Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis — Econometrics and Economics

Quantifying News Sentiment During the COVID-19 Crisis: A Zero-Shot Classification Approach and Its Impact on U.S. Household Stock Market Expectations

Mia Sin (561495)

Abstract

This study investigates the impact of news sentiment during the first year of the COVID-19 pandemic on U.S. households' expectations regarding the one-year-ahead stock market returns. Given the unprecedented nature of the pandemic and its extensive media coverage, understanding the influence of news sentiment in shaping household stock market expectations is crucial. This research employs a novel sentiment analysis approach using zero-shot classification with a pre-trained Bidirectional and Auto-Regressive Transformers (BART) model on two distinct corpora of news articles, focusing respectively on economic and health-related news. Two methods are explored to construct sentiment indices: one based on the proportion of articles labelled as positive, and another incorporating the polarity and sentiment intensity as measured by the confidence scores assigned by the zero-shot classifier. Subsequently, reduced-form ordinary least squares (OLS) regressions are employed to examine the relationship between news sentiment and the population average of the level, the population average of uncertainty, and cross-sectional heterogeneity of households' stock market expectations, using data from the Health and Retirement Study (HRS). The findings reveal that positive economic and health news sentiment tends to lower expectations for significant returns, while also reducing disagreement and decreasing uncertainty among households. This research contributes to the literature by introducing a novel method for constructing sentiment indices using zero-shot classification and validating its application, and providing empirical evidence on the influence of news sentiment on household stock market expectations during the COVID-19 crisis.

Supervisor:	Prof. dr. Robin Lumsdaine
Second assessor:	Dr. Erik Kole
Date:	1st July 2024

1 Introduction

The onset of the COVID-19 pandemic in March 2020 precipitated a historical crash of the Dow Jones Industrial Index, dropping over $6,400$ points ($\sim 26\%$ decline) in less than two weeks. This crash was a reaction to the U.S. government's response to the outbreak of the coronavirus (COVID-19), which included strict quarantines for the population and a shutdown of the majority of business activities to curb the spread of the virus. Monetary authorities also implemented intensive policy measures to relieve the stress on financial markets, among which the zero-percent interest rate policy announced by the Federal Reserve (Zhang et al., 2020). The pandemic and the policy responses of the government and other institutions created a high level of uncertainty due to scepticism about their effectiveness, and the associated economic losses caused markets to become highly volatile. Gormsen & Koijen (2020) find that financial policy interventions may create disparities between short-term and long-term expectations of investors. Similarly, Giglio et al. (2021) find that while stock markets began to recover following the crash, short-term pessimism among investors persisted, leading to increased disagreement regarding future stock market returns. Thus, while stock markets, including the Dow Jones, rebounded in the following months, uncertainty among investors remained. As these investor sentiments evolved, this uncertainty spilt over to households through the widespread media coverage, impacting their behaviour and expectations regarding the stock market. Both investors and households were influenced by the information disseminated during this period (Salisu & Vo, 2020), making it important to study how media narratives shaped their expectations regarding the stock market.

Understanding the dynamic between information flows and expectation formation requires examining the role of media narratives. Following the initial market crash, the rapid dissemination of information–and misinformation–by news outlets and on social media became a pivotal factor in shaping public perception and opinion about the government's response to the pandemic (Malecki et al., 2021). Moreover, Broadstock & Zhang (2019) show that, although investors traditionally turned to financial news, extracting information from social media posts and news articles has become an increasingly large part of financial decision-making. Particularly, due to the widespread and readily available nature of online media, it can easily magnify market sentiment and cause extreme price fluctuations (Costola et al., 2023). The COVID-19 pandemic has highlighted the critical role of online media in shaping economic behaviour (Broadstock & Zhang, 2019). Media coverage during the pandemic influenced not only investor sentiment but also household expectations regarding the economy. As public discourse rapidly adapted to new information about the virus, evolving government policies, and changing global economic conditions, understanding the extent of this influence on household expectations is crucial.

Unlike the 2008 financial crisis, which primarily centred around financial instability, the COVID-19 crisis was fundamentally a health crisis, characterised by the rapid spread of the virus, high mortality rates, and unprecedented public health measures. During the 2008 financial crisis, the government's response focused largely on stabilising financial institutions and restoring market confidence (Spatt, 2020). In contrast, the COVID-19 crisis induced widespread panic due to its immediate threat to public health, prompting extreme measures such as nationwide lockdowns, social distancing mandates, and travel bans. This difference significantly influenced public perception of the crisis and affected the level of fear and anxiety regarding both personal and public health. Moreover, the economic fallout of the COVID-19 stock market crash necessitated swift and comprehensive policy actions. Governments worldwide implemented substantial monetary policies and fiscal intervention; in the U.S., the Federal Reserve instituted a zero-percent interest policy (Zhang et al., 2020), and adopted the Coronavirus Aid, Relief, and Economic Security (CARES) Act, which was an economic stimulus package providing direct payments to individuals and businesses. While the effect of these measures on the public helped mitigate immediate economic distress, economic anxieties about the future remained. Therefore, when examining the effects of the COVID-19 crisis, it is crucial to account for both economic and health-related fears, as this dual impact inherently differentiates the pandemic from the financial crisis of 2008.

This paper analyses how news sentiment during the COVID-19 crisis shaped U.S. households' expectations regarding the future of the stock market by employing a two-part methodology. Firstly, sentiment indices are constructed from news articles using large language models (LLMs) and a novel zero-shot classification approach. Secondly, simple regression analyses are estimated using survey data from the Health and Retirement Study[1] (HRS) to examine the relationship between households' stock market expectations and news sentiment.

The sentiment indices are constructed using two corpora of news articles sourced during the first year of the COVID-19 crisis from five major U.S. newspapers. The corpora each have a distinct subject focus to capture more variation in news sentiment: the first corpus focuses on the U.S. economy and markets during the COVID-19 crisis, and the second corpus relates to health developments in the U.S. during the COVID-19 crisis. The sentiment analysis is performed using zero-classification, which leverages an advanced large language model BART (Bidirectional Encoder and Auto-Regressive Transformer) to classify news articles into positive and negative sentiment categories without needing to train the model on extensive labelled data. Then, similar to the analysis of Hudomiet et al. (2011) on the 2008 mortgage crisis, this study examines the effect of news sentiment on proxies of the subjective mean, subjective uncertainty, and cross-sectional heterogeneity of households' stock market expectations. In particular, the probabilistic survey questions regarding expectations on the stock market are related to the variables of interest and utilised as proxy variables. Reduced-form regressions are employed to examine the relationship between the constructed sentiment indices, other relevant control variables, and households' expectations of future stock returns.

This study demonstrates the potential for the direct use of zero-shot classification in sentiment analysis, particularly for news articles. The zero-shot classifier shows higher confidence in identifying positive articles but generally classifies more articles as negative compared to the benchmark index. This suggests that zero-shot classification might be more sensitive to affirmative language in their confidence to classify articles. Additionally, constructing sentiment indices based on sentiment intensity, as measured by the confidence score outputted by the zero-shot classifier, seems to be a valid measure. Furthermore, the construction of sentiment indices appears domain-specific, influenced by the expressiveness of the language used. Economic news sentiment was better captured by the score-based sentiment index, possibly reflecting stronger language, whereas health news sentiment was more accurately captured by the indicator-based

[1] The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

index, which measures the proportion of positive articles. This suggests that the indicator-based approach, which measures the proportion of positive articles, may be more reliable when language is less affirmative, while the score-based approach, which measures both polarity and intensity of sentiment, performs better with stronger language. This research contributes to the literature by exploring a novel approach to sentiment analysis using zero-shot classification on datasets comprised of news articles from major U.S. newspapers. It considers two different measures of sentiment, namely the proportion of positive articles (i.e., polarity) and both polarity and sentiment intensity, providing evidence that zero-shot classification could be a valid approach to sentiment analysis.

The regression analyses provide an indication of how news sentiment influences household expectations. The economic and health sentiment indices are both negatively related to the level, uncertainty, and disagreement of expectations for large stock market gains, suggesting that positive news may lead to lower expectations for substantial returns, but also less disagreement and uncertainty among households. Additionally, while economic sentiment negatively correlates with the level, uncertainty, and disagreement of expectations for overall gains and losses, health sentiment shows positive relationships, underscoring the duality of news sentiment in affecting household stock market expectations during the COVID-19 crisis. These findings contribute to the literature by highlighting the distinct impacts of economic and health news on household expectations, thus enhancing the understanding of sentiment analysis in the context of news media.

The remainder of the paper is structured as follows. Section 2 provides a comprehensive overview of existing literature, exploring the effects of the COVID-19 crash on both investors' and households' beliefs and expectations regarding the stock market, delving into sentiment analysis and elaborating on large language models and zero-shot classification. The methods are structured similarly to Costola et al. (2023), namely by first discussing the construction of the sentiment indices based on news articles published during the first year of the COVID-19 crisis in Section 3, and then exploring the relationship between the COVID-19 news sentiment indices and the U.S. households' stock market expectations in Section 4. Section 5 describes the news corpora used for sentiment analysis and delves into the probabilistic survey data and the control for the regression analysis. Section 6 presents empirical findings on the applicability of zero-shot classification for sentiment analysis, the validity of the approaches to constructing sentiment indices, and the influence of news sentiment on households' stock market expectations. Finally, Section 7 concludes on the insights gained from the research, discusses its limitations, and suggests directions for further research.

2 Literature Review

The COVID-19 crisis had a widespread impact, affecting not only individuals and households but also financial institutions and the government. Particularly, the pandemic had broad implications on household behaviour and expectations, highlighting the role of information in shaping beliefs about stock market recovery. Hanspal et al. (2021) discuss the implications of U.S. households' exposure to the COVID-19 stock market crash on the formation of expectations. They find that beliefs about the recovery of the stock market are causally related to households' expectations regarding their wealth and investment plans. Furthermore, Hanspal et al. (2021) find that the effects of the crash on stock market expectations are not homogeneous across different groups, which underscores the finding of Hudomiet et al. (2011) for the 2008 mortgage crisis. In particular, Hudomiet et al. (2011) delve into the differences of expectations of future stock market returns between stockholders and non-stockholders, informed and uninformed respondents, and respondents with above and below-average cognition. The findings suggest that the expectations of stockholders and informed respondents tend to be higher and less uncertain, although the expectations of stockholders are more strongly affected by the stock market crash than the expectations of respondents who follow the stock market. Additionally, Hanspal et al. (2021) find that respondents who are biased with information regarding the duration of a previous longer stock market crash become more pessimistic regarding the returns on the stock market in the coming years compared to the control group who did not receive any information. The opposite was the case for respondents who received information on the duration of a shorter historical stock market crash as they became more optimistic regarding future outcomes on the stock market. The findings of Hanspal et al. (2021) highlight the power of information in terms of historical events to stimulate individuals' expectations and beliefs regarding the future of the stock market and suggest that this could be due to the straightforward and understandable nature of the information provided.

Directly following the interest rate cuts by the Federal Reserve at the beginning of March 2020, Binder (2020) surveyed consumers to gauge their concerns regarding the pandemic, their awareness of the announcement made by the Federal Reserve, and their macroeconomic expectations. She finds that consumer characteristics affect concerns regarding the COVID-19 crisis. For instance, individuals who own stocks or follow the stock market appear to stay informed about coronavirus-related news, exhibit greater concern about the COVID-19 crisis, and are more likely to participate in the survey. Similarly, individuals who read newspapers tend to show heightened levels of concern about the pandemic. Binder (2020) theorises that informed individuals, who tend to be more associated with such characteristics, are more concerned about the implications of the pandemic. Binder (2020) also shows the power of information in affecting individuals' beliefs. She provided half of the respondents with brief information on the coronavirus and found that these respondents had elevated health-related concerns. Simply knowing about the effects of the coronavirus negatively affects individuals' beliefs.

