and Mamaysky, but their emphasis was on the short-term impacts of news. However, like Calomiris and Mamaysky, Fraiberger et al.. also used the bag-of-words approach, which cannot adapt well to the texts that are parsed through this approach, and relied on English news available for each country.

This paper aims to enhance accuracy by utilizing newer NLP techniques that move beyond the traditional bag-of-words approach. By incorporating news in local languages and using daily frequency data, this methodology is expected to offer a more nuanced understanding of the impact of news across different stock markets. Literature has shown that news sentiment, as a proxy for news, can serve as an excellent predictor for financial instruments, often surpassing the accuracy of many traditional models.

Additionally, in line with behavioral finance, it acknowledges the irrationality many investors exhibit due to the psychological influence of news. This approach underscores the interplay between investor psychology and market dynamics, providing deeper insights into how news shapes financial behaviors and outcomes. Furthermore, the possibility that news does not have the same impact everywhere opens avenues for exploring previously unforeseen aspects.

Methodology

This paper aims to establish a relationship between news and returns in the stock markets of North and South America, specifically observing how the different markets in these regions react to news published in their local media. To achieve this, I employ a newer method of news sentiment analysis, advancing beyond previous approaches.

I use a model that incorporates news sentiment as a variable to explain market movements in terms of returns. This research seeks to build upon the work of Calomiris and Mamaysky (2019) and Fraiberger et al. (2021) while also contributing to the limited body of literature on news sentiment analysis in Latin America.

The study focuses on mutual funds from the major stock markets in the Americas, including Argentina, Brazil, Canada, Chile, Mexico, and the USA. These markets were selected due to their size and the availability of robust data. Although other important stock markets, such as those in Colombia and Peru, were considered, their data limitations and the additional resources required for data collection led to their exclusion from this study.

While other papers have used stock market indexes, such as the S&P 500, this paper will be, to the best of my knowledge, one of the first to compare mutual funds' reactions to news across different countries. Mutual funds are particularly interesting because they are considered to hold a diversified portfolio of assets, spreading risk across different sectors and asset classes (Cresson, 2002; Agapova and Kaprielyan, 2023). This diversification means that the reaction of a mutual fund to news sentiment can reflect a broader market reaction rather than the reaction of a single asset or sector, providing a more comprehensive view of how news affects different components of the market. Additionally, mutual funds attract a diverse group of investors, from retail investors to institutional investors, leading to varied reactions to news and making them an intriguing subject for studying collective market sentiment and behavior.

The study primarily uses open-end mutual funds. These funds can be bought and sold at any time during the day, with the transaction price set at the net asset value of a share at the end of the trading day (Elton and Gruber, 2013). This liquidity allows investors to react quickly to new information, potentially amplifying the impact of news sentiment on mutual fund prices. Given their significant assets under management, mutual funds can have a substantial impact on market prices. Understanding how mutual funds react to news sentiment can shed light on broader market movements and trends.

Many mutual funds are benchmarked against market indices, and it has been shown that movements in mutual funds may be present in market indexes too (Goetzmann and Massa, 2003). By studying their reaction to news sentiment, the research can assess how news affects fund performance relative to these benchmarks. Since mutual funds encompass various sectors and asset classes, analyzing their reaction to news sentiment can reveal sector-specific impacts and broader market trends. This analysis can help identify which types of news have the most significant effects on different parts of the market.

Ultimately, this study aims to provide a more nuanced understanding of the impact of news on different stock markets. By integrating newer Natural Language Processing (NLP) techniques, local language news, and daily frequency data, this approach not only enhances the accuracy of predictions but also acknowledges the psychological influences of news on investor behavior. It offers deeper insights into the interplay between investor psychology and market dynamics. The development of this research design aims to contribute significantly to the literature on news sentiment analysis and its effects on financial markets, particularly in the context of the Americas.

Sources

This paper aims to explain the returns of various mutual funds across different countries in terms of news sentiment. While the recent boom in news sentiment research has enriched the literature, it has often lacked comparative analysis across different markets. For this study, the required data includes the returns of mutual funds. Although there is no single source for obtaining the returns of all assets, many financial websites provide daily prices for mutual funds. Well-known sources include Bloomberg and Yahoo Finance. For this paper, I have sourced mutual fund data from Yahoo Finance, which offers available data for the stock markets of Argentina, Brazil, Canada, Chile, Mexico, and the USA—the countries under study.

I focused on open-end mutual funds, which can be bought anytime. While the number of mutual funds in the US and Canada is extensive, that is not the case for other countries. The numbers are way lower in the case of other countries for that reason I have limited the list of mutual funds for the US and Canada to just a few just to equate to the number of mutual funds from the rest of the countries. The list of mutual funds included in this study is presented in the following table.

Argentina	Brazil	Canada	Chile	Mexico	USA
COME	ALZR11	0P0000P0K1	CFICI81I-E	BLKPATA	FSREX
MTR	0P0001L2GV	0P0000TF96	CFIFALCGLA	EQUITYA	HHDFX
VALO	0P0000VV0T	0P0001CP88	CFIFALCTAC	FT-BOND	MRDAX
MACRO	OUFF11	0P0001HQM7	CFIFYDIA-E	HSBCMEXBFV	PJEAX
IRSA	GOLD11	0P0001IPFZ	CFIFYNRF2A	MONEXCRBEC	VDEQX
	0P000163OY	0P0001N8Q5	CFIHMCRGPB	NTEIPC+APIF	VDIGX
	NVHO11	0P000074QW	CFMESGIPSA	PORTMANA	VGSTX
	0P0001AJBM	0P0000770X	CFIMBIDL-D	REGIOBA	VIPIX
	0P0001KOYM	0P0000820B	CFINRENTAS	SAM-APA	VSCGX
	0P0000ZU26	0P0000820E	CFIMDCHD	SURAGRA	VWEAX

Table 1: Selected Mutual Funds Tickers, Author

I have selected a diverse set of mutual funds, including equity mutual funds, fixed-income mutual funds, and balanced mutual funds. This selection ensures a comprehensive analysis across different types of financial instruments. After obtaining the price data, the calculation of returns becomes a straightforward process. This can be achieved by choosing either simple returns or logarithmic returns.

$$SimpleReturn = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{2}$$

$$logReturn = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{3}$$

Simple returns are calculated by dividing the difference between the current price and the previous price by the previous price. In contrast, logarithmic returns, which can often provide a more nuanced view of price changes, are calculated by taking the natural logarithm of the ratio of the current price to the previous price. The choice between these two methods depends on the specific characteristics and requirements of the analysis. For this study, logarithmic returns appear to be the most appropriate measure to understand the impact on mutual fund performance.

Simple returns offer an intuitive and straightforward measure of performance, as noted by Panna (2017); if it is feasible to use the simplest option, it is often preferable due to ease of interpretation. However, logarithmic returns have advantageous statistical properties that make them suitable for many financial time series studies. For instance, log returns are additive over time, which simplifies the analysis of compounded returns and the modeling of financial data. Despite these benefits, a drawback of log returns is that they can introduce complications to the risk measures due to missing data or the nature of the analysis.

In this study, incorporating logarithmic returns allows for a robust examination of how news sentiment influences mutual fund performance. This approach ensures that the analysis captures the detailed statistical insights provided by logarithmic returns, leading to a comprehensive understanding of the impact on mutual fund performance.

While other studies have used sources of news sentiment such as Thomson Reuters News Analytics (TRNA) (Sinha, 2016; Heston and Sinha, 2017), I have opted to conduct my own sentiment analysis. Previous sources have primarily focused on English-language news, which overlooks the rich nuances found in news reported in their original languages. This limitation can lead to a less accurate representation of sentiment, particularly in the context of diverse linguistic regions.

Therefore, the next crucial step is to gather news data, which will be used for sentiment analysis. Given the daily frequency of the study, a large volume of news articles is necessary. Collecting daily news from multiple newspapers and media outlets over several years can be time-consuming. Fortunately, modern technology allows access to vast amounts of constantly updated data through APIs (Aitamurto and Lewis, 2013).

Each country in this study presents unique challenges in finding APIs that provide relevant information. The easiest case is the USA, where free APIs, such as the New York Times API, allow access to news dating back to 1851 according to their website. However, this level of accessibility is not common for the other countries in this study. Although it is possible to obtain news data for countries like Argentina, Brazil, Canada, Chile, and Mexico, the process can be significantly more time-consuming without the use of APIs.

I have obtained daily news from recent years using APIs. While news data for the USA is available dating back to almost 200 years ago, I decided to focus on a period of around three years to maintain consistency with the data availability for the majority of the studied countries. This timeframe ensures a sufficient amount of data for robust analysis while also aligning with the accessible news data for Argentina, Brazil, Canada, Chile, and Mexico.

Using APIs to search for news allows for the exclusion of articles that are unlikely to impact the stock market, for example news about celebrity concerts or other entertainment events. This filtering capability is crucial for focusing on news that is pertinent to financial markets and investor sentiment. By honing in on relevant economic, political, and corporate news, the analysis can more accurately assess the impact of significant information on mutual fund returns. Below is a list of some of the keywords used during the search:

Keywords:

- Stock Exchange
- Economy
- Finance
- Recession
- Market Basket
- Exports
- Imports
- Inflation
- Balance of Payments
- commodities

In addition to filtering by topics and keywords, the API search can be refined using several other criteria to ensure the relevance and precision of the news data. For example, news can be filtered by the country of origin, such as ar (Argentina), br (Brazil), ca (Canada), cl (Chile), mx (Mexico), and us (United States). Also, language filters can be applied to include only news articles in English (en), Spanish (es), or Portuguese (pt). News categories, such as business and politics, can also be specified to further narrow down the search. These types of specifications enable the collection of news from local media sources in each studied country, ensuring that the data is both relevant and comprehensive for the analysis.

This targeted approach not only improves the efficiency of data collection but also enhances the quality of the sentiment analysis. It ensures that the news data used in the study is directly related to market movements and investor behavior, thereby providing more meaningful insights into how different types of news influence mutual fund performance. Additionally, this method helps in managing the volume of data, making the analysis more manageable and precise.

Furthermore, using APIs with the specified criteria, I was able to retrieve a total of 44,142 news articles from various sources. This large dataset ensures a comprehensive coverage of relevant news that could impact mutual fund returns. The articles were filtered based on country, language, and news category, ensuring that only pertinent economic, political, and corporate news was included in the analysis (Hamilton, 1995; Caporin et al., 2019; Carlini et al., 2020). The table below, presents the distribution of the number of articles collected from each country.

Country	Number of Articles
Argentina	1358
Brazil	5645
Canada	3842
Chile	4677
Mexico	12734
United States	15886
Total	44142

Table 2: News Articles by Country

Obtaining the news and mutual fund returns was not the only data necessary for this analysis. Since the aim is to examine the effect of news on returns, it is important to acknowledge that not all variations in returns can be explained solely by news sentiment. Therefore, I have also gathered additional variables that have been shown in previous studies to be related to returns. These variables will complement the model and provide a more comprehensive analysis, which will be explained in further sections.

Incorporating these supplementary variables helps to control for other factors that may influence mutual fund returns. This approach ensures that the impact attributed to news sentiment is not confounded by other market dynamics. By including a range of economic and financial indicators, the model aims to isolate the specific contribution of news sentiment to mutual fund performance, thereby providing a more robust and accurate understanding of the relationship between news and market returns. These variables include, local stock market indices, and government bond yields. By incorporating these factors, the model can better account for other influences on mutual fund performance, thereby isolating the effect of news sentiment more accurately.

One of the reasons for considering these factors is the difficulty in obtaining other variables, such as GDP, unemployment rates, and even some Fama-French factors, at a daily frequency. Local stock market indices, such as the S&P 500 in the USA, the IBOVESPA in Brazil, and the IPC in Mexico, serve as benchmarks for overall market performance. These indices offer essential context for evaluating mutual fund returns against broader market trends, allowing for better assessment of the impact of news sentiment For instance, by examining whether mutual funds respond to news in a manner similar to their local stock markets, we can determine if their reactions are of the same magnitude and aligned with broader market movements.

