Visual and textual content of social media and the stock market

Chi-Hsiou D. Hung[1,*] and *Ruipei Sun*[2]

Adam Smith Business School, University of Glasgow

Abstract

We study the effects of valence-arousal-dominance emotion dimensions on S&P 500 index returns using images, linguistic words, and emojis. By applying machine learning, we construct daily imagery emotion dimension indexes with PNG, JPG, and GIF images in StockTwits posts. Our approach of modeling behavioral phenomena demonstrates that imagery emotion indexes positively predict stock market returns. This predictive power holds in the presence of linguistic words and emojis. Realistic images elicit more intense responses than those including cartoons, thereby strongly impacting returns. The integration of realistic images heightens the emotional resonance within the associated textual content.

JEL classification: G02, G11, G12

Keywords: Behavioral modelling, Visual content, Social media, Machine learning, Emotion dimensions

* Corresponding author: Chi-Hsiou D. Hung, E-mail: chi-hsiou.hung@glasgow.ac.uk. We thank Constantinos Antoniou, Cathy Chen, Anthony Cookson, Jarrad Harford, Michael Weber, Yeqin Zeng and participants at the 4th Frontiers of Factor Investing Conference and International Risk Management Conference for helpful comments.

1. Introduction

Financial research shows that the sentiment expressed in news content interacts with stock returns. For example, Tetlock (2007) shows that media pessimism measured from a popular *Wall Street Journal* column predicts downward pressure on market prices, followed by a reversion, and that low market returns lead to high media pessimism. Garcia (2013) finds that the predictability of stock returns using news content in the *New York Times* is concentrated during NBER recessions. Obaid and Pukthuanthong (2022) show that sentiment conveyed in imagery content in news articles predicts stock returns. Baker and Wurgler (2006) use capital market characteristics to construct investor sentiment index and show that sentiment predicts the cross-section of stock returns. Another strand of research studies the impact of sentiment on stock returns by using exogenous shocks to proxy for sentiment. Hirshleifer and Shumway (2003) find that investor sentiment proxied by the amount of sunshine has positive correlations between high valence and stock market performance across 26 countries. Edmans, Garcia, and Norli, (2007) show that investor sentiment, transmitted from sports events, affects asset prices.

Crucially, however, the sentiment with optimism and pessimism characterizations is only one dimension of emotions—valence. Existing financial studies that focus on sentiment is the unidimensional valence-based cognitive modeling to investor emotions, and do not consider the role of emotion dimensions other than sentiment. Russell and Mehrabian (1977) and Sheth and Pham (2008) demonstrate that emotions encompass various dimensions and are not singular, unified processes. Psychology theory succinctly characterizes affective states with three independent emotion dimensions, as in the valence–arousal–dominance (VAD) model of Russell and Mehrabian (1977) and Mehrabian (1996). In the VAD model, valence represents the subjective evaluation of an individual's emotional state, ranging from pleasant to unpleasant. Arousal pertains to the intensity or activation level of an emotional experience and

represents the degree of psychological activation triggered by stimuli. Dominance reflects the perceived control or power dynamics within an emotional experience.

Despite all three emotion dimensions play important roles in influencing individuals' behavior and decision-making (Russell and Mehrabian, 1977; Lerner and Keltner, 2001; Lerner, Li, Valdesolo, and Kassam, 2015), the extant research of modelling behavioral phenomena has ignored the effects of the arousal and dominance dimensions of emotions on financial markets. Barberis (2018) emphasizes the importance of employing psychology-based models in asset pricing, highlighting the value of integrating psychological insights into financial research. The purpose of this paper is to fill this gap by studying the effects of the three fundamental emotion dimensions, expressed in the visual and textual content on the social media investing platform StockTwits, on the stock market return.

The content of StockTwits has been shown empirically to have impact on markets (e.g., Cookson and Niessner, 2020; and Renault, 2017). By applying machine learning (e.g., Garcia, Hu, and Rohrer, 2023), we construct imagery and textual VAD emotion-dimension indexes, using 2,046,190 images and 9,456,627 textual posts in our StockTwits sample covering 2,512 trading days from January 2012 to December 2021. Our wide coverage of image formats including PNG, JPG, and GIF images, and covering still photographs, and animated GIFs, allows us to capture a broad spectrum of visual content in memes. The valence index captures investor sentiment, whereas the arousal and dominance indexes, respectively, capture the dimensions of investor arousal emotion and investor dominance emotion.

The main contribution of this study is to show evidence that all three daily emotion dimensions expressed in the visual stimuli of images and textual content of linguistic words and visual text (that is, emojis) on StockTwits strongly predict U.S. stock market returns, indicating that all the three emotion dimensions influence investors' decision-making processes. To our knowledge, this study is the first to find evidence that the different facets of emotion

dimensions expressed on social media can predict U.S. stock market returns. Our evidence is consistent with behavioral approach that investor psychology impacts asset prices (Hirshleifer, 2001; and Barberis, 2018). This study contributes to the economics of social media literature (see Chen et al., 2014; Pedersen, 2022; Chen and Hwang, 2022; and Kogan, Moskowitz, and Niessner, 2022 and references therein) by offering important insights into how social media influences investor behavior and financial markets.[1]