Regarding investor beliefs and portfolio dynamics, Giglio et al. (2021) analyse investor behaviour before, during, and after the COVID-19 stock market crash. They show that, following the crash in March 2020, investors' expectations regarding the one-year-ahead stock returns declined, while beliefs regarding the long-run outcomes on the stock market remained stable. Furthermore, Giglio et al. (2021) show that around the time of the crash, investors anticipated a greater likelihood of extreme stock market shocks. Finally, they show that disagreement among investors about their beliefs increased following the crash, persisting even through the partial recovery of the stock market in subsequent months. The presence of persisting disagreement after the COVID-19 crash is similar to the pattern observed following the mortgage crash in late 2008. Hudomiet et al. (2011) find a significant long-term increase in disagreement, particularly among stockholders.

These findings underscore the impact of the COVID-19 crisis and information dissemination on investor and household expectations about the stock market. Recent research has built on this foundation by analysing sentiment to further explore the dynamics of market expectations. Since the COVID-19 crisis, there has been much research relating to the effects of media sentiment on stock market performance during the pandemic, which is inadvertently related to investor and household expectations of the stock market. While social media sentiment is widely researched, there is ample research on news sentiment and stock market performance. Salisu & Vo (2020) delve into the effect of health news sentiment on stock market returns during the COVID-19 crisis, and find that health-related news has a significant negative effect on stock market returns. Moreover, they find that combining financial and health news when analysing news sentiment improves forecasting results on the stock market. Costola et al. (2023) examine the relationship between financial news sentiment during the COVID-19 pandemic and stock market returns by analysing sentiment from online newspapers. They find that there is a positive correlation between financial news sentiment and the S&P 500, which implies that when news outlets are more positive about the stock market, the stock market experiences higher returns. This suggests that positive stock market news improves the expectations of investors and households on stock returns. The effects of news sentiment on investors' beliefs seem to be specific to the news domain considered.

Although sentiment analysis has been actively researched for decades, it has become more popular and widely used in recent years due to the rise of opinion-rich social media platforms such as Twitter (Broadstock & Zhang, 2019; Mehta et al., 2021; Monselise et al., 2021; Shofiya & Abidi, 2021). The purpose of sentiment analysis is to analyse, identify, and extract thoughts, opinions, and subjectivity to categorise text based on its polarity, which indicates whether the expressed sentiment is positive, negative, or neutral. Sentiment analysis is a widely used application in natural language processing (NLP) and in recent years has often employed large language models to handle large textual datasets. The literature on sentiment analysis primarily divides its approaches into three categories: lexicon-based, machine learning based, and deep learning based (Wankhade et al., 2022).

Lexicon-based methods utilise a predefined set of words with associated sentiment scores to determine the polarity in a text. Although these methods are straightforward and widely applied, they often struggle with contextual understanding and do not generalise well across different domains. In recent years, combining lexicon-based approaches with machine learning models has improved their effectiveness by leveraging the interpretability of lexicons and the predictive power of machine learning techniques (Sánchez-Rada & Iglesias, 2019). Currently, machine learning techniques dominate the field of sentiment analysis (Araque et al., 2017). These methods involve training supervised learning models on domain-specific labelled data to classify sentiment in text in the same domain. Commonly used machine learning techniques include Naive Bayes, Support Vector Machines, and k-Nearest Neighbours, which typically exhibit high accuracy and robust performance across various metrics (Kadhim, 2019). However, a major limitation of these methods is their reliance on substantial amounts of labelled data for training supervised classifiers. Labelled training data is expensive and labour-intensive to obtain, especially when manual labelling and human verification are required (Wankhade et al., 2022).

In recent years, deep learning has demonstrated strong performance in natural language processing tasks, including sentiment analysis. Before zero-shot classification was introduced, Collobert et al. (2011) investigated the use of unlabelled training data to learn internal representations capable of performing a multitude of NLP tasks, including sentiment analysis. Their deep learning models exhibit excellent performance, prompting further research into the application of deep learning in sentiment analysis. Recently, the focus has shifted from deep learning methods such as Long Short-Term Memory Networks and Recurrent Neural Networks to large language models like Transformer models. The Transformer model, introduced by Vaswani et al. (2017), employs an attention mechanism that significantly improves contextual understanding. Since their introduction, Transformer models have been widely adopted, most notably for the Bidirectional Encoder Representations from Transformers (BERT) model and the Bidirectional and Auto-Regressive Transformers (BART) model.

The BERT model, introduced by Devlin et al. (2018), captures context from both directions, enhancing its understanding of language. The model is pre-trained on large amounts of unlabelled data and subsequently fine-tuned on downstream tasks using (smaller) labelled datasets. In recent years, BERT has been applied to many natural language processing tasks, such as sentiment analysis. For example, Sousa et al. (2019) utilise a BERT model fine-tuned on a dataset with labelled stock news articles to perform sentiment analysis and provide insights into how sentiment in news articles can predict movements on the stock market. A notable adaptation of the BERT model is RoBERTa, which is an optimised variant that addresses the undertraining issues in BERT introduced by Liu et al. (2019). After improving the pre-training of the model, Liu et al. (2019) report state-of-the-art performance that is superior to BERT on various tasks. Such findings also hold for NLP tasks such as sentiment analysis. For instance, Adoma et al. (2020) report that RoBERTa outperforms BERT in terms of accuracy for emotion detection on their dataset. The BART model, introduced by Lewis et al. (2019), combines elements from encoder models such as BERT and decoder models like Generative Pre-training Transformers (GPT) to understand and generate text. Lewis et al. (2019) demonstrate that while BART is most effective when fine-tuned for generation tasks, it also exhibits strong performance in comprehension tasks. Additionally, they show that BART performs comparably to RoBERTa for most discriminative tasks, suggesting that the additional decoder layers do not comprise the classification performance of BART.

Recently, researchers have been exploring transfer learning techniques, where a model pre-trained on a specific task in one domain can be applied to a different task in another domain. Zero-shot classification is such a transfer learning technique: it leverages pre-trained models, such as BART, to classify text without the model having previously seen the classes during training. Early literature on zero-shot classification primarily focused on applications in image recognition and classification. However, in recent years, zero-shot learning has gained attention in the context of natural language processing. Notably, Pushp & Srivastava (2017) proposed a novel zero-shot learning approach for text categorisation. By reframing text categorisation as a binary classification task, their model learns the relatedness between sentences and tags, enabling generalisation to unseen classes and datasets. This approach facilitates scaleable and flexible text categorisation systems that can adapt to new data distributions and categories, reducing the need for extensive retraining or manual annotation. Although zero-shot classification does not yet achieve the accuracy of fully supervised models (Pushp & Srivastava, 2017), it represents a significant step towards more generalised and adaptable deep learning systems for NLP.

Yin et al. (2019) further explore zero-shot learning for text classification (0SHOT-TC), expanding its application to tasks such as emotion detection and situation detection, demonstrating that 0SHOT-TC can be applied to various dimensions of text interpretation. They propose treating 0SHOT-TC as a textual entailment problem, which involves determining if a text entails a hypothesis generated from a label. Their paper serves as a benchmark for zero-shot text classification and introduces a novel approach using textual entailment to structure the classification. Currently, many pre-trained models, including the BART large model, treat 0SHOT-TC as a textual entailment problem.

While Yin et al. (2019) lay the groundwork for expanding zero-shot learning applications beyond emotion detection and situation detection, there is limited literature available on applying zero-shot classification to sentiment analysis. Tesfagergish et al. (2022) propose a novel two-stage method combining both zero-shot learning for emotion detection with supervised learning methods for sentiment classification. Their research demonstrates that first detecting emotions can significantly improve the accuracy of sentiment analysis. Although their results are robust for sentiment databases characterised by a high degree of emotional content, such as reviews and tweets, their relevance for news articles remains limited. News articles typically aim to be relatively objective in reporting and often do not convey strong emotions, making emotion detection more challenging. This research contributes to the literature by exploring zero-shot classification for sentiment analysis on datasets comprised of news articles from major U.S. newspapers.

3 COVID-19 news sentiment construction

This section explores the application of large language models in sentiment analysis, focusing on the Transformer-based models BERT and BART. These models are leveraged to perform zero-shot classification, enabling sentiment classification in the text corpora without using labelled training data. Data collection involves sourcing news articles related to the COVID-19 pandemic from major U.S. newspapers, which serve as the input for the sentiment analysis. This research constructs sentiment indices from these text corpora, employing both indicator-based and score-based approaches to capture sentiment polarity and intensity. The resulting indices are validated against a benchmark sentiment index derived from the *COV19Tweets* dataset to ensure the reliability of the constructed indices in reflecting sentiment during the COVID-19 crisis (Lamsal, 2021).

3.1 Data collection

To capture news sentiment during the COVID-19 crisis, this study compiles a set of news articles published in the U.S. spanning from March 2020 to May 2021, focusing exclusively on news articles instead of social media posts due to the age range of the HRS respondents (51 to 100 years). Generally, individuals in this age group use social media less, particularly platforms such as Twitter that are often used for sentiment analysis (Leist, 2013), thus, it is more relevant to include news articles as the likelihood is higher that the respondents read the news than follow social media. While excluding social media data limits the extent to which the sentiment can capture public opinion, news articles are an important source of information for individuals to form their views and expectations on (Mutz & Soss, 1997).

This study focuses on five major U.S. newspapers, namely The New York Times, New York Post, The Wall Street Journal, The Washington Post, and USA Today. These outlets were selected for their national focus, broad readership, and significant influence on public opinion based on circulation (Rick et al., 2011; Tetlock, 2007), thereby increasing the likelihood that HRS participants have read articles from these sources. While focusing on these five newspapers might introduce selection bias, their wide circulation and strong reputation make them representative of mainstream news consumption in the U.S. Furthermore, these outlets cover a diverse range of topics, ensuring a comprehensive coverage of economic and health-related news during the COVID-19 crisis. To further mitigate potential selection bias, future research could include a wider array of news sources, such as regional newspapers and online news platforms. Including more news sources would provide improve the diversity of the media content and potentially capture a broader spectrum of public sentiment. By focusing on widely read newspapers, the likelihood increases that the content aligns with the information sources of the HRS respondents, enhancing the representativeness of the sentiment indices.

To specify a search query that adequately captures sentiment about the pandemic, this study searches for keywords directly related to the COVID-19 crisis, including 'COVID', 'COVID-19', 'COVID19', 'corona', 'coronavirus', 'pandemic', or 'epidemic'. These keywords were chosen based on their prevalence and relevance during the first year of the pandemic, particularly on social media, in news articles, and during press conferences (Asif et al., 2021). While there are many additional words related to the COVID-19 crisis, such as 'quarantine' or 'lockdown', they were not included in the primary search terms. The rationale behind this decision is that articles that discuss these topics generally also mention terms like 'COVID-19' or 'coronavirus', ensuring that relevant articles are sourced without explicitly including every coronavirus-related term. Including too many keywords could dilute the focus of the corpus and introduce noise, potentially reducing the accuracy of the sentiment analysis.