Government bond yields are included as they reflect the risk-free rate in each country, influencing investment decisions and the cost of capital. These last two factors have been included in various asset pricing models, such as the Capital Asset Pricing Model (CAPM) or the Arbitrage Pricing Theory (APT) (Elbannan, 2015; Siahaan et al., 2018). In essence, the data collection process for this study has been carefully designed to include all relevant variables that might influence mutual fund returns. By leveraging APIs to gather a wide array of news articles and incorporating supplementary economic and financial indicators, this approach aims to provide a comprehensive analysis.

News Sentiment Analysis

In the previous section, I described how news articles were obtained using APIs, allowing the collection of vast amounts of news relevant to the stock market and mutual fund returns. However, the most challenging aspect is determining the news sentiment. Earlier in this paper, I discussed how traditional models often neglected to incorporate news sentiment due to the complexities involved in analyzing various textual aspects, such as context and tone. It is only in recent times, with the development of machine learning techniques like Natural Language Processing (NLP), that we have been able to enhance the accuracy of news sentiment analysis.

The next step in analyzing the news that was obtained is to define a model to check if the news is optimistic (positive news sentiment) or pessimistic (negative news sentiment); there are many approaches, for example, Lasso-Ridge (Lin, 2021), where regression is done, and words are weighted according to specific rules. Alternatively, the bag-of-words approach, as seen before, while simpler to implement, tends to be less accurate as it does not account for context and semantic nuances in the text.

The choice of model is crucial because it can introduce biases into the analysis. For example, the phrase "... like Apple" differs significantly from "I like Apple" when referring to the company. The former involves a comparison and its sentiment depends on the preceding context, which could be either positive or negative. In contrast, the latter conveys a clear positive sentiment. In the bag-of-words approach, both sentences would be treated similarly, thereby ignoring the contextual differences and potentially misinterpreting the sentiment.

Many studies on the stock market have employed VADER (Valence Aware Dictionary and Sentiment Reasoner) for news sentiment analysis (Singh et al., 2022). VADER is one of the most popular models in this field due to its ease of implementation and the fact that it does not require training data. However, it has certain limitations, such as difficulties with long texts and a heavy reliance on its lexicon. VADER has proven effective for analyzing short messages on social media platforms like X (formerly Twitter). Given that this paper considers longer and more complex news articles, I have chosen to use a different NLP model that is also popular in this research field, which offer better analysis of news sentiments for this study.

The model I propose for this research is BERT (Bidirectional Encoder Representations from Transformers), an advanced NLP model built on the Transformer architecture. Unlike traditional language models that consider only the left or right context, BERT comprehensively understands the context of words in a sentence by considering both sides. Notably, BERT has been employed by Costola et al. (2023) in their research on the reaction of the US stock market to Covid-19 news. In the field of Finance, BERT has found application in creating sentiment-based portfolios, as demonstrated by the work of Hung et al. (2024), where BERT was used to measure news sentiment and formulate investment portfolios.

The primary objective of my research is to compare the effects of news across various countries and uncover heterogeneity and differences arising from regional distinctions; my focus lies in exploring these variations rather than emphasizing the optimal method for analyzing news sentiment. Moreover, the widespread use of the BERT model in numerous finance papers underscores its effectiveness in financial research (Xing et al., 2020; Kabbani and Duman, 2022). Given its established track record, I deem the BERT model sufficiently robust for application in my research.

Another advantage of employing the BERT model is its open-source nature and extensive pre-training on millions of data. This eliminates the need to start from scratch, requiring only fine-tuning for specific tasks (Costola et al., 2023). In the context of this research, the specific task at hand involves conducting sentiment analysis on news articles that could potentially impact the stock market. This entails focusing on news explicitly related to the stock market, as well as broader topics such as the economy and politics, which are known to influence market behaviors.

Furthermore, BERT has demonstrated efficacy and has versions available for various languages, including English, Spanish, French, Portuguese, German, Italian, and more. This broad language support ensures the feasibility of researching the effects of news in the Americas across the continent's primary languages. For instance, BERT can be used to analyze English-language news in Canada and the USA, Spanish-language news in Argentina, Chile, and Mexico, and Portuguese-language news in Brazil. This multilingual capability of BERT is particularly advantageous for this research as it allows for a comprehensive analysis of news sentiment across different linguistic and cultural contexts.

To fine-tune BERT for sentiment analysis in the context of this research, which is finance-related, the model can be specifically adapted using words identified as positive and negative within the field of finance. Notably, a valuable resource for such fine-tuning is the dataset provided in the study by Garcia et al. (2023), where the authors shared their relevant data. Considering the objectives of this study, the most suitable version of BERT is FinBERT. FinBERT has been pre-trained on 4,840 sentences from financial news (Malo et al., 2014). FinBERT adopts the same underlying approach as BERT but is specifically tailored for the finance domain. This NLP model stands out as the most accurate choice for the objectives of this study.

Figure 2 shows the process of handling the obtained news. After acquiring the text from news sources via APIs, the next steps involve tokenization. Where tools like NLTK tokenizers are employed to break down the text into shorter segments. Each segment or sentence begins and ends with special tokens: [CLS] ("Classification") and [SEP] ("Separation"). These segments are then further divided into individual words, each assigned a unique numerical ID representing that specific word. These numerical IDs serve as input for the BERT model, which has been fine-tuned to analyze sentiment. Ultimately, a dense neural network layer performs the final classification, resulting in the sentiment prediction. The model processes the input and outputs sentiment predictions, indicating whether the sentiment is positive or negative, along with a corresponding score.

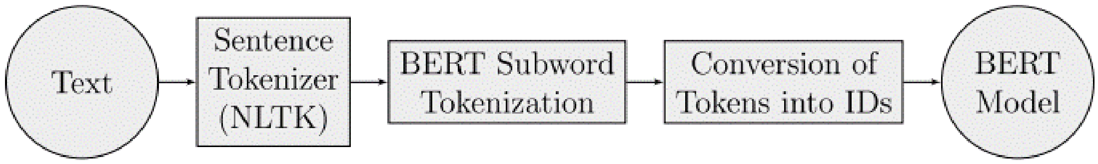


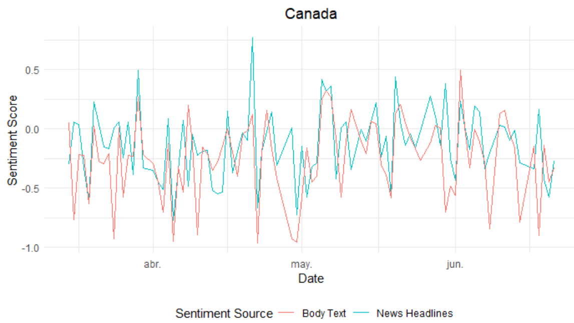
Figure 2: The Preprocessing steps using the Natural Language Toolkit. From Costola et al. (2023)

The model classifies all tokenized sentences as positive, negative, or neutral, along with their respective probabilities. The sentiment score of the text can be calculated as follows: $SentimentScore = PositiveProbability - NegativeProbability$ The scores of sentiment analysis range between -1 and 1, with 0 interpreted as neutral. A score closer to 0 indicates a more neutral sentiment, a negative score indicates pessimistic news and a positive score indicates optimistic news. This sentiment score is particularly suitable for implementation in our final econometric model. By quantifying the sentiment of news articles, I can incorporate these scores as explanatory variables in the final model.

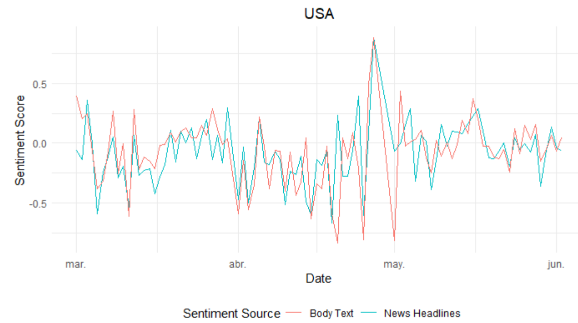
The news data obtained includes various fields of information such as publication date, author, and more. However, the most important and useful fields for this study are the title and body text. The title corresponds to the news headlines, while the body text encompasses the entire article with all its details.

These fields are particularly interesting because, as Yoon et al. (2021) explain, news stories can sometimes be misleading due to incongruence between headlines and the body text. To address this, I conducted sentiment analysis on both the news headlines and the body text. Furthermore, following the classification approach of Calomiris and Mamaysky (2019) and Fraiberger et al. (2021), I categorized the news by developed countries (Canada and the USA) and developing countries (Argentina, Brazil, Chile, Mexico). This dual classification allows for a more nuanced understanding of news sentiment across different economic contexts.

Figures 3 and 4 illustrate the sentiment trends over time for the most recent days, with Figure 3 focusing on developed countries (Canada and the USA) and Figure 4 on developing countries (Argentina, Brazil, Chile, and Mexico). The selection of recent days was intentional to ensure that the dataset includes contemporaneous information across all countries. This approach mitigates discrepancies in the data collection periods, as the span of data varies slightly between countries.



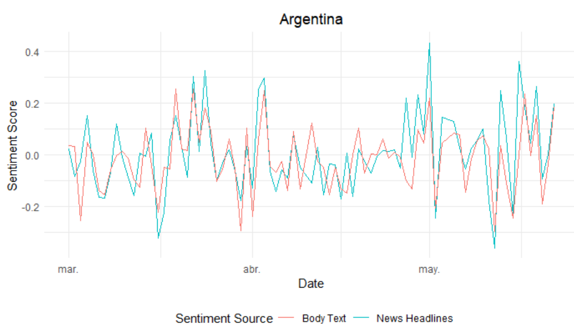
(a) Canada



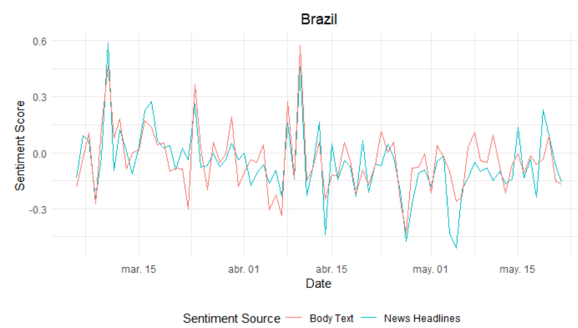
(b) USA

Figure 3: Developed Countries

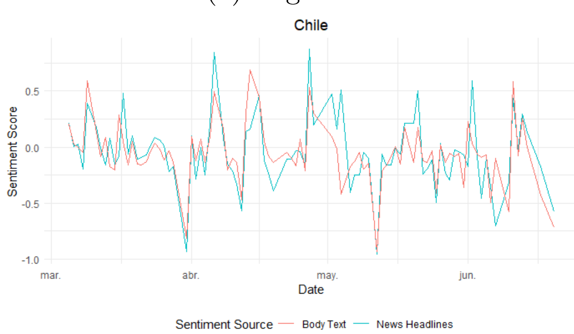
It is evident that there are differences between the sentiment expressed in news headlines and the body text. However, for the majority of the time span depicted, both components tend to follow similar trends. When the sentiment of the headlines is positive, the sentiment of the body text is also positive, often with close or similar scores. This consistency suggests a general alignment between the headline and the detailed content of the articles.



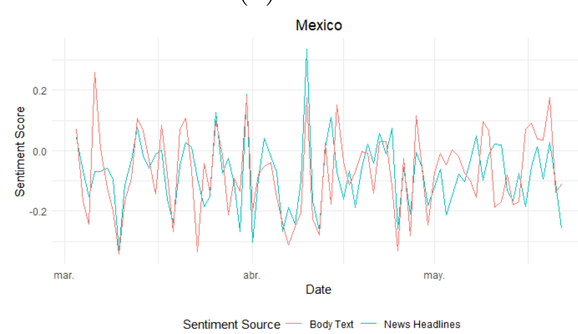
(a) Argentina



(b) Brazil



(c) Chile



(d) Mexico

Figure 4: Developing Economies

Nonetheless, there are instances where the sentiment scores of the headlines and the body text diverge. Although these discrepancies are relatively infrequent, they highlight the importance of analyzing both headlines and body text to capture a comprehensive view of the news sentiment. These occasional divergences underscore the complexity of sentiment analysis and highlight the importance of analyzing both headlines and body text. By examining both components, I think it is possible to achieve a more comprehensive understanding of news sentiment and its potential impact on mutual fund returns.