The news articles are sourced from Factiva, which is an online archive of over 33,000 news, data, and information sources. Although its focus is primarily on business-related news, it has a comprehensive coverage of a variety of topics, including politics, current events, and trends. To capture a broader range of sentiment in news articles, this study focuses on two main subjects: economic news and health-related news. Articles categorised by Factiva under 'economic news' are selected to capture sentiment related to financial markets, economic policies, and business impacts. Focusing on economic news is crucial for understanding economic sentiment that could influence household expectations regarding the future of the stock market. Furthermore, articles categorised by Factiva under 'health' are chosen to capture sentiment related to public health measures, medical developments, and healthcare responses. The dual focus ensures a sentiment is captured comprehensively, reflecting both the economic and health-related impacts of the pandemic, thereby providing a holistic view of the COVID-19 crisis.

Additionally, articles are limited to a maximum of 512 words to align with the model requirements of BERT and to maintain computational feasibility. Accounting for article length is necessary due to the tokenisation constraints of BERT and to streamline data processing. Although focusing on shorter articles might exclude some detailed analyses, it ensures consistency across the dataset. Furthermore, duplicate articles are identified and manually removed by the author, ensuring that each article in the dataset is unique.

3.2 Sentiment analysis

3.2.1 Large language models

Large language models are advanced machine learning models capable of understanding and generating human language by processing extensive amounts of data. LLMs leverage deep learning techniques, specifically Transformer architectures, to process and analyse vast amounts of textual data. The Transformer model, introduced by Vaswani et al. (2017), is a model architecture that uses an attention mechanism to establish dependencies between inputs and outputs. Compared to traditional recurrent architectures, the Transformer model allows for more parallelisation. Since its introduction, the Transformer architecture has become a prominent choice for building LLMs due to its ability to capture the context of all words in a sentence, not only neighbouring words, using the self-attention mechanism. By assigning attention weights to weigh the importance of the words in the input sequence relative to a particular word in the sequence, the model can determine the relevance of each word in the input sequence, regardless of its position in the text. The model learns these weights during pre-training, allowing it to selectively focus on certain words more than on others when making predictions, facilitating a deeper understanding of the context of a text. Additionally, Transformers use positional encoding to understand the sequential order of words in the input text. They generate positional tags that represent each word's positions within the sequence and integrate these tags with the word embeddings. Consequently, the model is able to understand both the semantic meaning of each word and its respective position in the sequence. This study delves into the model specifics of two Transformer-based large language models, namely BERT in Section 3.2.2 and BART in Section 3.2.3.

3.2.2 BERT

BERT, or Bidirectional Encoder Representations from Transformers, is a Transformer-based model introduced by Devlin et al. (2018). It has revolutionised natural language processing by introducing a method for pre-training deep bidirectional models that simultaneously consider contextual information from both directions throughout all layers. Essentially, this enables BERT to capture both left and right contexts, which improves its understanding of the semantics and syntax of the language. BERT is pre-trained on two unsupervised tasks: masked learning modelling (MLM) and next sentence prediction (NSP). MLM removes a word in a sequence and the objective is to predict the missing word based on its context. The bidirectional element stems from its use of the words preceding and succeeding the masked word to infer the word and outputs a probability that this word fits given the context. MLM is illustrated on a sentence drawn randomly from a news article in the economic corpus[2]:

Sequence: "Top economists with the Federal Reserve Bank are predicting that the worst of the economic $[\text{MASK}]_1$ from the coronavirus outbreak could result in a $[\text{MASK}]_2$ unemployment rate than during the Great Depression."

Labels: $[\text{MASK}]_1$ = damage (0.182), $[\text{MASK}]_2$ = higher (0.792)

The aim in this example is to mask the words 'damage' and 'higher' and have the model predict which word would most likely fill the masked word. While the model is confident that $[\text{MASK}]_2$ should be 'higher', it is not as confident that $[\text{MASK}]_1$ should be 'damage', even though it predicted 'damage' with the highest probability.

Following MLM, BERT is trained on next-sentence prediction, which randomly draws a sequence of two sentences from the corpus and predicts whether the second sentence follows the first. NSP is illustrated below on the same news article as above:

Sentence 1: "Top economists with the Federal Reserve Bank are predicting that the worst of the economic damage from the coronavirus outbreak could result in a higher unemployment rate than during the Great Depression."

Sentence 2: "Economists from the Fed's St. Louis district said in a recent analysis that the United States will lose 47 million jobs, resulting in an unemployment rate of 32.1 percent, outpacing the 24.9 percent unemployment rate during the worst point of the Great Depression in the 1930s."

Sentence 3: "The projections also go beyond the 30 percent unemployment rate originally predicted by St. Louis Fed Chair James Bullard."

Sequence pair (1,2) will be passed to BERT labelled that the sentences follow each other since Sentence 2 follows Sentence 1 in the article. Contrarily, pairs (1,3) and (2,3) will be labelled that the sentences do not follow each other, since Sentence 3 is drawn from a random location in the article.

In their paper, Devlin et al. (2018) pre-train BERT on the BookCorpus and English Wikipedia, which contain 800 million and $2,500$ million words respectively. Then, the model is initialised with the pre-trained parameters, which are consequently fine-tuned using labelled data on different downstream tasks. Furthermore, Devlin et al. (2018) propose two adaptations to the model size: the BERT base and the BERT large, which differ in the number of layers, the hidden size, the number of attention heads, and the number of parameters. The selection between these two models typically involves a trade-off between the superior performance of the larger models and the faster computation time of the base models. BERT is designed to handle sequences of up to 512 tokens for both model sizes, since handling longer sequences incur significantly higher costs due to the quadratic relationship between attention and sequence length (Devlin et al., 2018).

[2] Jacobs, E. (2020, March 31). Federal Reserve predicts 32% unemployment rate thanks to coronavirus. *New York Post*

3.2.3 BART

BART, or Bidirectional and Auto-Regressive Transformers, is a Transformer-based encoder-decoder model introduced by Lewis et al. (2019). It uses a sequence-to-sequence Transformer architecture that leverages the strengths of BERT, incorporating the bidirectional encoder, and Generative Pre-training Transformer (GPT), applying the left-to-right decoder. This dual capability allows BART to excel in tasks requiring both understanding and generating text. For this research, only the elements from encoder models are relevant, as sentiment analysis is a discriminative task, thus, generative elements relating to GPT models are not discussed. BART is designed to reconstruct the original document from its corrupted version, forcing the model to learn and understand the context even when parts of the input are distorted or missing.

BART is pre-trained by corrupting text using noising functions and subsequently learning to reconstruct the original text by optimising a reconstructing loss (Lewis et al., 2019). Traditional models that correct errors in texts are trained on specific noising functions, such as the masked learning modelling used in BERT. Contrarily, BART can handle any type of corruption scheme, including hiding or deleting words, filling in missing words, and shuffling sentence (Lewis et al., 2019), which enhances its understanding of the surrounding context. This flexibility allows BART to grasp the overall meaning of the text more effectively than BERT. Lewis et al. (2019) find that the highest performance of pre-trained objectives is achieved by randomly shuffling the sentence order and using an in-filling scheme where arbitrary lengths of text are masked by a single token. Combining these two noising functions generalises the masked learning modelling and next-sentence prediction objectives used in BERT, requiring BART to learn about sentence structure and coherence in greater detail. The object of the training phase is to allow BERT to understand the context and structure of language comprehensively. The noising flexibility is an advantage of BART as it helps the model learn robustly across various types of textual corruption.

While the pre-trained BART model can perform many NLP tasks, fine-tuning is often necessary to adapt the model to specific applications such as zero-shot classification. This study uses the *bart-large-mnli* model, which is fine-tuned on the Multi-Genre Natural Language Inference (MNLI) dataset. The MNLI dataset is a corpus trained on a collection of over 400 thousand sentence pairs labelled with entailment information, enabling the model to understand and predict the relationship between sentences (Williams et al., 2017). Its size and data collection techniques are modelled similarly to the Stanford Natural Language Inference dataset (Bowman et al., 2015). The *bart-large-mnli* model is publicly available through the open-source machine learning platform Hugging Face, which provides access to thousands of pre-trained Transformer models.

3.2.4 Zero-shot classification

Generally, natural language processing such as sentiment analysis requires large amounts of labelled data for training a supervised learning model. However, labelled data is generally costly to obtain as it is often time-consuming and labour-intensive to construct, and in recent years, transfer learning has become a popular approach to solving this data constraint (Ruder et al., 2019). Zero-shot classification is an instance of transfer learning; it leverages pre-trained language models to classify unseen data in a different domain, which is particularly useful in cases where the amount of labelled data is small or nonexistent. While zero-shot learning was initially applied to image processing to predict unseen images, it has been adapted to handle NLP tasks such as text classification in recent years (Ruder et al., 2019). In particular, this research applies an approach to text classification using zero-shot classification introduced by Pushp & Srivastava (2017).

This study utilises a Hugging Face pipeline, which is a high-level abstraction that simplifies the application of pre-trained models for various tasks, such as text classification. The pipeline handles loading the model, pre-processing the input, executing the model, and post-processing the output. During the pre-processing phase, the pipeline converts the input text and candidate labels into a suitable format for the model. Typically, this involves tokenisation and encoding of the textual data. Tokenisation is the process of breaking down a text into smaller units called tokens and subsequently, encoding converts these tokens into numerical values that the model can process. Then, the encoded input is fed into the *bart-large-mnli* model, which evaluates the input text against each candidate label by constructing a hypothesis for each label. For example, for the label 'positive', the model might construct the hypothesis 'This text expresses positive sentiment' and then assess the likelihood of this hypothesis given the input text. For each hypothesis, the model generates a score that represents the likelihood that the input text corresponds to each label. Subsequently, the pipeline converts each score into a probability, providing each label with the corresponding probability that indicates how likely the text belongs to that category. The label with the highest probability is generally considered the most appropriate classification. By using a pipeline, this study leverages the capabilities of the BART large model for classifying text as 'positive' or 'negative' without the necessity to train a new model specific to this domain and classification task.

To perform a sentiment analysis using zero-shot classification, the candidate labels are given as 'positive' and 'negative' as input for the pipeline. I choose to neglect including the label 'neutral' in the set of candidate labels, for two reasons: (1) to achieve more distinct and precise classification outcomes, which simplifies the interpretation, and (2) to optimise the performance of the model by limiting the number of candidate labels, which reduces computation time. The range of each probability score is from 0 to 1, and the scores assigned to each category sum to 1. A score closer to 1 indicates that the model is more certain that the text should be classified as that category. Note that all scores are larger than 0.5, otherwise a class cannot be assigned.

The example below shows zero-shot classification using the *bart-large-mnli* large model. The aim is to classify the sentence as either positive or negative. The output scores indicate the probability that a sentence belongs to a certain candidate class.

> **Sequence:** "The coronavirus poses evolving risks to economic activity."
> **Candidate classes:** ['negative', 'positive']
> **Output:** [0.985, 0.015]

The model identifies 'negative' as the most likely class to categorise the sequence, despite not being trained on data from the COVID-19 crisis.

3.3 Construction sentiment indices

Leveraging the pre-trained *bart-large-mnli* model for zero-shot classification, each news article is classified as either 'positive' or 'negative', with an associated score that measures the probability this article will fall into the given class. The sentiment indices are constructed for every month in the sample, which reduces the granularity of the sentiment indices and disregards nuanced changes within a month. However, due to the periodicity of the HRS survey responses (see Section **??**), daily or weekly sentiment scores have no interpretation. In this section, I explore two approaches to constructing sentiment indices.