Upon acquiring the news sentiment data, it becomes feasible to conduct an analysis concerning mutual fund returns, leveraging the ability of NLP techniques to quantify news sentiment. The outcomes of this analysis can reveal the statistical significance of news sentiment as a predictive factor for asset returns. Additionally, the results may indicate whether positive news sentiment correlates with positive returns and, conversely, whether negative news sentiment is associated with lower returns or potential losses. While, classifying the data between developed and developing countries, as well as differentiating between news headlines and body text, can provide deeper insights and highlight subtle nuances in the impact of news sentiment across different economic contexts and textual components.

Modeling Financial Dynamics: VAR and Vector Error Correction Model (VECM)

Some studies have demonstrated that the lagged effects of news can significantly influence returns, with the persistence of these effects varying by the frequency of the data. For instance, Costola et al. (2023) found that at lower frequencies, such as monthly data, the effects of news can persist over longer periods. News from one or two months ago can continue to impact the returns in the current month. Conversely, at higher frequencies, such as daily data, the impact of news is more immediate but short-lived. In these cases, only the news from the previous one or two days tends to have a significant effect on current-day returns. This distinction underscores the importance of incorporating lagged news data in a model that analyzes the effects of news on the returns of financial instruments, such as mutual funds.

A Vector Autoregressive (VAR) model appears to be the most widely accepted approach for this analysis. This is primarily because it accounts for endogeneity (Uhl, 2014), wherein past news can influence returns, and past returns can, in turn, affect future news sentiment. Additionally, a VAR model can incorporate other factors, such as the local stock index, exchange rates, and government bond yields, providing a comprehensive framework for analyzing the interplay between news and financial returns.

There have been approaches to stock returns like the one made by Campbell (1991), which has been known as one of the simplest and most common, where he used a first-order VAR model to forecast expected returns, while (Calomiris and Mamaysky, 2019) used different panel VAR models depending on if it was developed or emerging economies. While Uhl (2014) explains that they started with VAR but considering the data they have it was necessary to use a vector error correction model (VECM) that it adds error correction features to VAR model.

First, I collected extensive data that, based on frequency and findings from previous studies, are well-suited for the model. The variables used for the final model are as follows:

- $R_{i,t}$ Mutual Fund returns at time t
- NS_t News sentiment at time t
- $R_{m,t}$ Stock market return at time t
- $R_{f,t}$ Bond yield at time t

Mutual fund returns can be modeled as a function of news sentiment, stock market returns, and bond yields using a Vector Autoregression (VAR) framework. The relationship between these variables can be expressed through the following equation, which represents the VAR model with p lags: Returns can be expressed in terms of the other variables through the next equation:

$$\mathbf{Y}_t = \mathbf{c} + \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2} + \cdots + \Phi_p \mathbf{Y}_{t-p} + \epsilon_t \tag{4}$$

Where:

- $\mathbf{Y}_t = \begin{pmatrix} R_{i,t} \\ R_{m,t} \\ NS_t \\ R_{f,t} \end{pmatrix}$ is the vector of variables at time t

- $\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}$ is the vector of intercept terms
- Φ_i are the coefficient matrices for the lagged values
- $\epsilon_t = \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \end{pmatrix}$ is the vector of error terms

Morevover, the previous model can be expanded as a system of equations that accounts for the interdependencies among mutual fund returns $R_{i,t}$ stock market returns $R_{m,t}$ news sentiment$R_{m,t}$ and bond yields $R_{f,t}$ over time. The Vector Autoregression (VAR) model with p lags is then expressed as follows:

$$\begin{pmatrix} R_{i,t} \\ R_{m,t} \\ NS_t \\ R_{f,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} + \Phi_1 \begin{pmatrix} R_{i,t-1} \\ R_{m,t-1} \\ NS_{t-1} \\ R_{f,t-1} \end{pmatrix} + \cdots + \Phi_p \begin{pmatrix} R_{i,t-p} \\ R_{m,t-p} \\ NS_{t-p} \\ R_{f,t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \end{pmatrix} \tag{5}$$

Where each Φ_i is a 4×4 coefficient matrix corresponding to the i-th lag (from 1 to p) These matrices contain the coefficients that describe how past values of each variable in the system affect the current values. Each ϕ_i matrix is structured as follows:

$$\Phi_i = \begin{pmatrix} \phi_{11}^{(i)} & \phi_{12}^{(i)} & \phi_{13}^{(i)} & \phi_{14}^{(i)} \\ \phi_{21}^{(i)} & \phi_{22}^{(i)} & \phi_{23}^{(i)} & \phi_{24}^{(i)} \\ \phi_{31}^{(i)} & \phi_{32}^{(i)} & \phi_{33}^{(i)} & \phi_{34}^{(i)} \\ \phi_{41}^{(i)} & \phi_{42}^{(i)} & \phi_{43}^{(i)} & \phi_{44}^{(i)} \end{pmatrix} \tag{6}$$

Each element $\phi_{jk}^{(i)}$ within the matrix denotes the impact of the $k-th$ variable's lagged value on the $j-th$ variable at time t. A Vector Autoregression (VAR) model addresses the potential issue of reverse causality between news sentiment and returns by modeling each variable in terms of its own past values as well as the past values of other variables in the system. By including lagged values of all variables, the VAR framework allows for the examination of dynamic relationships and causal interactions over time. This approach helps to mitigate concerns about reverse causality because it does not assume a predetermined direction of causality but rather captures the bidirectional influences among the variables.

Before implementing a VAR model, it is essential to ensure that the data series are stationary. Stationarity implies that the statistical properties of the series, such as mean and variance, remain constant over time. Non-stationary data can lead to spurious regression results, where relationships between variables appear significant when they are not. Therefore, the first step is to perform unit root tests, such as the Augmented Dickey-Fuller (ADF) test. If the data series are found to be non-stationary, they can be differenced to achieve stationarity.

Argentina	Brazil	Canada	Chile	Mexico	USA
-3.6583	-10.2559	-10.2176	-9.6125	-9.3376	-18.8432
-2.8633	-9.9021	-9.709	-8.193	-10.7899	-20.2782
-2.8985	-12.7469	-10.5074	-3.3834	-10.5414	-19.6299
-2.8514	-9.9999	-9.9257	-10.991	-9.5757	-18.7995
-3.0654	-10.3225	-9.8924	-11.8973	-9.4469	-20.5182
	-3.5124	-9.8287	-7.8045	-9.4221	-20.5711
	-9.3042	-10.0498	-7.6167	-9.9813	-19.8048
	-13.3047	-9.6331	-7.9999	-10.484	-21.4509
	-13.3047	-10.0712	-10.8361	-7.6139	-18.8267
	-10.8486	-9.6567	-6.4975	-10.0673	-15.8728

Table 3: ADF Test for Funds returns with critical values -2.58(1pct) -1.95(5pct) -1.62(10pct)

All the mutual fund returns exhibited stationarity, as indicated by the results of unit root tests such as the Augmented Dickey-Fuller (ADF) test. Additionally, the other variables in the model, including news sentiment and stock market returns, also demonstrated stationarity. However, the bond yield variable did not initially show stationarity. To address this issue, the first difference of the bond yield was taken, which involves subtracting the previous period's value from the current period's value. This transformation effectively removed any trends or unit roots, and subsequent testing confirmed that the differenced bond yield series was stationary.

Once stationarity is confirmed, the next step is to determine the optimal number of lags to include in the VAR model. The choice of lag length is crucial because including too few lags may omit important dynamics, while too many lags can overfit the model and reduce its forecasting power. Various information criteria, such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Criterion (HQ), can be used to select the appropriate lag length.

	Average Number of lags per country
Argentina	8
Brazil	6
Canada	4
Chile	7
Mexico	5
USA	5

The number of lags required for the VAR model varies depending on the country, as indicated by the results of the lag selection criteria. For South American countries (Argentina, Brazil, and Chile), the optimal number of lags tends to be higher compared to North American countries (Canada, Mexico, and the USA). This difference suggests that markets in South American countries exhibit more complex and prolonged dynamics, requiring additional lagged terms to adequately capture the relationships among the variables.

After determining the appropriate number of lags for each country, cointegration tests were conducted to examine the long-term relationships among the variables. The Johansen cointegration test revealed evidence of cointegration for all mutual funds across the countries studied. This finding indicates that, despite short-term fluctuations, there exist long-term equilibrium relationships among mutual fund returns, news sentiment, stock market returns, and bond yields.

Given the presence of cointegration, it is appropriate to employ a Vector Error Correction Model (VECM). The VECM extends the VAR framework by incorporating error correction terms, which capture the long-term equilibrium relationships while still allowing for short-term dynamics. This approach enables the analysis of both the short-term and long-term interactions among the variables of how news sentiment and other factors influence mutual fund returns over time.

A general and simple Vector Error Correction Model (VECM) with one error correction term can be represented as follows:

$$\Delta \mathbf{Y}_t = \sigma + \sum_{i=1}^{k-1} \mathbf{B}_i \Delta \mathbf{Y}_{t-i} + \lambda \mathbf{ECT}_{t-1} + \epsilon_t \tag{7}$$

Where:

- $\Delta \mathbf{Y}_t = \begin{pmatrix} \Delta R_{i,t} \\ \Delta R_{m,t} \\ \Delta NS_t \\ \Delta R_{f,t} \end{pmatrix}$ is the vector of differenced variables at time t

- $\sigma = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \end{pmatrix}$ is the vector of constants

- $\mathbf{B}_i$ are the coefficient matrices for the lagged differenced values

- λ is the error correction coefficient vector

- $\mathbf{ECT}_{t-1}$ is the error correction term at time $t-1$

- $\epsilon_t = \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \end{pmatrix}$ is the vector of error terms

Understanding that the Vector Error Correction Model (VECM) will utilize the lagged values of the variables—namely, mutual fund returns, news sentiment, market returns, and government bond yields—the VECM for mutual fund returns can be broken down and specified as follows:

$$\Delta R_{i,t} = \sigma_1 + \sum_{i=1}^{k-1} B_{1,i} \Delta NS_{t-i} + \sum_{i=1}^{k-1} B_{2,i} \Delta R_{m,t-i} + \sum_{i=1}^{k-1} B_{3,i} \Delta R_{f,t-i} + \lambda_1 \mathrm{ECT}_{t-1} + \epsilon_{1,t} \quad (8)$$

$$\Delta NS_t = \sigma_3 + \sum_{i=1}^{k-1} B_{7,i} \Delta R_{i,t-i} + \sum_{i=1}^{k-1} B_{8,i} \Delta R_{m,t-i} + \sum_{i=1}^{k-1} B_{9,i} \Delta R_{f,t-i} + \lambda_3 \mathrm{ECT}_{t-1} + \epsilon_{3,t} \quad (9)$$

$$\Delta R_{m,t} = \sigma_2 + \sum_{i=1}^{k-1} B_{4,i} \Delta NS_{t-i} + \sum_{i=1}^{k-1} B_{5,i} \Delta R_{i,t-i} + \sum_{i=1}^{k-1} B_{6,i} \Delta R_{f,t-i} + \lambda_2 \mathrm{ECT}_{t-1} + \epsilon_{2,t} \quad (10)$$

$$\Delta R_{f,t} = \sigma_4 + \sum_{i=1}^{k-1} B_{10,i} \Delta NS_{t-i} + \sum_{i=1}^{k-1} B_{11,i} \Delta R_{i,t-i} + \sum_{i=1}^{k-1} B_{12,i} \Delta R_{m,t-i} + \lambda_4 \mathrm{ECT}_{t-1} + \epsilon_{4,t} \quad (11)$$

Initially, a Vector Autoregression (VAR) model was selected as the primary method for examining the dynamic relationships between mutual fund returns and news sentiment. However, as the research progressed, it became evident that the Vector Error Correction Model (VECM) was more appropriate. The VAR model, while useful for examining the interactions between lagged values of all variables, does not inherently account for the presence of cointegration—a situation where variables share a common long-term trend. In contrast, the VECM specifically addresses this by incorporating cointegration relationships, thereby allowing for a more nuanced analysis of how deviations from equilibrium are corrected over time.

Both the VAR and VECM frameworks enable the examination of dynamic interactions by treating and expressing each variable in terms of its own lags and the lags of other variables. However, the VECM is particularly robust in capturing equilibrium relationships, which is essential for understanding how mutual fund returns adjust in response to news sentiment and other financial indicators. The subsequent chapter provides a detailed analysis of the VECM findings and their implications. It explores how these findings enhance our understanding of market behavior and the response of mutual funds to news sentiment.