Firstly, a sentiment score is calculated by aggregating the number of positive articles in a given month and dividing that by the total number of articles in a month. The range of scores is between 0 and 1, where a score closer to 1 indicates more positive sentiment in the month. This index gives a straightforward measure of sentiment polarity in a given month, using the proportion of positive sentiment. The indicator-based sentiment index is formalised in Equation 1.

$$I_t = \frac{\sum_{i=1}^{N_t} I_{p,i}}{N_t} \quad \text{for } t \in \{1, ..., 15\}, \tag{1}$$

where I_t is the indicator-based sentiment score for month t, $I_{p,i}$ is a dummy variable taking on value 1 if article i is labelled as 'positive', and N_t is the total number of articles in month t.

Secondly, I introduce a novel approach to sentiment scoring, leveraging the probability score from zero-shot classification as a measure of sentiment polarity. The premise is that the score, which reflects the classifier's certainty about categorising a text into a specific sentiment category, correlates with the degree of that sentiment within the text. For instance, if the zero-shot classifier assigns a probability score of 0.95 to a text being categorised as 'positive', this confidence suggests that the content and tone of the text are indicative of positive sentiment. Consequently, this text can be considered more positive than another text with a lower probability score, e.g., 0.60. This approach implies that a higher score reflects both the classifier's certainty and signifies a stronger presence of the corresponding sentiment. A zero-shot classifier utilises contextual and linguistic cues to assess and predict the sentiment of unseen texts. A higher confidence score indicates that the text aligns more closely with the learned attributes of the target category. Therefore, the intensity of the sentiment can be inferred from the probability score itself. This approach not only captures the sentiment polarity but also reflects the intensity of the sentiment, providing a more nuanced measure. The score-based sentiment index is formalised in Equation 2.

$$S_t = \frac{\sum_{i=1}^{N_t} p_{i,t} + \sum_{j=1}^{N_t} (1 - q_{j,t})}{N_t} \quad \text{for } t \in \{1, ..., 15\}, \tag{2}$$

where S_t is the score-based sentiment score for month t, $p_{i,t}$ is the score of positive article i in month t, $q_{j,t}$ is the score of negative article j in month t, and N_t is the total number of articles in month t.

To validate the constructed sentiment indices, this research compares them with an established sentiment index to ensure their reliability and accuracy. Specifically, this research uses the *COV19Tweets* dataset[3], which contains sentiment scores of millions of tweets related to the pandemic (Lamsal, 2021). This dataset has been widely used in recent literature for sentiment analysis and opinion mining during the COVID-19 crisis (Monselise et al., 2021; Shofiya & Abidi, 2021). The *COV19Tweets* sentiment scores are generated using the Sentiment Analysis module of TextBlob, which is a lexicon-based approach utilising a predefined lexicon to determine the sentiment of each token based on its associated polarity and subjectivity scores (Lamsal, 2021).

To benchmark the constructed sentiment indices against the *COV19Tweets* sentiment scores, this study constructs sentiment indices for the *COV19Tweets* dataset in a similar manner and then compares the correlations between the new sentiment indices and the benchmark index. The indices, both for the economic and the health-related corpus, that show the strongest correlation with the benchmark index will be used for the subsequent regression analyses. A strong correlation indicates that the indices capture similar sentiment patterns as the benchmark index, which highlights their reliability in capturing the overall sentiment during the first year of the COVID-19 crisis. Given the widespread application of the *COV19Tweets* dataset, this study presumes its validity as a reliable benchmark for sentiment classification.

Each file in the *COV19Tweets* dataset represents approximately one day of tweets, each file containing over one million tweets. The sentiment scores are assigned on a scale of -1 to 1, where scores below 0 indicate that the tweet is labelled 'negative', scores above 0 indicate that the tweet is labelled 'positive', and scores of 0 are labelled as 'neutral'. Since the sentiment indices constructed in this research are on a scale of 0 to 1, the *COV19Tweets* scores must be transformed to this range. The following transformation is applied:

$$T_{i,t} = \frac{s_{i,t} + 1}{2} \quad \text{for } t \in \{1, ..., 439\}, \tag{3}$$

where $s_{i,t}$ is the sentiment score for tweet i on day t and $T_{i,t}$ is the transformed sentiment score for tweet i on day t.

A preliminary analysis of the *COV19Tweets* dataset reveals that there are disproportionately more tweets categorised as positive than negative. Thus, this study applies a similar approach to construct the benchmark index as the score-based index since an indicator-based index would not be as representative of the overall sentiment. Hence, the monthly sentiment score of the benchmark index corresponds to the mean of the transformed sentiment scores of the *COV19Tweets* dataset. As noted by Lamsal (2021), the files containing tweets between 1 March 2020 and 20 March 2020, and on 27 October 2020, 9 January 2021, and 17 April 2021 do not contain any sentiment scores and therefore are disregarded in this analysis.

4 COVID-19 news sentiment and households' stock market expectations

This section employs a reduced-form regression model to investigate the relationship between COVID-19 news sentiment and proxies for the average of the level, uncertainty, and cross-sectional heterogeneity of households' stock market expectations. The analysis focuses on prob-

[3] The files are available at https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset.

abilistic survey questions regarding one-year-ahead stock market returns from the Health and Retirement Study, and it provides a discussion of the various control variables included in the regression analyses.

4.1 Data collection

This research studies the impact of news sentiment during the COVID-19 crisis on U.S. households' expectations regarding future stock market returns using data from the 15^{th} wave of the Health and Retirement Study. This wave contains responses recorded between March 2020 and May 2021. The HRS is a national survey targeting individuals aged 51 and above, providing comprehensive data on retirement, health insurance, savings, and economic well-being. Specifically, this research employs the RAND HRS Longitudinal File 2020 and the 2020 RAND HRS Fat File. The longitudinal file is a cleaned dataset containing public information from the HRS Core and Exit interviews, in which the variables are consistently derived across survey waves. The fat file includes recorded responses to all survey questions, including the section on expectations. Both datasets record the survey responses with information on the interview dates tracked on a monthly basis.

Similar to Hudomiet et al. (2011), this research considers the answers to the probabilistic questions on the one-year-ahead returns in the stock market. These questions were introduced by the HRS in 2002 and have been asked in each wave since. The expectations on the future returns of the stock market are asked as a percentage on a scale from 0 to 100, where answering 0 indicates no chance and 100 indicates absolute certainty. Respondents were asked the following question, denoted by P_0:

'By next year at this time, what is the percent chance that mutual fund shares invested in blue chip stocks like those in the Dow Jones Industrial Average will be worth more than they are today?'

The answers to subjective probability questions are often centred '50' because respondent are unsure of the probability they would assign, thus, they settle on the average probability. However, similar to the 9^{th} survey wave used by Hudomiet et al. (2011), P_0 is left unanswered much more often than other probability questions in the survey, which could be due to two reasons: either respondents believe it is equally likely for the stock market to increase in the next year or they are simply not sure. To account for the uncertainty of '50' answers, the HRS included a follow-up question for the respondents who answered '50' to P_0:

'Do you think that it is about equally likely that these mutual fund shares will increase in worth as it is that they will decrease in worth by this time next year, or are you just unsure about the chances?'

If the respondent is unsure about their answer or does not know how to answer this question, their answers to the stock market questions are disregarded, as their uncertainty could be an indication of insufficient knowledge of the stock market to reasonably answer these questions.

Stock market questions are difficult to answer, and if an individual does not have stocks nor follows the stock market, they have not had to form an opinion on the one-year-ahead returns on the Dow Jones (Hudomiet et al., 2011), thus, they are not representative for the population in this study. Respondents who were sure about their response to P_0, including respondents who did not answer '50' to P_0 were asked two additional questions, denoted by P_1 and P_2 respectively, on a scale of '0' to '100':

'By next year at this time, what is the percent chance that mutual fund shares invested in blue chip stocks (like those in the Dow Jones Industrial Average) will have gained in value by more than 20 percent compared to what they are worth today?'

'By next year at this time, what is the percent chance that mutual fund shares invested in blue chip stocks (like those in the Dow Jones Industrial Average) will have fallen in value by more than 20 percent compared to what they are worth today?'

Essentially, P_1 and P_2 capture expectations on substantial market movement. A higher probability assigned to P_1 indicates that the respondent perceives a greater likelihood of significant gains in the stock market. Conversely, a higher probability assigned to P_2 suggests that the respondent anticipates a greater likelihood of large losses.

Following the premises of the analysis by Hudomiet et al. (2011), this study uses several variables to control for the variation in the dataset. Aside from common controls such as *gender*, *marital status*, *race*, *age*, and *years of education*, this research also accounts for *above-average cognition*, whether a respondent follows the stock market (*stockmarket*), and whether a respondent is a stockholder *stockholder*. Such variables are relevant to control for differences in knowledge about the stock market when dissecting the effects of the COVID-19 crisis on expectations regarding the future of the stock market. Most controls are included as dummy variables, whereas others are continuous variables (*age*, *years of education*). Excluding individuals below age 51 is an important consideration. While the HRS targets individuals over the age of 50, spouses and partners of respondents are also interviewed, regardless of age. This research excludes any individuals whose age is below 51 to keep the sample representative of the target group of the HRS, namely individuals nearing or in retirement. Moreover, *stockmarket* is a categorical variable taking on values 1 (very closely), 2 (somewhat closely), or 3 (not at all). Hudomiet et al. (2011) suggest analysing whether a respondent is informed about the stock market by aggregating the categories in which respondents follow the stock market, since the percentage of respondents in the first category is very small. The *informed* variable, which is a dummy variable that takes on value 1 if a respondent answered very closely or somewhat closely to the question on following the stock market, is highly correlated with *stockmarket* (Hudomiet et al., 2011), and has a better interpretability due to its binary nature. Finally, *above-average cognition* and *stockholder* are imputed from the HRS dataset. The variable *above-average cognition* is constructed as a dummy variable that takes on value 1 if the cognition score of a respondent is above the sample mean, and, since asset ownership is determined at the household level, the variable *stockholder* is constructed as a dummy variable that takes on value 1 for all members of the same household if the respondent owns assets.

Finally, the interview date of each respondent is mapped to the corresponding sentiment score of each sentiment index in a given month and year. This study utilises the endpoint of the interview date, as, according to the HRS, interviews not completed on the same day are predominantly conducted on the endpoint date.