Results

Analyzing the returns of mutual funds across different countries reveals significant variations, reflecting the diverse economic conditions and market dynamics in each region. Comparing these returns can be challenging due to factors such as varying market structures, regulatory environments, and levels of market development. However, a useful starting point is to examine the cumulative returns of the respective stock market indices over the past decade. Figure 5 shows the cumulative returns of five major stock market indices:

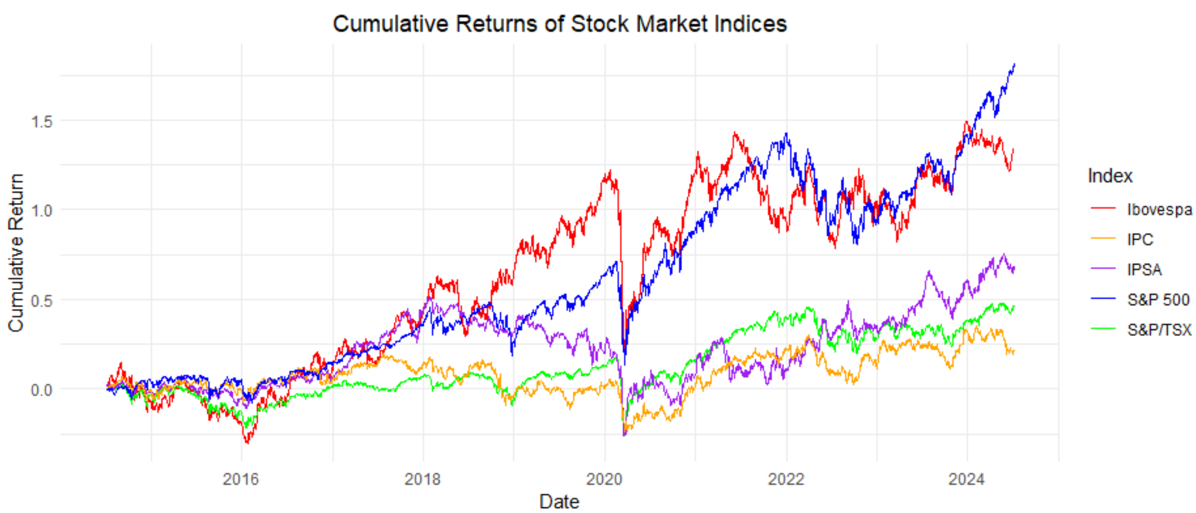


Figure 5: Cumulative Returns of Stock Market Indices

The best-performing markets, as shown in Figure 5, are the IBOVESPA (Brazil) and the S&P 500 (USA). These indices have demonstrated higher cumulative returns over the past decade, indicating stronger market performance relative to their counterparts. In contrast, the IPC (Mexico), S&P/TSX (Canada), and IPSA (Chile) show lower cumulative returns, with these three indices exhibiting roughly similar levels of performance.

The case of S&P Merval, the stock index of the Buenos Aires Stock Market (Argentina), stands out with increases that are astonishing and unmatched by any other stock market, as shown in Figure 6. This remarkable performance can be partly attributed to Argentina's position as one of the top three countries with the highest inflation rates in the world. The high inflation environment has, as expected, transferred to its stock market, leading to nominal gains that significantly outpace those of other markets. However, it is crucial to interpret these gains within the context of inflationary pressures, which can distort the real value of returns. While nominal returns appear exceptionally high, the real returns—adjusted for inflation—may present a different picture.

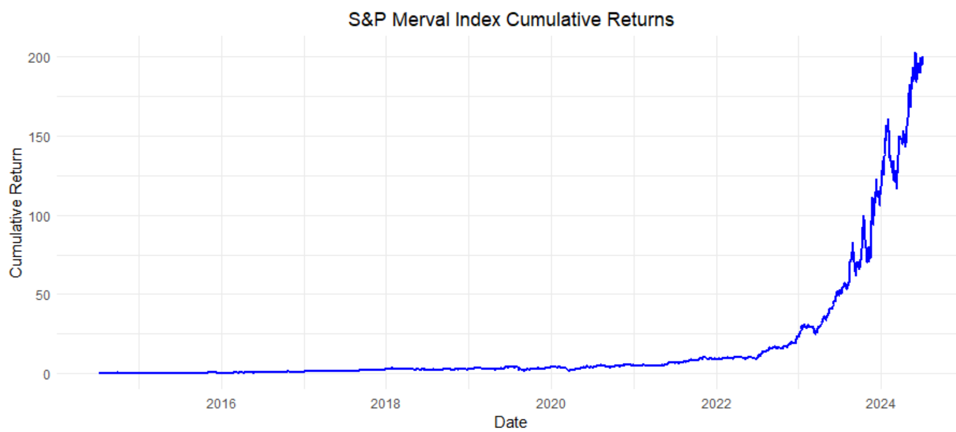


Figure 6: S&P Merval Index Cumulative Returns

When comparing the performance of the best and worst-performing mutual funds to their respective stock market indices, it is evident from Table 4 that, aside from the S&P 500, the best-performing mutual funds analyzed in the study have outperformed their respective benchmark indices. This indicates a robust ability of these funds to generate returns exceeding those of the broader market.

Index	Return	Best Performing	Return	Worst Performing	Return
S&P Merval	19883.19	COME.BA	32167.63	MTR BA	-99.57
Ibovespa	134.69	0P000163OY	171.77	OUFF11.SA	-38.61
S&P/TSX	45.38	0P0000TF96.TO	71.30	0P0000770X.TO	-12.21
IPSA	67.97	CFICI81I-E.SN	118.38	CFIHMCRGPB.SN	-6.04
IPC	20.44	SAM-APA.MX	44.87	FT-BONDBE5.MX	-10.98
S&P 500	181.51	VDIGX	71.91	FSREX	-15.07

Table 4: Percentage Returns of the Best and Worst Performing Mutual Funds

Specifically, mutual funds linked to the IBOVESPA (Brazil), IPSA (Chile), and Merval (Argentina, despite its current economic situation) have demonstrated superior performance relative to their benchmarks. Conversely, the worst-performing mutual funds are associated with the Buenos Aires Stock Exchange (S&P Merval) and the São Paulo Stock Exchange (IBOVESPA). This underperformance may be due to significant challenges faced by these funds, potentially due to heightened market volatility and economic instability in their respective regions.

Continuing with results, Table 5 presents the coefficients of the first-lag news sentiment effect on mutual fund returns for Canada and the USA, representing developed countries.

News Headlines		Body Text	
Canada	USA	Canada	USA
6.954	-7.19	10.9*	-3.15
(0.2447)	(0.23469)	(0.094799)	(0.62303)
2.13	-0.0974	11.9*	23.1*
(0.713811)	(0.92547)	(0.067942)	(0.03558)
6.227	2.58	16	8.27
(0.555223)	(0.6612)	(0.169478)	(0.1836)
-1.764	-13.577	3.08	17.2
(0.76802)	(0.355892)	(0.64333)	(0.26887)
16.431	9.8	-12.4	39.5***
(0.21943)	(0.43452)	(0.39988)	(0.003009)
5.229	-4.03	1.36	19.6*
(0.797791)	(0.67672)	(0.95265)	(0.0561)
-1.76	4.732	0.548	11.1
(0.75773)	(0.573139)	(0.93109)	(0.208665)
6.101	-2.49	-7.98	-3.86
(0.45339)	(0.535022)	(0.374938)	(0.3611)
10.64	0.071	23.3**	5.82
(0.25268)	(0.88457)	(0.021505)	(0.2581)
7.142	5.26	15.6	-1.51
(0.746684)	(0.116238)	(0.525675)	(0.669704)
16.083	8.12	31.6***	29.2**
(0.117402)	(0.454998)	(0.0042)	(0.0115)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
All coefficients are scaled by a factor of 10^{-4}.

Table 5: Impact of First-Lag News Sentiment on Mutual Fund Returns in Developed Countries

The number of lags included in the model varies across mutual funds and countries, reflecting the different dynamics and market conditions in each region. On average, the optimal number of lags for mutual funds in North American countries, such as Canada and the USA, is lower than that for mutual funds in South American countries, including Brazil, Argentina, and Chile. This variation suggests that market adjustments and reactions to news occur more rapidly in developed markets compared to emerging markets.

As shown in Table 5, the effects of the first lag of news sentiment are divided into News Headlines and Body Text. The table reveals that many coefficients for the first lag are not statistically significant. However, among the statistically significant coefficients, most pertain to the News Body Text rather than the News Headlines. This suggests that at the first lag the detailed content of the news articles exerts a more substantial influence on mutual fund returns compared to the headlines alone.

News Headlines		Body Text	
Canada	USA	Canada	USA
26.5***	17.2**	33.1***	16.7**
(0.001045)	(0.02497)	(0.0005)	(0.03532)
-11.2	-4.75	21.8***	11.5
(0.148296)	(0.71876)	(0.007169)	(0.39435)
27.9*	-10.2	43.4***	-1.46
(0.053386)	(0.1697)	(0.003061)	(0.8486)
-8.5	-29.95	-2.52	-15.3
(0.293732)	(0.107201)	(0.76205)	(0.426276)
34.8**	6.69	40**	42.6***
(0.04979)	(0.67301)	(0.026908)	(0.009838)
-24.3	-16.7	-17.2	6.52
(0.38015)	(0.17193)	(0.55081)	(0.6033)
-6.75	4.83	-2.21	-1.89
(0.38196)	(0.649486)	(0.78117)	(0.862273)
1.05	-4.19	-17.4	9.62*
(0.92294)	(0.40802)	(0.120363)	(0.0648)
-28.1**	0.182	41.5***	2.64
(0.02387)	(0.97651)	(0.000963)	(0.6769)
82.2***	-9.48**	-45.1	-6.58
(0.006813)	(0.025803)	(0.144887)	(0.132918)
50.5***	3.49	66.1***	23.7*
(0.000313)	(0.799977)	(0.00002)	(0.0965)

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
All coefficients are scaled by a factor of 10^{-4}.

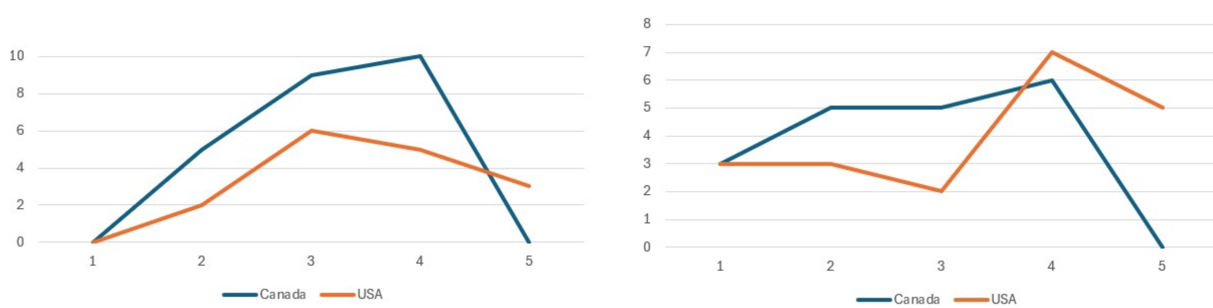
Table 6: Impact of Second-Lag News Sentiment on Mutual Fund Returns in Developed Countries

Moreover, as shown in Table 6, the number of statistically significant coefficients increases with higher lags, indicating that the influence of news sentiment on mutual fund returns becomes more pronounced over time. This lagged effect suggests that the market's reaction to news is not immediate, and the information embedded in news sentiment takes time to be fully absorbed and reflected in mutual fund returns. This delay in market response may be due to a variety of factors, such as the time required for investors to process and act on new information or the gradual dissemination of news across different segments of the market.

Interestingly, the statistically significant coefficients are predominantly positive, indicating that positive or detailed news content tends to have a favorable impact on mutual fund returns. Conversely, while there are some negative coefficients, these are generally not statistically significant.

It is also observed that the effect of news sentiment on the first lag is generally, on average, lower than that of their respective stock market indices for both Canada and the USA. However, as we move to the subsequent lags, such as the second lag, the patterns diverge between the two countries. For Canada, the coefficients representing the effects of news sentiment remain consistently lower than those of its benchmark index across all lags. This suggests that news sentiment has a relatively muted impact on mutual fund returns compared to the broader market movements reflected in the stock market index.