4.2 Regression analysis

To examine the impact of news sentiment during the COVID-19 crisis on U.S. households' expectations regarding future stock market returns, this research utilises reduced-form ordinary least squares (OLS) regressions. The analysis regresses proxies for the subjective mean, subjective uncertainty, and cross-sectional heterogeneity (also referred to as disagreement) on the constructed sentiment indices and a set of control variables. The approach and proxies used in this study are similar to those utilised in the descriptive analysis conducted by Hudomiet et al. (2011). They explore descriptive results by estimating OLS regressions in a reduced-form model under the assumption of normally distributed returns. A reduced-form model directly relates the dependent variables to the independent variables, making it suitable for describing observed changes under various controls, but not for understanding the underlying mechanisms that drive those changes. In other words, while reduced-form regressions can identify correlations, they cannot establish causality. In contrast, structural estimation models aim to capture the fundamental theoretical relationships and mechanisms that drive changes in the dependent variables as a result of the independent variables, and are able to analyse causal effects. The results from structural estimation provide a comprehensive understanding of how changes in the independent variables influence the dependent variables through underlying mechanisms. Hudomiet et al. (2011) argue that the reduced-form model cannot estimate the impact of the stock market crash on variables of interest because the response variables are proxies for the variables of interest and the model does not account for survey noise. To address this limitation, they employ a structural model that accounts for rounding bias and uncertainty in survey responses. Their findings indicate that while the structural estimation results generally align with the reduced-form results, the structural estimation results provide more nuanced insights into the effects of the stock market crash on households' expectations by accounting for biases and noise that the reduced-form model overlooks

This research uses reduced-form OLS regressions to estimate the direct effects between the proxy response variables and the control variables due to the absence of a theoretical framework underpinning the regression model. Hudomiet et al. (2011) find that the estimates from the structural estimation were qualitatively similar to the reduced-form results in identifying key trends and implications, suggesting that reduced-form models can offer valid descriptive insights despite their limitations. Leveraging this finding, I argue that directly modelling the relationship between probabilistic questions and a set of control variables is appropriate when interpreting the results descriptively and indicatively. This approach is justified because, in the context of the analysis in this study, the primary goal is to capture the observed patterns and changes in expectations without the assumptions required for structural estimation.

The dependent variables in the OLS regressions in this study follow the premises of the descriptive analysis framework established by Hudomiet et al. (2011). Hudomiet et al. (2011) delve into the relationship between the probabilistic answers to P_0, P_{x+}, and P_{x-} and the variables of interest, namely the subjective mean, subjective uncertainty, and the heterogeneity of expectations. While this study does not further explore these relationships, it adopts their premises and constructs the same independent variables utilised by Hudomiet et al. (2011) in the OLS regressions. Note that P_{x+} and P_{x-} correspond to P_1 and P_2 in this study.

The level regressions in Equation 4 approximate how the control variables and the sentiment indices influence the population average of the level of expectations about stock market returns using the probability answers P_0, P_1, and P_2 directly. If individuals become more pessimistic on average about future stock returns, I expect the answers to P_0 and P_1 to decrease on average, and the answers to P_2 to increase on average.

Disagreement reflects the variability among respondents' expectations regarding future stock market returns. The absolute values of the residuals from the regression in Equation 4 are used as a proxy for the cross-sectional heterogeneity, or disagreement, of expectations denoted by $|u_{P_i}|$ for $i \in \{0, 1, 2\}$. The disagreement regressions are specified in Equation 5. An increase in disagreement implies that respondents have more dispersed expectations, and consequently, the residuals from the regression in Equation 4 would become more scattered on average, leading to an increase in their absolute values.

$$P_{i,j} = \alpha + \mathbf{w}_j'\boldsymbol{\beta}_{P_i} + \mathbf{x}_j'\boldsymbol{\gamma}_{P_i} + \epsilon_j \quad \text{for } i \in \{0, 1, 2\}, \tag{4}$$

$$\left|u_{P_{i,j}}\right| = \mathbf{w}_j'\boldsymbol{\beta}_{\left|u_{P_{i,j}}\right|} + \mathbf{x}_j'\boldsymbol{\gamma}_{\left|u_{P_{i,j}}\right|} + \epsilon_j \quad \text{for } i \in \{0, 1, 2\}, \tag{5}$$

where $\mathbf{w}_j$ is the (9×1) vector of control variables, including gender, marital status, race, ethnicity, age, education, cognition, informed about the stock market, stockholder, for respondent j and $\mathbf{x}_j$ is the (2×1) vector of sentiment indices for respondent j.

Uncertainty about future stock returns is inferred from the differences in respondents' probabilistic answers regarding overall expectations and expectations of large gains and losses. Specifically, uncertainty is derived from the probabilistic answers through the difference between P_0 and the expectations of gains and losses, given a threshold of 20%, represented by P_1 and P_2, respectively. Due to the relationship between these probabilistic answers and the variables of interest, the dependent variables are the inverse differences. For uncertainty regarding gains, this is represented by P_1–P_0, and for uncertainty regarding losses, it is represented by $(P_0 + P_2$–$100)$ (Hudomiet et al., 2011). The uncertainty regressions, specified in Equation 6, approximate the effects of the control variables on the population average of uncertainty about return. Higher values of uncertainty may indicate that respondents perceive the stock market as more volatile on average or expect more extreme price movements.

$$P_k = \mathbf{w}_j'\boldsymbol{\beta}_{P_k} + \mathbf{x}_j'\boldsymbol{\gamma}_{P_k} + \epsilon_j, \tag{6}$$

where P_k is defined as $(P_1 - P_0)$ for uncertainty about gains denoted by P_{k+}, and $P_0 + P_2 - 100$ for uncertainty about losses denoted by P_{k-}, $\mathbf{w}_j$ is the (9×1) vector of control variables, including gender, marital status, race, ethnicity, age, education, cognition, informed about the stock market, stockholder, for respondent j and $\mathbf{x}_j$ is the (2×1) vector of sentiment indices for respondent j.

Note that while $\mathbf{x}_j$ comprises both an economic and a health-related sentiment index, each regression model includes only one of these indices. Consequently, the coefficient for the excluded sentiment index is constrained to zero.

5 Data

5.1 Corpus data

This section analyses the summary statistics and trends in the economic and health news corpora during the first year of the COVID-19 crisis. Table 1 provides an overview of the articles in the corpora after removing duplicates to ensure the uniqueness of all articles. Noticeably, the sample size of the economic corpus is significantly smaller than that of the health corpus, 962 and 7052 articles, respectively. The smaller economic corpus may not be able to capture the full range of economic news sentiment during the COVID-19 crisis, which can limit the generalisability of the findings. Additionally, due to the smaller size, the corpus may not be representative of the entire scope of articles related to economic news during the first year of the pandemic and can be more sensitive to outliers. Importantly, comparing results between the economic and health corpora could be challenging due to the difference in size, as the imbalance may skew the comparative analysis of sentiment effects on household expectations. For instance, the smaller economic corpus might lead to more volatility in the sentiment indices, making it hard to draw conclusions.

Table 1: Summary statistics economic corpus and health corpus

	Mean	St Dev	Min	Max	N
Economic corpus	64.13	15.75	44	90	962
Health corpus	470.1	232.5	210	1038	7052

Figure 1 plots the number of articles for each month. The figure is scaled due to the large size difference between the economic and health corpus. The left y-axis corresponds to the economic corpus and the right y-axis corresponds to the health corpus. For both corpora, most articles were published in the first three months following the COVID-19 crash, after which the number of articles steadily declined until the end of the time period. This trend likely reflects the initial surge of media coverage in response to the immediate impact of the COVID-19 crisis, followed by a gradual stabilisation as the pandemic evolved.

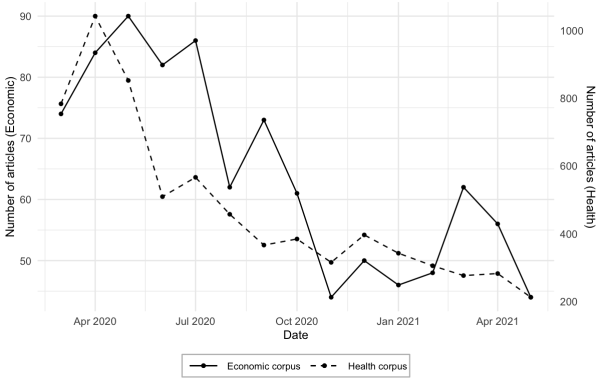

Figure 1: Number of articles for each corpus scaled

The sample size of the economic corpus reaches its maximum in May 2020 and its minimum in both November 2020 and May 2021. The size fluctuates strongly over the year, with peaks in September 2020 and March 2021. Following these peaks are strong drops, dropping from 86 to 62 from July to Augustus, and from 73 to 44 from September to November. Although the corpus is small, it can provide specific insights into economic trends and highlight important events during the first year of the COVID-19 crisis. The health corpus reaches its maximum size in April 2020 and its minimum size in May 2021. Following the announcement of a global pandemic in March 2020, the number of articles published on health-related topics peaked in April 2020, but fell sharply to 509 articles in June 2020, followed by a steady decline in subsequent months until May 2021. This steady decline indicates a gradual stabilisation of health-related news during the pandemic. Both corpora experienced a general decline in the number of articles over the year, indicating a reduction in media coverage as the initial urgency of the pandemic among the public lessened over time.

5.2 HRS data

This study considers responses from the 15$^{\text{th}}$ wave of the 2020 HRS Core, integrating longitudinal data on respondent characteristics that have been tracked consistently across survey years. After merging the RAND HRS Longitudinal File 2020 and the 2020 RAND HRS Fat File, the sample consists of $15,723$ individuals from $11,052$ households. Individuals included in the analysis should be representative of the sample, thus, this study excludes responses that may not provide meaningful insights. Making an educated guess on the future of the stock market is not easy for most individuals, especially if they do not have a stock portfolio or do not follow stock market developments. Therefore, it is prudent to exclude their responses to the probability questions as they might not have well-informed expectations regarding the stock market. Out of the $15,723$ individuals, the answers of $2,768$ ($\sim 18\%$) individuals who did not answer P_0 are removed, which leaves a sample of $12,955$ individuals. Out of these individuals, $4,390$ answered '50' to the P_0 question; these individuals were subsequently asked the follow-up question to understand this answer. $3,403$ ($\sim 78\%$) people answered that they were unsure about their answers to P_0 of '50'. This research also excludes individuals who did not answer P_1 and P_2, which leaves a sample of $7,426$ individuals. Finally, excluding individuals of ineligible ages and accounting for missing values in the controls, the final sample consists of $7,155$ individuals ($\sim 47\%$ of the full sample) from 6,073 households. Table 2 provides an overview of the variables included in the final sample.

Table 2: Summary statistics HRS data

	Mean	St Dev	Min	Max
Female	0.553	0.497	0	1
Single	0.355	0.478	0	1
Black	0.200	0.400	0	1
Hispanic	0.141	0.348	0	1
Age	67.55	9.63	51	100
Years of education	13.57	2.95	0	17
Above-average cognition	0.595	0.491	0	1
Follows the stock market	2.34	0.634	1	3
Informed stock market	0.569	0.495	0	1
Stockholder	0.235	0.424	0	1
P_0	48.59	25.63	1	99
P_1	35.93	23.76	0	99
P_2	33.53	22.41	0	100
N	7155			

Note: P_0, P_1, and P_2 represent answers to probabilistic expectation questions regarding the stock market.

Similar to the HRS data from the 9th survey wave, only 8.8% of the respondents in the 15th survey wave answered that they follow the stock market 'very closely' (Hudomiet et al., 2011). Additionally, the *informed* variable exhibits a high correlation with *stockmarket*, which can be seen in Appendix 5. Therefore, this study chooses to exclude *stockmarket* in favour of *informed* as a control in the regression analyses.

The summary statistics provide a comprehensive overview of demographic, cognitive, and financial characteristics of the sample. The sample comprises 7,155 individuals, of which 55.3% are female. A substantial portion, 35.5%, of the sample is single, and the racial composition includes 20% Black and 14.1% Hispanic participants. The mean age is close to 68 years, spanning from 51 to 100 years, indicating a wide age range within the dataset. Individuals have almost 14 years of education, which suggests that on average participants went to college, reflecting a relatively well-educated population. Additionally, approximately 60% of the sample is classified as having above-average cognition. While 56.9% of the respondents are considered informed about the stock market, only 23.5% of the sample owns stocks. The mean values of P_0, P_1, and P_2, 48.59, 35.93, and 33.53 respectively, suggest that respondents do not believe that the stock market will have strong positive or negative movement in the next year, and the standard deviations indicate a broad distribution of probabilities.

6 Results

This study analyses the effect of news sentiment during the COVID-19 crisis on U.S. households' expectations regarding the one-year-ahead returns on the Dow Jones Industrial Average. News sentiment is captured by constructing and validating monthly sentiment indices from an economic corpus and a health-related corpus, leveraging the pre-trained BART large model for zero-shot classification. The model is implemented in Python 3.11 using a Hugging Face pipeline for the *bart-large-mnli* model. Subsequently, reduced-form regressions examine the relationship between households' expectations, using responses to probabilistic survey questions as proxies for the expectations, and a set of control variables, including demographic characteristics and socio-economic factors, and the constructed sentiment indices.

6.1 Sentiment analysis

This section discusses the outcomes of the novel approach to sentiment analysis using zero-shot classification. For each corpus, two sentiment indices are constructed: an indicator-based and a score-based index. The indicator-based index captures the proportion of positive articles in a given month, and the score-based index captures the intensity of the sentiment in a given month. The constructed indices are benchmarked against the sentiment scores from the *COV19Tweets* dataset (Lamsal, 2021). Figure 2 plots the sentiment indices over the time period.

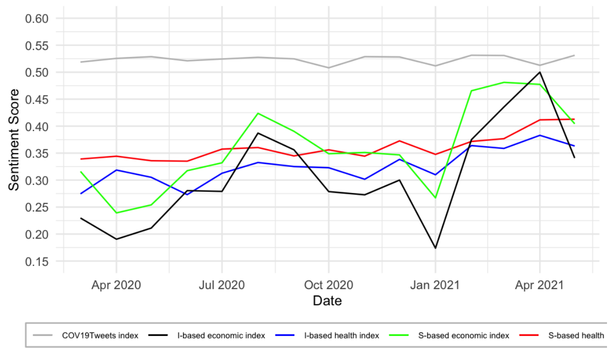

Figure 2: Sentiment Scores over Time

Note: I-based refers to the indicator-based index and S-based refers to the score-based index

The sentiment indices constructed in this research are noticeably lower than the benchmark index. The maximum score of the constructed indices, 0.500, is lower than the minimum score of the *COV19Tweets* index, 0.508, see Appendix 6. This discrepancy could suggest that the zero-shot classifier tends to assign more negative labels than the TextBlob module used by Lamsal (2021) assigns negative scores, or that news articles during this time contained more language associated with negative sentiment than the tweets. Given the lower sentiment scores for the constructed indices, it is important to consider that the coefficients in the regression analysis could be biased downwards. Additionally, score-based sentiment indices generally report higher sentiment scores than their indicator-based counterparts. The mean of the constructed indices ranges from 0.307 for the score-based economic index to 0.361 for both the indicator-based economic index and the indicator-based health index, see Appendix 6. This finding suggests that, although fewer articles are classified as positive, the zero-shot classifier is more confident in labelling those that are positive. The classifier's confidence in labelling positive articles could be due to its adeptness at recognising positive language or because positive news articles during the first year of the COVID-19 crisis contained more frequent and stronger affirmative language, making the class more apparent to the classifier.

The sentiment indices constructed from the economic corpus fluctuate significantly. As discussed in Section 5.1, the smaller size of the economic corpus relative to the health corpus could affect the smoothness of the sentiment scores. With fewer articles, outliers have a larger impact on sentiment scores in a particular month. However, these fluctuations could also reflect economic events during the time period. For instance, the initial drop in sentiment in early 2020 likely corresponds to the declaration of a global pandemic and nationwide lockdowns. Subsequent fluctuations could reflect positive developments such as economic stimulus measures and stock market recovery, as well as negative factors like rising unemployment rates and significant recession risks. Overall, the trend shows an increase from the start to the end of the time period, aligning with expectations that news sentiment would improve as the initial shock of the pandemic wore off and restrictions were gradually lifted.

In contrast to the economic sentiment indices, the health sentiment indices remained fairly consistent over the time period, with no significant spikes or drops, indicating reliable sentiment tracking. Similar to the economic indices, the overall trend in health sentiment was positive, showing that sentiment improved as the pandemic continued. Slight fluctuations could be attributed to news articles reflecting both positive developments such as vaccine announcements and negative developments such as rising case numbers and the emergence of new virus variants.

The correlation heat map in Figure 3 provides a comprehensive visualisation of the relationship between different sentiment indices. Sentiment indices are selected for the regression analyses based on their correlation with the benchmark index, as a higher correlation indicates that the constructed index more accurately reflects the sentiment patterns during the pandemic.

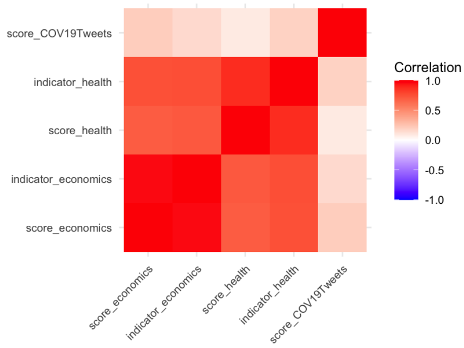

Figure 3: Correlation Heat Map of Sentiment Indices

Figure 3 shows strong correlations among economic sentiment indices, suggesting a high degree of consistency between the two construction methods in capturing sentiment trends in economic news. For the health corpus, the correlation between the indicator-based and score-based sentiment indices is moderate, indicating some consistency but also notable variation in how sentiment is captured.

The correlation between the sentiment indices from different corpora is mild. This observation supports the notion that the corpora capture different sentiment trends and highlights the importance of considering news sentiment from distinct domains. Additionally, the score-based economic index and the indicator-based health index exhibit the strongest correlation with the *COV19Tweets* index, with correlation values of approximately 0.211 and 0.191, respectively. The correlation matrix can be found in Appendix 7. This indicates that these indices are most adept at capturing the sentiment trends reflected in the benchmark index, suggesting that they adequately reflect the sentiment in their respective corpora and are suitable for further analysis.

These observations suggest that different domains may require distinct construction techniques for sentiment indices to effectively capture sentiment patterns in news articles. The score-based sentiment index is more reliable for the economic corpus, potentially due to the use of strong, affirmative language in economic articles. This language enables the zero-shot classifier to be more confident about its classification, which allows the score-based index to capture sentiment patterns more reliably, and potentially, more accurately. In contrast, health-related articles may employ less definitive language, resulting in a lower certainty in sentiment classification by the zero-shot classifier. Consequently, the indicator-based sentiment index, which focuses on the proportion of positive articles rather than the intensity of sentiment, demonstrates superior performance for the health corpus. Thus, these findings suggest that there is no dominant approach to the construction of sentiment indices following zero-shot classification, but rather that the approach should be tailored to a domain specifically.

6.2 Regression analysis

This section discusses the results of the reduced-form OLS regressions, examining the effect of news sentiment, captured by the score-based economic and indicator-based health indices, on the level, uncertainty, and disagreement of households' stock market expectations. Table 3 presents the results of the regressions where the score-based economic sentiment index is included as an independent variable. Columns 1, 2, and 3 report the regressions for the population average of the level of expectations, and columns 4, 5, and 6 display the regressions for the proxies indicating disagreement among respondents. Finally, columns 7 and 8 provide the results for the regressions related to the population average of uncertainty in expectations.

Table 3: OLS regressions for the level (columns 1-3), disagreement (columns 4-6), and uncertainty (columns 7 and 8) of expectations including the score-based economic index

Dependent variable	1	2	3	4	5	6	7	8
	P_0	P_1	P_2	$\lvert u_{P_0}\rvert$	$\lvert u_{P_1}\rvert$	$\lvert u_{P_2}\rvert$	P_{k+}	P_{k-}
Constant	51.10	45.38	49.73	19.42	23.20	23.68	-5.72	0.836
	(3.38)**	(3.27)**	(3.06)**	(1.72)**	(1.69)**	(1.67)**	(2.94)	(4.68)
S-based economic index	-0.820	-17.05	-3.67	-2.77	-9.05	-2.25	-16.23	-4.49
	(5.01)	(4.84)**	(4.53)	(2.55)	(2.50)**	(2.47)	(4.36)**	6.92
Female	-4.41	2.04	2.98	0.113	0.407	0.919	6.45	-1.43
	(0.608)**	(0.588)**	(0.550)**	(0.309)	(0.304)	(0.300)**	(0.529)**	(0.841)
Single	-2.15	-1.56	0.018	0.091	-0.275	0.898	0.588	-2.13
	(0.643)**	(0.622)*	(0.581)	(0.327)	(0.321)	(0.317)**	(0.559)	(0.889)*
Black	-6.09	2.07	0.294	0.957	0.350	1.13	8.16	-5.79
	(0.779)**	(0.753)**	(0.704)	(0.396)*	(0.389)	(0.384)**	(0.678)**	(1.08)**
Hispanic	-4.15	2.30	0.573	1.53	0.492	0.888	6.44	-3.58
	(0.927)**	(0.896)**	(0.838)	(0.471)**	(0.463)	(0.547)	(0.806)**	(1.28)**
Age	-0.187	-0.105	-0.297	0.092	0.017	-0.026	0.082	-0.484
	(0.032)**	(0.031)**	(0.029)**	(0.016)**	(0.016)	(0.016)	(0.028)**	(0.045)**
Years of education	0.688	0.056	0.147	-0.206	-0.118	-0.265	-0.632	0.835
	(0.112)**	(0.109)	(0.102)	(0.057)**	(0.056)*	(0.055)**	(0.098)**	(0.155)**
Above-average cognition	2.16	-0.556	0.920	-1.42	-0.357	-0.261	-2.72	3.08
	(0.644)**	(0.623)	(0.583)	(0.328)**	(0.322)	(0.318)	(0.561)**	(0.891)**
Informed stock market	6.60	2.90	2.22	-0.368	0.956	0.264	-3.70	8.82
	(0.627)**	(0.607)**	(0.567)**	(0.319)	(0.372)**	(0.309)	(0.546)**	(0.867)**
Stockholder	4.29	0.001	-2.21	-1.82	-0.322	-1.30	-4.29	2.09
	(0.744)**	(0.720)	(0.673)**	(0.378)**	(0.372)	(0.367)**	(0.647)**	(1.03)*
R^2	0.0904	0.0105	0.0266	0.0255	0.0051	0.0171	0.1152	0.0655
Adjusted R^2	0.0891	0.0091	0.0252	0.0241	0.0038	0.0157	0.1140	0.0642

Note: Standard errors in parentheses; * significant at 5%; ** significant at 1%

The score-based economic index has a significantly negative relationship with P_1 (-17.05), indicating that when economic news sentiment in a given month increases, the level of expectations for large stock market gains tends to decrease. This suggests that positive economic news might negatively affect stock market expectations. While this result seems counter-intuitive, it could indicate that positive economic news might be perceived as signalling market stabilisation rather than triggering substantial gains. Households might believe that the market has already absorbed the positive news, reducing the probability of large gains. Additionally, increased risk aversion during the crisis could persist despite positive news, leading to more cautious expectations regarding large gains. Compared to other coefficients for this regression, it seems that the economic index has a strong relationship with expectations on substantial gains. Positive economic news also appears to reduce expectations of losses, although this relationship is not significant, suggesting that positive news might mitigate fears of substantial losses.