In contrast, the USA exhibits a different pattern at the second lag. Specifically, the coefficients for News Titles show a higher average impact on mutual fund returns than the respective stock market index, while the coefficients for News Body Text remain lower. This indicates that the headlines alone might have a more immediate and pronounced effect on mutual fund performance in the USA, whereas the detailed content of the news articles continues to have a lesser influence.



(a) News Heading

(b) News Body Text

Figure 7: Number of Mutual Funds with Statistically Significant Coefficients from News Sentiment Effects Across Different Lags

Furthermore, Figure 7 illustrates the number of mutual funds with statistically significant coefficients from news sentiment effects across different lags. The figure distinguishes between news headlines and news body text, revealing distinct patterns for each type of news sentiment.

For news headlines, the number of mutual funds with statistically significant coefficients increases with subsequent lags, forming a bell-shaped pattern. In Canada, this number rises from zero to ten (out of ten mutual funds) before decreasing. This pattern suggests that the impact of news headlines on mutual fund returns intensifies initially, peaks, and then diminishes over time. The USA exhibits a similar but less pronounced bell-shaped pattern, indicating that the influence of news headlines on mutual funds also fluctuates with lags but to a lesser extent. This bell-shaped pattern could be indicative of market participants gradually incorporating the information from news headlines into their trading decisions, leading to an initial increase in significance followed by a decrease as the market adjusts to the new information.

In the case of news body text sentiment, the pattern is less consistent. While there is some movement up and down, the changes are not as recognizable or stable as those for news headlines. For some lags, the number of statistically significant coefficients remains steady or even decreases before rising again. This variability suggests that the impact of news body text sentiment on mutual fund returns is more sporadic and less predictable compared to news headlines. Empirically, this could be due to the more detailed and nuanced information contained in news body text, which may take longer for market participants to interpret and act upon, leading to a less uniform pattern over time.

The increasing number of statistically significant coefficients with higher lags, as shown in Figure 7, indicates that the influence of news sentiment on mutual fund returns becomes more pronounced over time. This lagged effect suggests that the market's reaction to the news is not immediate and that the information embedded in news sentiment takes time to be fully absorbed and reflected in mutual fund returns. The bell-shaped pattern observed for news headlines highlights the initial strong reaction followed by a gradual assimilation of the information, whereas the more erratic pattern for news body text suggests a more complex and delayed market response.

When considering the rest of the Americas' stock markets, particularly those of Argentina, Brazil, and Chile, distinct patterns emerge in comparison to developed countries like the USA and Canada. As previously mentioned, with the exception of Mexico, most developing countries tend to exhibit a greater number of lags in their models. This indicates that the market dynamics in these countries require a longer period to fully capture the influence of news sentiment on mutual fund returns compared to their developed counterparts.

As shown in Table 7, similar to the USA and Canada, there are fewer mutual funds showing statistically significant coefficients for these developing countries. However, as the number of lags increases, the significance of the coefficients becomes more apparent. Notably, Mexico stands out as an exception. While it has a number of lags comparable to that of the USA, it exhibits a higher number of statistically significant coefficients at earlier lags. This pattern suggests that Mexico's market behavior, in terms of responsiveness to news sentiment, aligns more closely with developed markets.

News Headlines				Body Text			
Argentina	Brazil	Chile	Mexico	Argentina	Brazil	Chile	Mexico
0.056	0.0002	-0.0013	0.001*	0.065	0.0003	-0.00049	0.00086**
(0.37)	(0.91)	(0.65)	(0.097)	(0.42)	(0.85)	(0.86)	(0.021)
-0.011	-0.0034	-0.0032	0.0023	-0.045	0.0024	-0.00043	0.0032*
(0.79)	(0.2)	(0.3)	(0.36)	(0.35)	(0.350)	(0.15)	(0.09)
0.078*	0.001789	0.0004	-0.0021	0.064	-0.0084	-0.0017	-0.0019
(0.08)	(0.75217)	(0.859119)	(0.23)	(0.22)	(0.12)	(0.45)	(0.15)
0.14	-0.0012	0.0025	-0.00018	0.019	-0.0012	0.0026	-0.00004
(0.11)	(0.532)	(0.15)	(0.63)	(0.82)	(0.53)	(0.129)	(0.88)
0.11	-0.00007	-0.0017	-0.00002	-0.019	-0.000971	-0.001424	0.00001
(0.2)	(0.97)	(0.3)	(0.53)	(0.82)	(0.54)	(0.36)	(0.48)
0.017	0.00004	0.0014	-0.00022	0.032	0.000035	0.0027**	0.00035
(0.62)	(0.52)	(0.3)	(0.7)	(0.55)	(0.52)	(0.036)	(0.43)
	-0.00023	0.0047	0.00054		-0.00052	0.0033	0.001
	(0.76)	(0.14)	(0.57)		(0.47)	(0.3)	(0.18)
	-0.00058	0.000649	0.00068		-0.00087	0.0009	0.0007
	(0.58)	(0.53)	(0.37)		(0.38)	(0.34)	(0.2)
	-0.000861	-0.002038	0.00025		-0.0005	-0.004368	0.0006141***
	(0.57)	(0.58)	(0.27)		(0.72)	(0.2)	(0.000526)
	0.0028**	0.00045	0.00057		0.0026473**	0.003041	-0.000145
	(0.04)	(0.85)	(0.65)		(0.0422)	(0.19)	(0.88)
	0.0018	0.0013408	0.0055*		0.00009	0.0021	0.0054**
	(0.5)	(0.56)	(0.05)		(0.97)	(0.33)	(0.016)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Impact of First-Lag News Sentiment on Mutual Fund Returns in Developing Countries

Interestingly, even though the number of lags is higher for South American countries, the coefficients for the news sentiment effects are, on average, larger than those observed in the USA and Canada. This indicates a stronger impact of news sentiment on mutual fund returns in these developing markets. For the specific case of Mexico, which has fewer lags akin to the USA and Canada, the average size of the coefficients is also larger, similar to what is observed in Argentina, Brazil, and Chile.

The larger coefficients in South American countries suggest a more pronounced reaction to news sentiment, reflecting the potentially higher volatility and sensitivity of these markets to external information. The higher number of lags indicates that the information dissemination and absorption processes are more prolonged, possibly due to less efficient market mechanisms or greater economic uncertainties. In contrast, the smaller number of lags in Mexico, USA, and Canada points to quicker information processing and market adjustments, characteristic of more developed financial markets.

	News Headlines				Body Text		
Argentina	Brazil	Chile	Mexico	Argentina	Brazil	Chile	Mexico
0.0129	-0.0022	-0.001312	0.001	-0.0016	-0.0016	-0.0028	0.0013***
(0.88)	(0.3181)	(0.72471)	(0.17)	(0.99)	(0.44)	(0.43)	(0.00432)
0.09	0.0075**	-0.0004014	0.00234	-0.08	-0.00091	-0.0055	0.0063***
(0.17)	(0.038)	(0.92)	(0.43)	(0.18)	(0.78)	(0.16)	(0.0057)
0.23***	0.0051	0.0028	-0.00089	-0.019	-0.0082	0.00086	0.0018774
(0.005)	(0.502)	(0.34)	(0.67)	(0.78)	(0.25)	(0.77)	(0.24)
0.19	0.00043	0.0051**	0.000731*	-0.012	-0.0012	0.0013	0.00006
(0.16)	(0.87)	(0.032)	(0.092)	(0.92)	(0.609)	(0.55)	(0.85)
0.123	0.0012	-0.0032	-0.00008	-0.05	-0.00062	-0.00198	0.0004*
(0.32)	(0.6)	(0.14)	(0.81)	(0.64)	(0.76)	(0.33)	(0.05752)
0.062	-0.00006	0.0013	0.00172***	0.053	-0.00003	0.003	0.00057
(0.204)	(0.42)	(0.46)	(0.0098)	(0.48)	(0.65)	(0.081)	(0.28)
	0.00045	0.00027	-0.0016		0.00023	0.0018	0.00129
	(0.65)	(0.95)	(0.16)		(0.80)	(0.66)	(0.18)
	-0.0008	0.00012	0.0009		-0.0011	0.00032	0.0017**
	(0.56)	(0.93)	(0.3)		(0.41)	(0.79)	(0.01)
	-0.00077	0.00975**	-0.00003		0.00007	-0.0079*	0.00069***
	(0.7)	(0.044)	(0.91)		(0.97)	(0.073)	(0.002)
	0.0025	0.0048	-0.00017		0.00063	0.0066**	0.00142
	(0.18)	(0.129)	(0.91)		(0.71)	(0.027)	(0.99)
	0.0062*	0.0002	0.0081**		0.00196	-0.0017	0.0096***
	(0.09)	(0.95)	(0.015)		(0.56)	(0.55)	(0.0008)

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

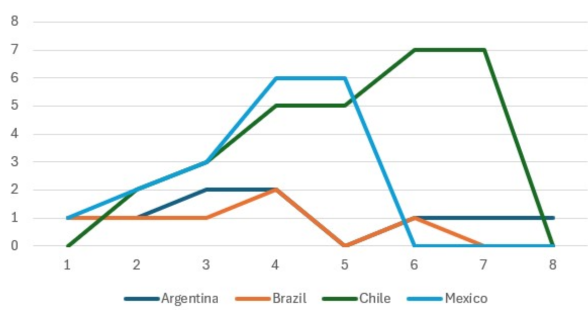
Table 8: Impact of Second-Lag News Sentiment on Mutual Fund Returns in Developing Countries

As observed in Table 8, titled "Impact of Second-Lag News Sentiment on Mutual Fund Returns in Developing Countries," there is a notable increase in the number of mutual funds with statistically significant coefficients, and these coefficients are, on average, larger compared to those in developed countries. This trend is particularly evident when examining the effect of news sentiment at the second lag. Developing countries, such as Argentina, Brazil, and Chile, exhibit a more pronounced and statistically significant response to news sentiment, reflecting greater market sensitivity compared to their developed counterparts.

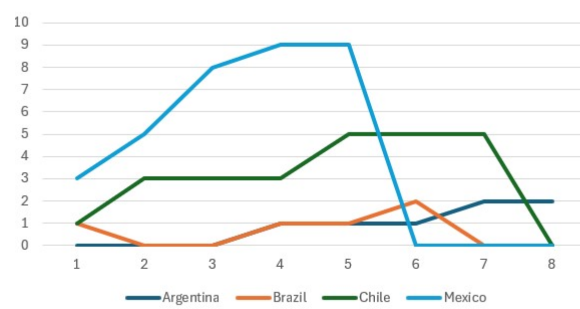
This trend persists in subsequent lags. The coefficients for news sentiment at higher lags are generally larger on average than those at earlier lags within the same dataset. Specifically, in developing countries, the impact of news sentiment on mutual fund returns becomes more pronounced over time, with subsequent lags showing greater statistical significance and higher magnitude coefficients compared to earlier lags. This indicates that the influence of news sentiment not only grows with time but also demonstrates a stronger effect in developing markets compared to developed ones.

This pattern suggests that mutual funds in developing countries react more robustly to news sentiment as time progresses, potentially due to higher market volatility and less efficient information dissemination. The increasing magnitude of coefficients with each lag could reflect the delayed but amplified market response to news sentiment, underscoring the dynamic nature of investor behavior in these emerging markets.

In contrast, the effect of news sentiment on mutual fund returns in developed countries shows a relatively stable pattern across lags, with fewer significant coefficients and generally lower magnitudes. This stability in developed markets suggests a more immediate and contained reaction to news, possibly due to more efficient markets and established mechanisms for integrating new information.



(a) News Heading



(b) News Body Text

Figure 8: Number of Mutual Funds with Statistically Significant Coefficients from News Sentiment Effects Across Different Lags Developing

Similar to the analysis conducted for developed countries, Figure 8 illustrates the number of mutual funds with statistically significant coefficients from news sentiment effects across different lags for developing countries, specifically Argentina, Brazil, Chile, and Mexico. This figure differentiates between the effects of news headlines and news body text, akin to Figure 7 for developed countries.

In the case of Brazil and Argentina, the pattern of significant coefficients appears more erratic, with relatively low numbers of mutual funds exhibiting statistically significant responses across various lags. This erratic behavior may reflect the higher market volatility and less mature financial systems in these countries. Conversely, Mexico and Chile display a more discernible pattern. For these countries, there is an increasing number of mutual funds with significant coefficients as the lags progress, indicating a more pronounced and stable response to news sentiment.

Chile, which has a higher level of development compared to other Latin American countries, shows a peak in the number of mutual funds with statistically significant coefficients at the 6th and 7th lags for news headlines. While Mexico, with its large and industrialized economy, demonstrates a notable increase in the number of mutual funds with significant coefficients, particularly in the 4th and 5th lags. This indicates that Chilean mutual funds react more strongly to news headlines at longer lags compared to Mexican funds. However, when considering news body text, Mexico surpasses Chile, with nine out of ten funds showing significant coefficients at the 4th and 5th lags. Chile, in contrast, maintains a consistent level of significance from the 5th to 7th lags, with fewer mutual funds reaching significance at higher lags.

The findings highlight that while Mexico and Chile demonstrate some similarities to developed countries in terms of the pattern of significant coefficients, there are notable differences. For instance, the number of lags required to capture the dynamics of news sentiment is generally higher in developing countries, with coefficients of greater magnitude compared to their developed counterparts. This suggests that the impact of news sentiment is more enduring and pronounced in the mutual funds of these developing markets.

The more erratic behavior observed in Argentina and Brazil contrasts with the relatively stable and pronounced patterns in Mexico and Chile. The latter countries exhibit similarities to developed markets in terms of the structure of significant coefficients but still require more lags to fully model the dynamics of news sentiment, reflecting the unique characteristics of emerging market economies.

As the approach to analyzing mutual fund returns explained by news sentiment utilizes the Vector Error Correction Model (VECM), all mutual funds in developed countries, such as Canada and the USA, incorporate an error correction term. Table 9 presents these error correction terms, which are consistently negative and statistically significant across the board. The magnitude of these coefficients ranges from 0 to less than -1, indicating that the short-term deviations from the long-term equilibrium are corrected over time. Specifically, the negative values suggest that any deviations from the equilibrium caused by news sentiment are promptly adjusted. Therefore, the presence of a significant negative error correction term highlights that the mutual funds in these developed markets exhibit a tendency to revert to their long-term equilibrium, with the rate of adjustment varying but remaining within the specified range.

News Headlines		Body Text	
Canada	USA	Canada	USA
-0.0313***	-0.0661**	-0.2658***	-0.0651**
(2.80E-12)	(0.02167)	(1.48E-14)	(0.02839)
-0.3167***	-0.7280***	-0.4281***	-0.7767***
(5.01E-7)	(2E-16)	(7.62E-7)	(2E-16)
-0.1449***	-0.0390	-0.2916***	0.0270
(9.90E-06)	(0.5383)	(0.000639)	(0.6758)
-0.0028*	-0.2536***	-0.0404	0.2546***
(0.000268)	(0.000835)	(0.13936)	(0.000767)
0.2327**	-0.1880***	-0.2459**	-0.8517***
(0.01032)	(2E-16)	(0.015269)	(2E-16)
-0.1207***	-1.4251***	-0.0248***	-1.22
(8.58E-7)	(2E-16)	(0.00387)	(2E-16)
-0.0154**	-0.7254***	0.0253	-0.7443***
(0.00139)	(2E-16)	(0.45584)	(2E-16)
-0.1507*	-0.1760***	-0.2247**	-0.197
(0.0216)	(1.86E-5)	(0.002031)	(3.75E-6)
-0.9861***	-0.7240***	-1.2613***	0.7389***
(5.18E-8)	(2E-162)	(1.10E-9)	(2E-16)
-0.3494***	-0.0674	-0.0246***	-0.0625
(1.23E-7)	(0.222973)	(7.69E-9)	(0.252687)
-0.3281***	-1.1906***	-0.6418***	-1.2075***
(4.20E-12)	(2E-16)	(2E-16)	(2E-16)

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
E-notation represents powers of 10

Table 9: Error Correction Term Developed Countries

In Table 10, which focuses on developing countries such as Argentina, Brazil, Chile, and Mexico, a similar pattern is observed but with notable differences in magnitude. The error correction terms for these countries are also negative and statistically significant but exhibit larger absolute values, with some coefficients falling below -1 (e.g., -1.16). These larger negative values indicate that the effects of news sentiment—both in the short run and the long run—are more pronounced in developing countries compared to their developed counterparts. This suggests a greater degree of volatility and a stronger need for correction in the face of deviations from equilibrium. The more negative error correction terms in developing countries imply that these markets experience more significant and abrupt adjustments in response to news sentiment, reflecting higher market instability and sensitivity.

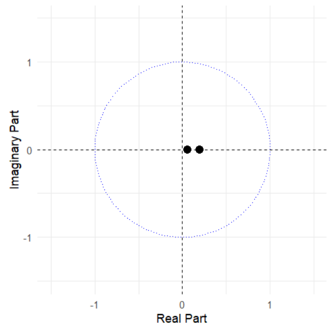
News Headlines				Body Text			
Argentina	Brazil	Chile	Mexico	Argentina	Brazil	Chile	Mexico
-0.235167*	-1.1749***	-0.5437***	-1.037***	-0.1521	-1.1693***	-0.5198***	-0.8305***
(0.055)	(2.21E-12)	(2.82E-08)	(2E-16)	(0.468)	(1.91E-12)	(2.46E-07)	(2E-16)
-1.5864**	-0.9000***	-0.0277	-1.0415***	-1.7381*	-0.5471*	3.09E-04	-0.8947***
(0.0188)	(1.23E-06)	(0.3862)	(2E-16)	(0.0484)	(0.0006)	(0.963)	(2E-16)
-0.8734	-0.2823***	-0.7175***	-1.1521***	-0.3397	-0.1168*	-0.0057**	-0.9391***
(0.1758)	(0.0096)	(2.26E-09)	(2E-16)	(0.14)	(0.0914)	(0.03)	(2E-16)
-1.8323	-1.1102***	-0.0651**	-1.1300***	-0.3751	-1.0882***	0.0232	-1.11***
(0.2873)	(1.67E-11)	(0.0012)	(2E-16)	(0.874)	(2.74E-11)	(0.316)	(2E-16)
-1.3303	-1.0300***	-1.1692***	-0.9060***	-1.6960**	-0.9933***	-1.3833***	-0.9034***
(0.1073)	(1.29E-07)	(6.32E-13)	(2E-16)	(0.0266)	(1.40E-07)	(8.86E-15)	(2E-16)
-1.6653***	-0.0345	-0.0704	-1.0800***	-0.1356*	0.0310	-0.0498*	-0.1295***
(0.0052)	(0.4410)	(0.1581)	(1.97E-08)	(0.076)	(0.2352)	(0.0928)	(0.0004)
	-0.0400**	-0.1044	-0.9503***		-0.0024	-0.0082	-0.2264***
	(0.0029)	(0.1266)	(2E-6)		(0.1560)	(0.2441)	(7.13E-08)
	-1.4285***	-0.1164**	-0.7270***		-1.4351***	-0.8429***	-1.0900***
	(1.12E-13)	(0.0104)	(2.97E-16)		(8.86E-14)	(4.19E-09)	(2E-16)
	-0.8053***	-1.1617***	-0.0107		-0.8150***	-1.3964***	-0.0647*
	(4.25E-08)	(6.87E-14)	(0.365)		(1.97E-08)	(2E-16)	(0.0044)
	-1.5006***	-0.5200***	-0.5033***		-1.4196***	-0.5012***	-0.0099
	(2.47E-13)	(1.10E-07)	(8.70E-11)		(1.03E-12)	(2.77E-07)	(0.141)
	-0.9862***	-0.0072*	-0.0052**		-0.9580***	-0.1606***	-0.0020*
	(2.86E-07)	(0.06)	(0.02)		(2.86E-07)	(0.0017)	(0.08)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
E-notation represents powers of 10

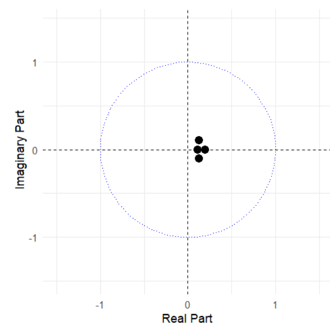
Table 10: Error Correction Term Developing Countries

The Error Correction Terms (ECT) typically involve lagged values of the variables included in the model. For example, in the case of the USA, where the lag length is set at 5, the ECT incorporates returns lagged by 6 periods, news sentiment lagged by 6 periods, index returns lagged by 6 periods, and government bond yields lagged by 6 periods. This setup indicates that the model captures the long-term equilibrium relationships by considering the past values of these variables, which helps in understanding how past deviations from equilibrium influence current adjustments. The inclusion of these lagged terms ensures that the model accounts for the delayed effects of news sentiment and other variables on mutual fund returns, thereby providing a more comprehensive view of the dynamic interactions and adjustment processes.

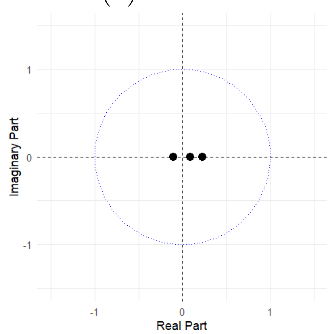
Figure 9 illustrates the stability plots for the stock market indices of the countries under study. These plots display the eigenvalues of the companion matrices from the Vector Error Correction Model (VECM) analysis, which are crucial for assessing the stability of the estimated models. The results indicate that all eigenvalues are located within the unit circle, confirming that the models for the stock market indices are stable. This stability suggests that the dynamic relationships modeled by the VECM are robust and do not exhibit explosive behavior over time.



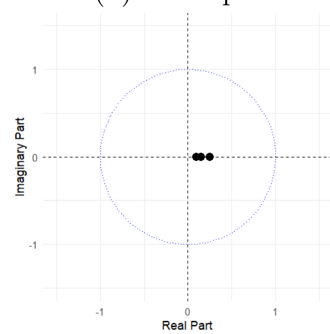
(a) Merval



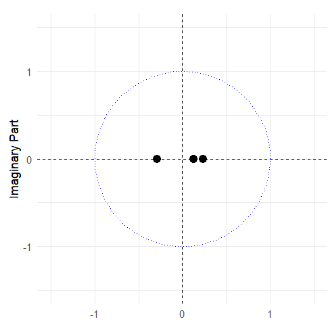
(b) Ibovespa



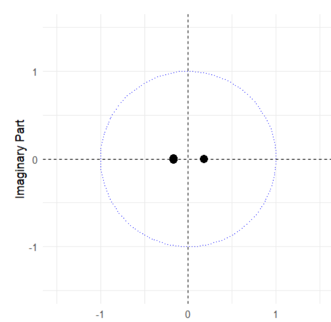
(c) S&P TSX



(d) Ipsa



(e) IPC



(f) S&P 500

Figure 9: Developing Economies

Specifically, the eigenvalues, which are plotted in the complex plane, are all positioned inside the unit circle. This is a key criterion for stability in time series models, ensuring that the estimated VECM coefficients lead to stable and convergent behavior in the long run. The fact that none of the eigenvalues lie on or outside the unit circle implies that the stock market indices' relationships with news sentiment are stable and reliable for forecasting and interpretation.

For brevity, the stability analysis shown is on the stock market indices only. The stability plots for mutual funds, while similar in concept, are not presented here but exhibit analogous stability characteristics. The underlying principle is that if the stock market indices are stable, the mutual funds, which are influenced by these indices, are likely to exhibit similar stability patterns. This ensures that the results and insights derived from the VECM analysis for mutual funds are grounded in a stable and well-behaved framework.