Furthermore, the economic sentiment index is negatively related to $|u_{P_1}|$ with a coefficient of -9.05, suggesting that when economic news in a month is perceived as more positive, disagreement among households tends to decrease, leading to more uniform responses. This coefficient also has the largest coefficient, suggesting that the relationship between economic sentiment and disagreement about large gains is the strongest. While the coefficients for $|u_{P_0}|$ and $|u_{P_2}|$ are not significant, their negative signs indicate that positive economic sentiment may generally reduce uncertainty across all expectations. The relationship between economic sentiment and uncertainty regarding large gains is also negative (-16.23), indicating that more positive economic news seems to decrease respondents' uncertainty about their expectations for large gains. Similarly, while the coefficient for uncertainty about losses is not significant, the negative relationship suggests that positive news likely reduces uncertainty regarding potential losses. Compared to other coefficients in this regression, the economic index appears to have a particularly strong relationship with expectations regarding substantial gains.

Examining the control variables, women seem to expect a lower chance of stock market improvement in the next year, yet they have higher expectations for both large gains and large losses. On average, women exhibit more disagreement, which is significant for large losses. Single respondents tend to have lower expectations for stock market performance and large gains. They also demonstrate more disagreement but less uncertainty about their expectations of large losses. The results for Black and Hispanic respondents are similar, showing lower expectations for future stock returns, but higher expectations for both large gains and large losses. These groups also tend to exhibit more disagreement regarding expectations and are more uncertain about large gains than large losses. Older respondents generally have lower expectations, while more educated respondents tend to be more optimistic about future stock returns. Additionally, more educated respondents on average exhibit less disagreement and uncertainty, particularly regarding large gains. Respondents with above-average cognition show similar patterns; they tend to have more positive expectations, exhibit less disagreement for all expectations, and have greater uncertainty about large losses. Being informed about the stock market shows a strong positive relationship with the level of expectations, and also for large gains and large losses. Moreover, they tend to be less uncertain about large gains and more uncertain about large losses. Finally, stockholders tend to be more optimistic about the future of the stock market, with lower expectations of large losses on average. However, they tend to exhibit more disagreement with other stockholders and more uncertainty about large losses than about large gains.

Finally, the low adjusted R^2 values across all regressions indicate that the models explain only a small fraction of the variation in stock market expectations, which is common in studies with a large set of heterogeneous respondents. The model for $|u_{P_1}|$ only explains approximately 0.38% of the variation in the model, while the best model for P_{k+} explains approximately 11.4% of the variation in the model. This limitation suggests the need for further refinement of the model, possibly by including additional controls or exploring non-linear relationships, to better capture the dynamics of household expectations during the COVID-19 crisis.

Table 4 shows the results of the regressions where only the indicator-based health sentiment index is included as an independent variable. Each column denotes the same variables as in Table 3.

Table 4: OLS regressions for the level (columns 1-3), disagreement (columns 4-6), and uncertainty (columns 7 and 8) of expectations including the indicator-based health index

Dependent variable	1	2	3	4	5	6	7	8
	P_0	P_1	P_2	$\lvert u_{P_0}\rvert$	$\lvert u_{P_1}\rvert$	$\lvert u_{P_2}\rvert$	P_{k+}	P_{k-}
Constant	49.47	42.43	46.30	17.23	24.14	21.68	-7.04	-4.23
	(4.75)**	(4.59)**	(4.29)**	(2.41)**	(2.37)**	(2.34)**	(4.13)	(6.56)
I-based health index	3.97	-10.50	6.15	3.51	-13.13	3.46	-14.47	10.12
	(11.51)	(11.14)	(10.41)	(5.85)	(5.75)*	(5.68)	(10.02)	15.90
Female	-4.41	2.04	2.98	0.110	0.350	0.920	6.45	-1.43
	(0.608)**	(0.589)**	(0.550)**	(0.309)	(0.304)	(0.300)**	(0.530)**	(0.841)
Single	-2.16	-1.69	-0.025	0.053	-0.333	0.869	0.474	-2.18
	(0.642)**	(0.621)**	(0.580)	(0.326)	(0.321)	(0.317)**	(0.559)	(0.887)*
Black	-6.09	2.11	0.292	0.956	0.338	1.13	8.20	-5.80
	(0.779)**	(0.754)**	(0.705)	(0.396)*	(0.389)	(0.384)**	(0.678)**	(1.08)**
Hispanic	-4.16	2.31	0.556	1.50	0.431	0.883	6.47	-3.60
	(0.927)**	(0.897)**	(0.838)	(0.471)**	(0.463)	(0.547)	(0.807)**	(1.28)**
Age	-0.186	-0.094	-0.291	0.096	0.021	-0.022	0.091	-0.475
	(0.032)**	(0.031)**	(0.029)**	(0.017)**	(0.016)	(0.016)	(0.028)**	(0.045)**
Years of education	0.686	0.033	0.140	-0.212	-0.129	-0.270	-0.652	0.826
	(0.112)**	(0.109)	(0.101)	(0.057)**	(0.056)*	(0.055)**	(0.098)**	(0.155)**
Above-average cognition	2.17	-0.501	0.934	-1.41	-0.298	-0.271	-2.67	3.10
	(0.644)**	(0.623)	(0.583)	(0.328)**	(0.322)	(0.318)	(0.561)**	(0.890)**
Informed stock market	6.60	2.92	2.22	-0.368	0.995	0.271	-3.68	8.82
	(0.627)**	(0.607)**	(0.567)**	(0.319)	(0.314)**	(0.310)	(0.546)**	(0.867)**
Stockholder	4.29	-0.008	-2.21	-1.82	-0.306	-1.31	-4.30	2.08
	(0.744)**	(0.720)	(0.673)**	(0.378)**	(0.372)	(0.367)**	(0.648)**	(1.03)*
R^2	0.0904	0.0089	0.0266	0.0253	0.0039	0.0171	0.1137	0.0655
Adjusted R^2	0.0891	0.0075	0.0252	0.0240	0.0025	0.0157	0.1125	0.0642

Note: Standard errors in parentheses; * significant at 5%; ** significant at 1%

The results show that the indicator-based health index is only significant in the regression of disagreement about large stock market gains. Specifically, more positive health-related news is associated with significantly less disagreement about large gains, suggesting that when there is more favourable health news, households tend to have more aligned expectations regarding the stock market. This relationship is highlighted by a relatively large coefficient of -13.13, indicating a stronger relationship with disagreement than other coefficients. Although the other coefficients of the indicator-based health index are insignificant, considering their signs provides insights into the potential relationship between sentiment and the variables. The coefficients for the level of overall expectations and for large losses are positive, suggesting that more positive news correlates with higher overall market expectations, but also with higher expectations of large losses. Conversely, the negative coefficient for the level of expectations of large gains suggests that positive health news tends to lower respondents' expectations for large gains. Furthermore, the outputs of the disagreement regressions indicate that more positive health news tends to increase disagreement on average about overall expectations and large losses. Finally, health sentiment seems to be negatively related to uncertainty about gains and positively related to uncertainty about losses, which suggests that positive health news decreases uncertainty about potential gains while increasing uncertainty about potential losses.

Since the changes in the control variables between the two tables are minimal, and the sign and significance rarely change, I do not delve further into their interpretation for Table 4. The minimal changes in the coefficients of the control variables between the two tables indicate that including the sentiment indices does not substantially impact the relationship of these variables with the dependent variables. This suggests that the sentiment indices do not have a strong explanatory power. The overall lower model fit for both tables, indicated by the low adjusted R^2 values, further supports this, suggesting that the models explain only a small fraction of the variation in stock market expectations.

The results presented in Table 3 and Table 4 highlight several findings regarding the impact of news sentiment on households' stock market expectations. Notably, the indicator-based health index is only significant in one regression, while the score-based economic index is significant in three. Despite that the economic sentiment index is derived from a significantly smaller corpus and therefore may be less representative of economic news, the results suggest that the economic index shows a more consistent and significant impact on households' expectations. This finding could indicate that economic news might be more closely monitored by households when forming beliefs and expectations about the stock market. Furthermore, since the health sentiment index shows significance in fewer models, it could indicate that while there was much health news during the first year of the COVID-19 crisis, its translation into stock market expectations might be more nuanced or less straightforward.

Finally, it is important to acknowledge that this study does not have information on whether the respondents actually read the news articles used to construct the sentiment indices. Consequently, this research cannot be certain that the sentiment captured in these articles directly influences their expectations. Thus, this study is cautious in interpreting the signs and significance of the sentiment indices, since the significant relationships might not genuinely reflect the impact of news sentiment on expectations but rather result from unrelated factors.

7 Conclusion

This paper explores the influence of news sentiment during the first year of the COVID-19 pandemic on U.S. households' expectations regarding future stock market returns. Sentiment analysis was conducted using zero-shot classification with a pre-trained BART model to construct monthly sentiment indices from the two corpora: economic news and health-related news. Subsequently, reduced-form OLS regressions were estimated, regressing proxies for the variables of interest–subjective mean, subjective uncertainty, and cross-sectional heterogeneity–on the sentiment indices, alongside other control variables, to assess their impact on households' stock market expectations.

Zero-shot classification demonstrated higher confidence in scoring documents it perceived as positive but generally classified more documents as negative compared to the benchmark index, *COV19Tweets*. The first result suggests that zero-shot classification could be sensitive to the degree of affirmative language used in the articles and that positive articles might generally use more expressive language than negative articles, making the classifier more confident in labelling positive articles. This suggests that the sentiment indices should be validated against different benchmarks to ensure robustness. The second result indicates that the zero-shot classifier tends to classify more articles as negative than the lexicon-based approach used by Lamsal (2021). From this research it is not clear whether zero-shot classifiers classify more articles negatively or lexicon-based methods score fewer articles negatively, thus, further research should explore the application of zero-shot classification for sentiment analysis relative to established methods such as lexicon-based approaches and supervised learning classifiers. Additionally, it may be interesting to further explore the two-stage methodology proposed by Tesfagergish et al. (2022) to leverage zero-shot classification to label datasets for training supervised classifiers.

The approach to constructing sentiment indices seems to be domain-specific and dependent on the degree of expressive and affirmative language used. Economic news sentiment was better captured by the score-based sentiment index, which reflects both the polarity and intensity of sentiment associated with the text. Economic articles may use stronger language to express the content, which the zero-shot classifier seems to be more adept at confidently capturing. Contrarily, health news sentiment was more accurately captured by the indicator-based sentiment index, which measures the proportion of articles labelled as positive. This discrepancy may suggest that health news articles use less expressive language, which makes it difficult for classifiers to confidently label articles, thus, the proportion of positive articles better captures sentiment patterns in this domain. In general, these findings suggest that if language is less affirmative, the indicator-based approach may be more reliable, else the score-based approach may yield more reliable results. However, the lack of a dominant strategy in constructing sentiment indices may also be influenced by the selection bias of newspapers included in this research. This study focuses on major U.S. newspapers due to their wide circulation and broad readership in favour of regional or niche newspapers. Major newspapers often employ more general language to cater to a broader audience, which might not capture the specific nuances present in more specialised newspapers. To address this, future research should consider including a wider range of news sources and focus on additional domains to determine whether the observed domain-specific performance holds when incorporating different types of publications and domains. This allows researchers to better understand the extent to which the domain specificity of sentiment indices is influenced by the inherent linguistic characteristics of the articles or by the selection of news sources.