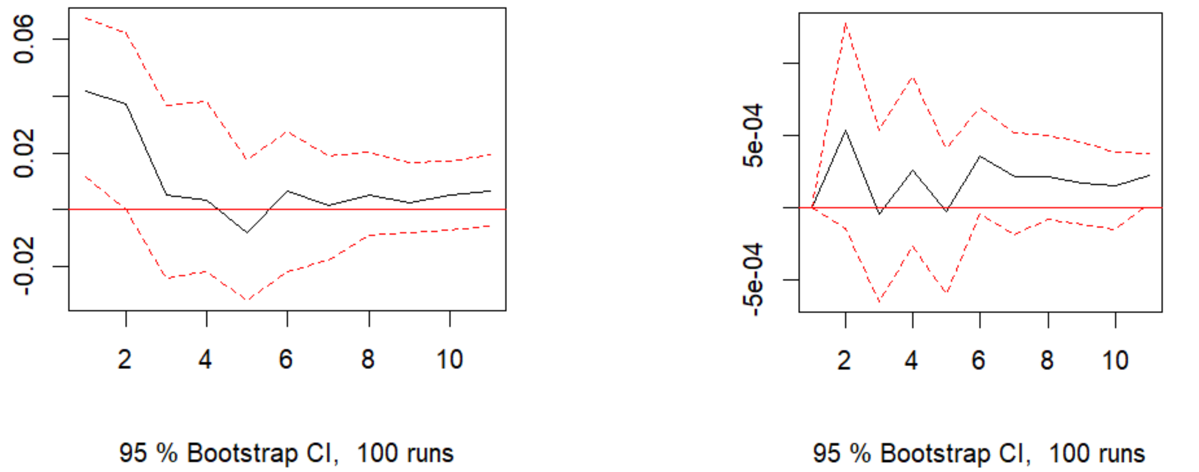
A critical aspect of this study is addressing endogeneity concerns, particularly reverse causality between news sentiment and mutual fund returns. To this end, we conducted a variance decomposition analysis to understand the extent to which news sentiment and past returns explain current mutual fund returns and vice versa.

The results indicate that news sentiment increasingly explains the variation in mutual fund returns over time. Initially, the lagged values of mutual fund returns themselves are the primary drivers of current returns, highlighting the momentum effect in financial markets where past performance heavily influences current performance. However, the influence of these lagged returns diminishes over time as the explanatory power of news sentiment grows.

On the other side, the decomposition also reveals that news sentiment is predominantly explained by its own past values, which is expected given the autoregressive nature of sentiment indices. However, it is noteworthy that the explanatory power of lagged returns on current news sentiment increases over subsequent periods. This suggests a feedback loop where market performance influences the tone and content of news reporting, which in turn, affects future market performance.

Specifically, in the initial periods, returns lags explain a significant portion of the variation in current returns, underscoring the inherent autocorrelation in financial returns. As the lags progress, the proportion of variance explained by news sentiment grows, indicating that the market gradually incorporates and reacts to new information conveyed through news sentiment.

Conversely, while news sentiment's current state is primarily driven by its past values, the increasing influence of lagged returns points to a scenario where market movements are reflected in news sentiment over time. This relationship suggests that while news sentiment can be an exogenous predictor of returns, it also endogenously responds to the market's past performance. Initially, returns are more self-referential, but over time, the impact of news sentiment becomes more pronounced. By accounting for this bidirectional relationship, the VECM approach provides a comprehensive understanding of the underlying mechanisms driving mutual fund performance. This nuanced view highlights the dynamic interplay between news sentiment and financial returns, offering valuable insights for market analysts.



(a) Canada: Impulse Response of Sentiment to Return Shocks

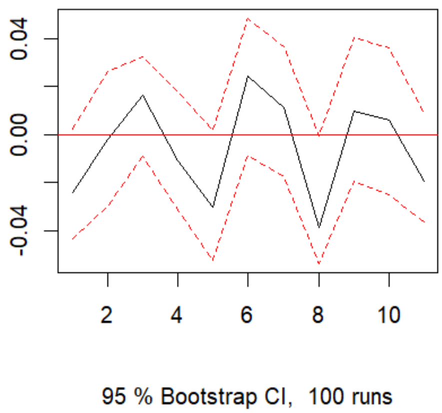
(b) USA: Impulse Response of Sentiment to Return Shocks

Figure 10: Impulse Response Analysis for Developed Economies: Return Shocks

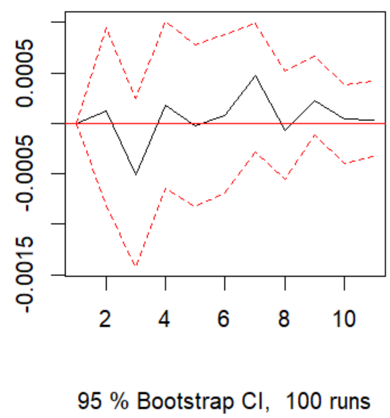
Figures 10 and 11 present the impulse response functions (IRFs) of news sentiment to return shocks for developing and developed countries, respectively. These figures illustrate the dynamic response of news sentiment to unexpected changes in returns over time, providing insights into how news sentiment reacts to market movements.

Figure 10 depicts the IRFs for developed countries like Canada and the USA. In Canada, the IRFs initially declined before stabilizing, indicating a delayed but steady incorporation of market information on returns into news sentiment. The USA shows a different pattern, with news sentiment experiencing ups and downs but remaining within a consistent range for peaks and valleys, suggesting a more robust and balanced reaction to return shocks.

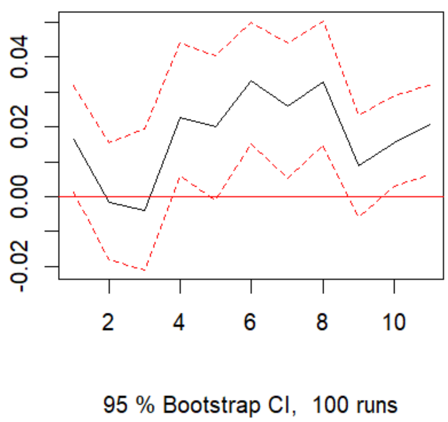
This behavior highlights the relative stability and efficiency of information processing in the USA's financial markets. Interestingly, Mexico's IRFs are similar to those of the USA, demonstrating even more stability and indicating a high level of market maturity compared to other developing countries.



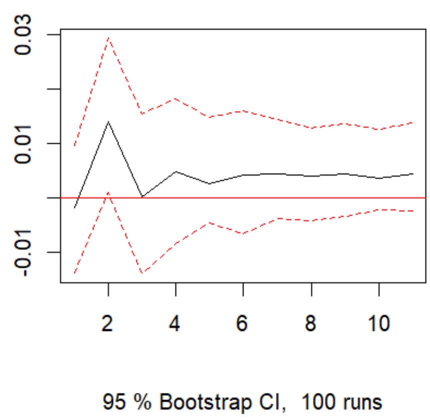
(a) Argentina: Impulse Response of Returns to Sentiment Shocks



(b) Brazil: Impulse Response of Returns to Sentiment Shocks



(c) Chile: Impulse Response of Returns to Sentiment Shocks



(d) Mexico: Impulse Response of Returns to Sentiment Shocks

Figure 11: Impulse Response Analysis for Developing Economies: Return Shocks

In Figure 11, which focuses on developing countries such as Argentina, Brazil, Chile, and Mexico, the impulse responses are characterized by greater volatility and less stability over time, with the notable exception of Mexico. This volatility reflects the higher market uncertainty and economic instability often present in these countries. Specifically, in South American countries like Argentina, Brazil, and Chile, the reactions to return shocks are more intense and erratic, demonstrating the challenges these markets face in processing and integrating new information. In contrast, Mexico's responses are more stable and akin to those observed in developed countries. Overall, in developing countries, particularly in South America, the reactions are more pronounced and unpredictable, whereas in developed countries, the responses tend to be more moderate and stable.

Ultimately, the final section of this chapter focuses on forecasting mutual fund returns based on news sentiment using the VECM approach. This study aims to predict future return movements by leveraging the relationships established between news sentiment and returns. To this end, Figures 12 and 13 present the forecasted mutual fund returns for Mexico and Canada, respectively. These forecasts are generated by applying the VECM framework to project how future news sentiment might influence returns, providing insights into potential market trends. For a comprehensive view, including forecasts for other countries in the study, please refer to the appendix. This section underscores the practical utility of the VECM approach in financial forecasting and highlights the potential for news sentiment to serve as a valuable predictor of market dynamics.

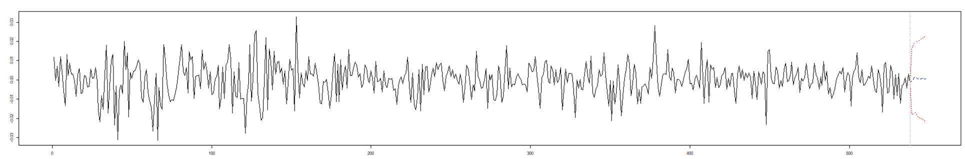


Figure 12: Canada: Forecasts of Returns

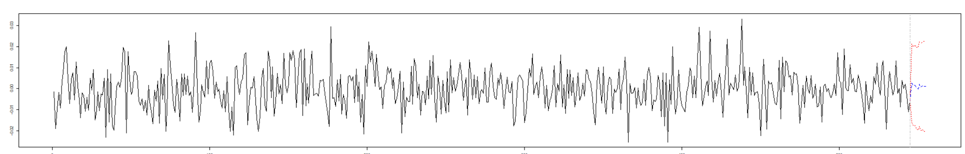


Figure 13: Mexico: Forecasts of Returns

In conclusion, this chapter presents a comprehensive analysis of the impact of news sentiment on mutual fund returns across different regions using the VECM approach. The results reveal that while developed markets like Canada and the USA show more stable and predictable responses, developing markets exhibit higher volatility and variability. The VECM models effectively capture these dynamics, offering valuable insights into the differential behavior of mutual funds across economic contexts and enhancing our understanding of the interplay between news sentiment and financial returns.

Conclusion

The results of this paper have shown that news sentiment impacts mutual funds differently across various countries, with developed markets exhibiting more uniform reactions to news compared to their counterparts in developing markets. In developed countries like Canada and the USA, mutual funds tend to respond in a more synchronized manner and less volatile. This suggests a more efficient integration of news into market pricing. Conversely, mutual funds in developing countries show a wider range of responses, indicating less predictable and more fragmented influences of news sentiment, potentially due to lower information dissemination efficiency and varying market sophistication levels.

The VECM model emerged as the most fitting approach based on both the evidence and existing literature, effectively addressing endogeneity and capturing the dynamic relationships between news sentiment and mutual fund returns. The model's ability to incorporate long-term equilibrium relationships and short-term dynamics proved crucial in understanding these interactions. Additionally, the lag selection process revealed notable differences across countries. Developed markets such as Canada and the USA required fewer lags to capture significant relationships, reflecting a more immediate and efficient response to news sentiment. In contrast, countries like Brazil, Chile, and Argentina required more lags on average, indicating that their markets exhibit slower adjustments to news and potentially more complex dynamics in integrating sentiment into returns. Notably, Mexico has shown characteristics of both, with a lag structure similar to that of the USA, suggesting a hybrid response pattern that blends traits of both developed and developing markets.

However, the coefficients revealed that the reactions to news sentiment, while more variable and hard to capture, were generally greater in magnitude in developing markets compared to their developed counterparts. This suggests that, although the responses in developing countries may be less synchronized and more heterogeneous, the impact of news sentiment tends to be more pronounced. This heightened sensitivity could be attributed to factors such as less efficient information dissemination, greater market volatility, and/or more pronounced investor reactions in these regions. Conversely, in developed markets, where information is processed more efficiently, and market structures are more stable, the reactions to news sentiment tend to be more moderate and consistent.

Some of the limitations of this study include the precision with which news sentiment is measured and the computational resources required for implementing the VECM models. The accuracy of news sentiment analysis is dependent on the effectiveness of Natural Language Processing (NLP) techniques, which can sometimes struggle with nuances in language or context, potentially impacting the reliability of sentiment scores. Additionally, VECM models, while powerful in addressing endogeneity and capturing dynamic relationships, are computationally intensive and can require substantial processing power, particularly when dealing with large datasets and multiple countries. This computational demand can limit the feasibility of applying such models in real-time or on a broader scale. Furthermore, there are potential issues related to the generalizability of the findings, as the model may not fully account for all regional or sectoral variations in news impact.

Overall, this study makes a significant contribution to the comparative analysis of how news sentiment affects returns across diverse regions and countries. By integrating advanced Natural Language Processing (NLP) and machine learning techniques into asset pricing models and utilizing a Vector Error Correction Model (VECM) approach, this research provides valuable insights into the dynamic interactions between news sentiment and market reactions. This contribution enhances understanding of regional disparities in market behavior and bridges the gap between theoretical advancements and practical applications. Consequently, the study paves the way for future research and advancements in the integration of news sentiment into financial modeling to further enhance the accuracy and effectiveness of financial analyses.