The corpora used in this study were limited to articles from five major newspapers, which may have implications for the validity and reliability of the sentiment analysis. Notably, the economic corpus was relatively small, potentially leading to the construction of less reliable sentiment indices. Although the sentiment indices derived from the economic corpus exhibited more fluctuations than those from the larger health corpus, their relatively strong correlation with the benchmark index suggests that they remain representative of economic sentiment. However, it would be prudent to further explore the effect of economic sentiment indices constructed from a larger corpus. To expand the corpus for economic and health-related news, it could be beneficial to allow the article length to exceed 512 words. While the BERT encoder would still truncate longer articles to 512 tokens, a more extensive range of articles could provide deeper insights. Although shorter articles tend to be more concise and focused, longer articles might contain additional insights. Often, the most important information of an article is conveyed in the initial paragraphs to catch the readers' attention, thus, despite the truncation, there should be minimal loss of content. Furthermore, constructing sentiment indices on a monthly basis smooths out variations in sentiment within a month and may overlook nuances that could be captured by daily or weekly sentiment scores. Since the indices were constructed monthly due to the frequency of the HRS survey responses, exploring alternative datasets that record survey responses at more granular intervals could be beneficial. This approach could potentially better capture the impact of sentiment on household expectations by reflecting more nuanced changes in news sentiment.

The regression analyses provide insights into how news sentiment may influence household expectations. Despite the limited explanatory power of the models due to the relatively poor fit, The relationships between control variables, particularly the sentiment indices and the proxies for variables of interest, yield interesting results. The results demonstrated a significant negative relationship between the score-based economic sentiment index and expectations for large stock market gains (P_1). This suggests that positive economic news might lead households on average to anticipate substantial gains in stock market returns less often, potentially due to persistent risk aversion as a consequence of the pandemic. This study finds a similar negative relationship for the health sentiment index, and, while not statistically significant, also shows that positive health news might lead to lower expectations of substantial market gains. These findings could be of interest to policymakers, as they indicate that more positive news could cause households to become more cautious about substantial gains. This caution might stem from a loss of trust in financial institutions or in the government, seemingly leading to scepticism about positive news. Additionally, economic sentiment is negatively related with related to disagreement and uncertainty about large gains. This suggests that, while households are cautious about large positive gains on the stock market, they seem to be more certain about their expectations, and that there is less disagreement among households. The health index also exhibits a significant negative relationship with disagreement and a similar, but not significant, relationship with uncertainty. These findings highlight that, despite differences in the content of the corpora and the construction techniques for the sentiment indices, both indices show consistent relationships with the proxy variables related to substantial stock market gains. However, the sentiment indices exhibit different behaviour for the regressions concerning overall expectations and large losses. The economic sentiment index shows negative relationships with both proxy variables, while the health sentiment index shows positive relationships. These results indicate that, for a given sentiment index, the direction of the relationship with the level, uncertainty, and disagreement of expectations remains consistent. However, this consistency in the direction of the relationship does not necessarily hold between different sentiment indices. These findings underscore the importance of considering both economic and health-related news in understanding and predicting households' stock market expectations during crises. The influence of news sentiment on the formation of stock market expectations highlights the power of information in shaping households' beliefs.

8 Acknowledgements

The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

References

Adoma, A. F., Henry, N.-M. & Chen, W. (2020). Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. In *2020 17th international computer conference on wavelet active media technology and information processing (iccwamtip)* (pp. 117–121).

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F. & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, *77*, 236–246.

Asif, M., Zhiyong, D., Iram, A. & Nisar, M. (2021). Linguistic analysis of neologism related to coronavirus (covid-19). *Social Sciences & Humanities Open*, *4*(1), 100201.

Binder, C. (2020). Coronavirus fears and macroeconomic expectations. *Review of Economics and Statistics*, *102*(4), 721–730.

Bowman, S. R., Angeli, G., Potts, C. & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326*.

Broadstock, D. C. & Zhang, D. (2019). Social-media and intraday stock returns: The pricing power of sentiment. *Finance Research Letters*, *30*, 116-123.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, *12*, 2493–2537.

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

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Giglio, S., Maggiori, M., Stroebel, J. & Utkus, S. (2021). The joint dynamics of investor beliefs and trading during the covid-19 crash. *Proceedings of the National Academy of Sciences*, *118*(4).

Gormsen, N. J. & Koijen, R. S. J. (2020). Coronavirus: Impact on stock prices and growth expectations. *The Review of Asset Pricing Studies*, *10*(4), 574-597.

Hanspal, T., Weber, A. & Wohlfart, J. (2021). Exposure to the covid-19 stock market crash and its effect on household expectations. *Review of Economics and Statistics*, *103*(5), 994–1010.

Hudomiet, P., Kézdi, G. & Willis, R. J. (2011). Stock market crash and expectations of american households. *Journal of Applied Econometrics*, *26*(3), 393–415.

Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review*, *52*(1), 273–292.

Lamsal, R. (2021). Design and analysis of a large-scale covid-19 tweets dataset. *Applied Intelligence*, *51*(5), 2790–2804.

Leist, A. K. (2013). Social media use of older adults: A mini-review. *Gerontology*, *59*(4), 378–384.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Malecki, K. M. C., Keating, J. A. & Safdar, N. (2021). Crisis communication and public perception of covid-19 risk in the era of social media. *Clinical Infectious Diseases*, *72*(4), 697–702.

Mehta, P., Pandya, S. & Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. *PeerJ Computer Science*, *7*, e476.

Monselise, M., Chang, C.-H., Ferreira, G., Yang, R. & Yang, C. C. (2021). Topics and sentiments of public concerns regarding covid-19 caccines: Social media trend analysis. *Journal of Medical Internet Research*, *23*(10), e30765.

Mutz, D. C. & Soss, J. (1997). Reading public opinion: The influence of news coverage on perceptions of public sentiment. *The Public Opinion Quarterly*, *61*(3), 431–451.

Pushp, P. K. & Srivastava, M. M. (2017). Train once, test anywhere: Zero-shot learning for text classification. *arXiv preprint arXiv:1712.05972*.

Rick, U. K., Boykoff, M. T. & Pielke, R. (2011). Effective media reporting of sea level rise projections: 1989–2009. *Environmental Research Letters*, *6*(1), 014004.

Ruder, S., Peters, M. E., Swayamdipta, S. & Wolf, T. (2019). Transfer learning in natural language processing. In *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: Tutorials* (pp. 15–18).

Salisu, A. A. & Vo, X. V. (2020). Predicting stock returns in the presence of covid-19 pandemic: The role of health news. *International Review of Financial Analysis*, *71*, 101546.

Shofiya, C. & Abidi, S. (2021). Sentiment analysis on covid-19-related social distancing in canada using twitter data. *International Journal of Environmental Research and Public Health*, *18*(11), 5993.

Sousa, M. G., Sakiyama, K., Rodrigues, L. d. S., Moraes, P. H., Fernandes, E. R. & Matsubara, E. T. (2019). Bert for stock market sentiment analysis. In *2019 ieee 31st international conference on tools with artificial intelligence (ictai)* (p. 1597-1601).

Spatt, C. S. (2020). A tale of two crises: The 2008 mortgage meltdown and the 2020 covid-19 crisis. *The Review of Asset Pricing Studies*, *10*(4), 759–790.

Sánchez-Rada, J. F. & Iglesias, C. A. (2019). Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. *Information Fusion*, *52*, 344-356.

Tesfagergish, S. G., Kapočiūtė-Dzikienė, J. & Damaševičius, R. (2022). Zero-shot emotion detection for semi-supervised sentiment analysis using sentence transformers and ensemble learning. *Applied Sciences*, *12*(17), 8662.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Wankhade, M., Rao, A. C. S. & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780.

Williams, A., Nangia, N. & Bowman, S. R. (2017). A broad-coverage challenge corpus for sentence understanding through inference. *arXiv preprint arXiv:1704.05426*.

Yin, W., Hay, J. & Roth, D. (2019). Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. *arXiv preprint arXiv:1909.00161*.

Zhang, D., Hu, M. & Ji, Q. (2020). Financial markets under the global pandemic of covid-19. *Finance Research Letters*, *36*, 101528.

Appendix

8.1 Code description

News articles from Factiva are organized into 'Economic corpus' and 'Health corpus' folders. For each corpus, the RTF files are combined to one RTF file using the Python 3.11 script 'append_rtf_files.py', which reads, extracts, and appends content. The 'zero_shot_classification.py' script then uses the *bart-large-mnli* model from Hugging Face for zero-shot classification, processing and classifying the articles as 'positive' or 'negative'. The output is saved in an Excel file.

The R script 'construction_sentiment_indices.R' processes this output, to construct score-based and indicator-based monthly sentiment indices, which are then written to a new Excel file. Sentiment scores from *COV19Tweets* dataset are processed using 'construction_benchmark.R'. The script transforms scores in each file to a 0-1 range, calculates the average of the transformed scores, which is considered the daily sentiment score for each file.

Data from the RAND HRS Longitudinal File 2020 and the 2020 RAND HRS Fat File is processed and cleaned in Stata using 'preprocessing.do'. It merges the datasets, constructs dummy variables, and maps sentiment indices to interview dates. The 'regression_analysis.do' file then performs regression analyses, including level, disagreement, and uncertainty regressions, using the economic and health sentiment indices.

8.2 HRS data

Table 5: Correlation Matrix of cleaned HRS variables

	Female	Single	Black	Hispanic	Age	Education
Female	1					
Single	0.230	1				
Black	0.056	0.170	1			
Hispanic	-0.002	-0.053	-0.171	1		
Age	-0.013	0.141	-0.112	-0.152	1	
Education	-0.024	-0.063	-0.049	-0.346	-0.008	1
Cognition	0.072	-0.115	-0.141	-0.071	-0.192	0.287
Informed	-0.184	-0.115	-0.101	-0.073	0.0163	0.227
Stockholder	-0.036	-0.121	-0.185	-0.167	0.126	0.254
P_0	-0.123	-0.114	-0.125	-0.079	-0.053	0.167
P_1	0.026	-0.032	0.025	0.030	-0.048	0.003
P_2	0.062	-0.008	0.021	0.023	-0.138	0.021

	Cognition	Informed	Stockholder	P_0	P_1	P_2
Cognition	1					
Informed	0.154	1				
Stockholder	0.137	0.255	1			
P_0	0.123	0.203	0.155	1		
P_1	0.007	0.049	0.001	0.583	1	
P_2	0.055	0.031	-0.042	0.054	0.167	1

8.3 Sentiment indices

Table 6: Summary statistics sentiment indices

	Mean	St Dev	Min	Max
Score-based economic index	0.361	0.076	0.239	0.481
Indicator-based economic index	0.307	0.089	0.174	0.500
Score-based health index	0.361	0.024	0.335	0.413
Indicator-based health index	0.326	0.031	0.273	0.383
COV19Tweets index	0.524	0.007	0.508	0.531

Table 7: Correlation matrix sentiment indices

	SE index	IE index	SH index	IH index	*COV19Tweets* index
S-based economic index	1	0.953	0.696	0.749	0.211
I-based economic index	0.953	1	0.712	0.756	0.156
S-based health index	0.696	0.712	1	0.878	0.089
I-based health index	0.749	0.756	0.878	1	0.191
COV19Tweets index	0.211	0.156	0.089	0.191	1

Note: S refers to score-based, I refers to indicator-based, E refers to economic, and H refers to health