Appendix

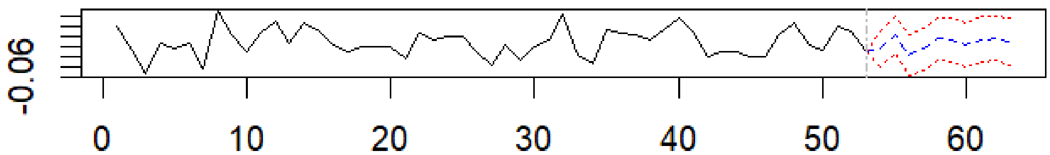


Figure 14: Argentina: Forecasts of Returns

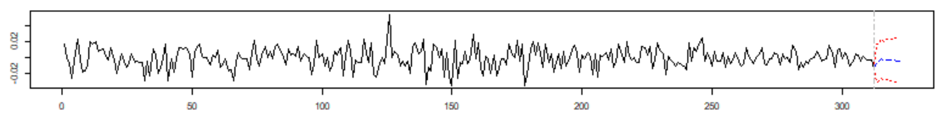


Figure 15: Brazil: Forecasts of Returns

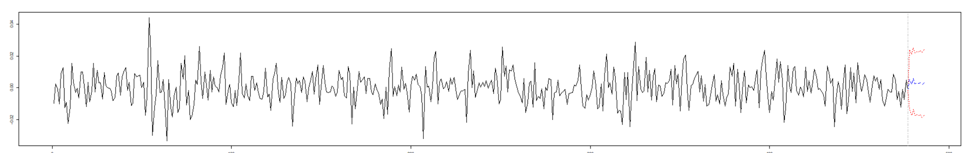


Figure 16: Chile: Forecasts of Returns to Sentiment Shocks

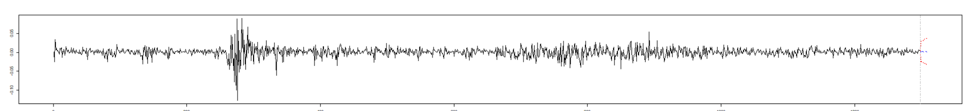


Figure 17: USA: Forecasts of Returns to Sentiment Shocks

Bibliography

Agapova, A. and M. Kaprielyan (2023). Diversification measures: Mutual fund family case. *International Review of Financial Analysis 90*, 102932.

Aitamurto, T. and S. C. Lewis (2013). Open innovation in digital journalism: Examining the impact of open apis at four news organizations. *New media & society 15*(2), 314–331.

Akey, P., V. Grégoire, and C. Martineau (2022). Price revelation from insider trading: Evidence from hacked earnings news. *Journal of Financial Economics 143*(3), 1162–1184.

Allen, D. E., M. McAleer, and A. K. Singh (2019). Daily market news sentiment and stock prices. *Applied Economics 51*(30), 3212–3235.

Arcuri, M. C., G. Gandolfi, and I. Russo (2023). Does fake news impact stock returns? evidence from us and eu stock markets. *Journal of Economics and Business 125*, 106130.

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *The journal of Finance 61*(4), 1645–1680.

Baloria, V. P. and J. Heese (2018). The effects of media slant on firm behavior. *Journal of Financial Economics 129*(1), 184–202.

Bengio, Y., R. Ducharme, P. Vincent, and C. Jauvin (2003). A neural probabilistic language model. *Journal of Machine Learning Research 3*, 1137—-1155.

Bernile, G., J. Hu, and Y. Tang (2016). Can information be locked up? informed trading ahead of macro-news announcements. *Journal of Financial Economics 121*(3), 496–520.

Brown, G. W. and M. T. Cliff (2005). Investor sentiment and asset valuation. *The Journal of Business 78*(2), 405–440.

Calomiris, C. W. and H. Mamaysky (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics 133*(2), 299–336.

Campbell, J. Y. (1991). A variance decomposition for stock returns. *The economic journal 101*(405), 157–179.

Caporin, M., L. Corazzini, and M. Costola (2019). Measuring the behavioural component of the s&p 500 and its relationship to financial stress and aggregated earnings surprises. *British Journal of Management 30*(3), 712–729.

Carlini, F., D. Cucinelli, D. Previtali, and M. G. Soana (2020). Don't talk too bad! stock market reactions to bank corporate governance news. *Journal of Banking & Finance 121*, 105962.

Chi, T.-L., H.-T. Liu, and C.-C. Chang (2023). Hedging performance using google trends–evidence from the indian forex options market. *International Review of Economics & Finance 85*, 107–123.

Chowdhury, S. G., S. Routh, and S. Chakrabarti (2014). News analytics and sentiment analysis to predict stock price trends. *International Journal of Computer Science and Information Technologies 5*(3), 3595–3604.

Cochrane, J. (2009). *Asset pricing: Revised edition*. Princeton university press.

Costola, M., O. Hinz, M. Nofer, and L. Pelizzon (2023). Machine learning sentiment analysis, covid-19 news and stock market reactions. *Research in International Business and Finance 64*, 101881.

Cresson, J. E. (2002). R^2: A market-based measure of portfolio and mutual fund diversification. *Quarterly Journal of Business and Economics 41*(3/4), 115–143.

Dougal, C., J. Engelberg, D. Garcia, and C. A. Parsons (2012). Journalists and the stock market. *The Review of Financial Studies 25*(3), 639–679.

Eachempati, P. and P. R. Srivastava (2021). Accounting for unadjusted news sentiment for asset pricing. *Qualitative Research in Financial Markets 13*(3), 383–422.

Elbannan, M. A. (2015). The capital asset pricing model: an overview of the theory. *International Journal of Economics and Finance 7*(1), 216–228.

Elton, E. J. and M. J. Gruber (2013). Mutual funds. In *Handbook of the Economics of Finance*, Volume 2, pp. 1011–1061. Elsevier.

Fang, L. and J. Peress (2009). Media coverage and the cross-section of stock returns. *The journal of finance 64*(5), 2023–2052.

Fernandez, R., B. P. Guizar, and C. Rho (2021). A sentiment-based risk indicator for the mexican financial sector. *Latin American Journal of Central Banking 2*(3), 100036.

Fraiberger, S. P., D. Lee, D. Puy, and R. Ranciere (2021). Media sentiment and international asset prices. *Journal of International Economics 133*, 103526.

Garcia, D., X. Hu, and M. Rohrer (2023). The colour of finance words. *Journal of Financial Economics 147*(3), 525–549.

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–574.

Goetzmann, W. N. and M. Massa (2003). Index funds and stock market growth. *The Journal of Business 76*(1), 1–28.

Goldman, E., N. Gupta, and R. Israelsen (2024). Political polarization in financial news. *Journal of Financial Economics 155*, 103816.

Gupta, K. and R. Banerjee (2019). Does opec news sentiment influence stock returns of energy firms in the united states? *Energy Economics 77*, 34–45. Energy Finance 2017.

Hamilton, J. T. (1995). Pollution as news: Media and stock market reactions to the toxics release inventory data. *Journal of environmental economics and management 28*(1), 98–113.

Heston, S. L. and N. R. Sinha (2017). News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal 73*(3), 67–83.

Hochreiter, S. and J. Schmidhuber (1997). Long short-term memory. *Neural computation 9*(8), 1735–1780.

Hsu, Y.-J., Y.-C. Lu, and J. J. Yang (2021). News sentiment and stock market volatility. *Review of Quantitative Finance and Accounting 57*(3), 1093–1122.

Hung, M.-C., P.-H. Hsia, X.-J. Kuang, and S.-K. Lin (2024). Intelligent portfolio construction via news sentiment analysis. *International Review of Economics & Finance 89*, 605–617.

Januário, B. A., A. E. d. O. Carosia, A. E. A. da Silva, and G. P. Coelho (2021). Sentiment analysis applied to news from the brazilian stock market. *IEEE Latin America Transactions 20*(3), 512–518.

Kabbani, T. and E. Duman (2022). Deep reinforcement learning approach for trading automation in the stock market. *IEEE Access 10*, 93564–93574.

Kamal, J. B. and M. Wohar (2023). Heterogenous responses of stock markets to covid related news and sentiments: Evidence from the 1st year of pandemic. *International Economics 173*, 68–85.

Karppi, T. and K. Crawford (2016). Social media, financial algorithms and the hack crash. *Theory, culture & society 33*(1), 73–92.

Knill, A., B. Liu, and J. J. McConnell (2022). Media partisanship and fundamental corporate decisions. *Journal of Financial and Quantitative Analysis 57*(2), 572–598.

Kogan, S., T. J. Moskowitz, and M. Niessner (2023). Social media and financial news manipulation. *Review of Finance 27*(4), 1229–1268.

Li, Q., J. Wang, and L. Bao (2018). Do institutions trade ahead of false news? evidence from an emerging market. *Journal of Financial Stability 36*, 98–113.

Li, Y. (2021, January 30). Gamestop, reddit and robinhood: A full recap of the historic retail trading mania on wall street. https://www.cnbc.com/2021/01/30/gamestop-reddit-and-robinhood-a-full-recap-of-the-historic-retail-trading-mania-on-wall-street.html. Retrieved from CNBC.

Malo, P., A. Sinha, P. Korhonen, J. Wallenius, and P. Takala (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology 65*(4), 782–796.

Marty, T., B. Vanstone, and T. Hahn (2020). News media analytics in finance: a survey. *Accounting & Finance 60*(2), 1385–1434.

Neal, R. and S. M. Wheatley (1998). Do measures of investor sentiment predict returns? *Journal of financial and quantitative analysis 33*(4), 523–547.

Panna, M. (2017). Note on simple and logarithmic return. *APSTRACT: applied studies in agribusiness and commerce 11*, 127–136.

Perold, A. F. (2004). The capital asset pricing model. *Journal of economic perspectives 18*(3), 3–24.

Shah, D., H. Isah, and F. Zulkernine (2018). Predicting the effects of news sentiments on the stock market. In *2018 IEEE International Conference on Big Data (Big Data)*, pp. 4705–4708. IEEE.

Shiller, R. J. et al. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review 71*, 421–436.

Siahaan, A. P. U. et al. (2018). Autoregression vector prediction on banking stock return using capm model approach and multi-factor apt (ijciet). *International Journal of Civil Engineering and Technology (IJCIET) 9*(9), 1093–1103.

Singh, A., J. Singh, and A. Ghosal (2022). Impact of social media on stock market-a case of sentiment analysis using vader. In *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, Volume 1, pp. 300–304. IEEE.

Sinha, N. R. (2016). Underreaction to news in the us stock market. *Quarterly Journal of Finance 6*(02), 1650005.

Smales, L. A. (2016). Time-varying relationship of news sentiment, implied volatility and stock returns. *Applied Economics 48*(51), 4942–4960.

Sun, F., S. Dutta, H. Huang, and P. Zhu (2021). News media and insider trading profitability: an emerging country perspective. *Emerging Markets Review 49*, 100795.

Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *science 211*(4481), 453–458.

Uhl, M. W. (2014). Reuters sentiment and stock returns. *Journal of Behavioral Finance 15*(4), 287–298.

Wan, X., J. Yang, S. Marinov, J.-P. Calliess, S. Zohren, and X. Dong (2021). Sentiment correlation in financial news networks and associated market movements. *Scientific reports 11*(1), 3062.

Xing, F., D. H. Hoang, and D.-V. Vo (2020). High-frequency news sentiment and its application to forex market prediction. In *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)*.

Xing, F. Z., E. Cambria, and R. E. Welsch (2018). Natural language based financial forecasting: a survey. *Artificial Intelligence Review 50*(1), 49–73.

Xu, Y., C. Liang, Y. Li, and T. L. Huynh (2022). News sentiment and stock return: Evidence from managers' news coverages. *Finance Research Letters 48*, 102959.

Yoon, S., K. Park, M. Lee, T. Kim, M. Cha, and K. Jung (2021). Learning to detect incongruence in news headline and body text via a graph neural network. *IEEE Access 9*, 36195–36206.

Zong, H., S. Wu, and G. Wei (2022). Sentiment analysis of news on the stock market. In *LISS 2021: Proceedings of the 11th International Conference on Logistics, Informatics and Service Sciences*, pp. 284–296. Springer.