Our use of visual content is in line with the cognitive load theory (Sweller, Van Merrienboer, and Paas, 1998) that imagery visual stimuli can speed up information perception and impact decision-making. Bazley, Cronqvist, and Mormann (2021) find that financial information presented in red color reduces individuals' risk preferences and willingness to invest in stocks. Nekrasov, Teoh, and Wu (2022) find that firms using visual stimuli in earnings announcements on Twitter experience higher levels of user engagement as reflected in increased retweets. Jiang, Kelly, and Xiu (2023) show that image-based signals using machine learning predict future stock returns. Gu, Teoh, and Wu (2023) construct firm-level sentiment using images in GIF format. For our analysis, we extract StockTwits posts containing $SPY, the symbol of the SPDR S&P 500 exchange-traded fund that tracks the Standard & Poor's Composite Index. The use of both imagery and textual content with emojis in online communications of StockTwits posts highlights a significant surge in their popularity and widespread adoption, with the annual count of images growing from 853 in 2012 to 539,450 in 2021.[2]

[1] Prior studies document stock return predictability of textual content on social media platform. For example, Cookson and Niessner (2020) find that the textual content on StockTwits exhibit stock return predictability. Renault (2017) shows that changes in investor sentiment expressed in the textual content on Stocktwits predict intraday return of the S&P 500 index ETF. Chen et al. (2014) show that textual content in social media Seeking Alpha can predict stock returns. Pedersen (2022) demonstrates that communications in social networks influence investors' trading.

[2] A growing number of StockTwits users choose the largest GIF search engine, Giphy.com, to insert images, mostly in GIF format, into their posts, streamlining the process of searching and uploading GIFs to express their investment opinions. Giphy.com is a popular website for discovering, sharing, and creating animated GIFs that offers a vast collection of GIFs in various categories and provides users with the means to express themselves

Consistent with the theory that individuals with high sentiments make optimistic judgments (Han, Lerner, and Keltner, 2007), our daily imagery valence index predicts positive changes in stock market returns. This finding indicates that on days when the average valence level of imagery content in StockTwits is high, signaling a greater degree of optimism, the stock market return increases the next day. A one standard deviation increase in the daily imagery valence index on day *t-1* predicts a return increase of 2.37 basis points, on average, in the S&P 500 index the next day. Our evidence is consistent with that of Obaid and Pukthuanthong (2022), who find that sentiment expressed in the visual content of news photos in still image formats (i.e., PNG and JPG) from the Wall Street Journal can predict daily market-level stock returns after controlling for textual content.

Arousal captures the emotional activation or engagement evoked in viewers by an image. High levels of arousal cause individuals to shift their attention and become more responsive to stimulus-based judgments, such as investment decisions (Han, Lerner, and Keltner, 2007; and Wilson and Brekke, 1994). Cookson et al. (2024) show that social media attention is distinct from sentiment, whereas Barber and Odean (2008) find that attention influences investors' stock-purchase considerations. Crucially, arousal affects attention and increases risk appetite because individuals with high-arousal emotional states tend to display risk-seeking behaviors (Mano, 1994; Porcelli and Delgado, 2009; and Galentino, Bonini, and Savadori, 2017). As expected, our imagery arousal index predicts positive changes in market returns, in which a one standard deviation increase in the daily imagery arousal index predicts, on average, an increase of 1.85 basis points in the S&P 500 index return the next day. This evidence shows that investors' elevated emotional arousal positively affects stock returns. Dominance reflects individuals' perceived level of influence or control, leading to a greater

through visual communication. Giphy.com also integrates with various messaging apps, social media platforms, and Internet communities, allowing users to share GIFs directly in conversations and posts easily.

determination to make decisions (Johnson, Leedom, and Muhtadie, 2012). Our imagery dominance index predicts upward increases in the market price, where a one standard deviation increase in the daily imagery dominance index predicts an increase of 3.37 basis points in the S&P 500 index return the next day. Moreover, this return increase is followed by further increases of 3.74 and 3.73 basis points in the subsequent two days.

We further explore the impact of imagery content in the presence of textual content in StockTwits posts on stock market returns. To this end, we construct daily textual emotion dimension indexes based on textual posts and use our LM-Renault-augmented word and emoji VAD lexicon to assign valence, arousal, and dominance scores to the textual content via machine learning methods. Importantly, the textual content in our study includes emojis, and linguistic words. We allocate valence, arousal, and dominance scores to the 1,114 most frequently used emojis in our StockTwits sample. Bai et al. (2019) demonstrate that emojis have become an integral part of digital communication by providing a fun and expressive way to enhance written messages and convey emotions. Arjaliès and Bansal (2018) show that emojis effectively communicate emotions, personal evaluations, and decisions. Messages can become affectively ambiguous without emojis, potentially leading to confusion between the communicators.

In our analysis, the correlations between the imagery and textual emotion dimension indexes are relatively weak, indicating that certain information captured in images is unique and not replicated in the textual content of StockTwits posts. We find that the effects of imagery emotion dimension indexes constructed based on all images remain strong, even when the influences of textual emotion dimension indexes are considered. We also find that all three textual emotion dimension indexes predict market-level return reversals in the trading week. Our evidence regarding the reversal of the impact of valence, expressed in textual content, on

stock market returns is consistent with Tetlock (2007), Garcia (2013), and Obaid and Pukthuanthong (2022).

We further construct imagery emotion dimension indexes specifically from realistic images. This is because realistic images impact persona perception more than cartoons (e.g., Zhao et al., 2019; and Tian and Zhang, 2021). Our evidence shows that, relative to the arousal-imagery emotion-dimension index constructed based on all images including cartoons, the realistic arousal-imagery emotion-dimension index has stronger effects on stock market returns. Furthermore, the effects of the emotion dimension indexes derived from realistic images are even more pronounced when textual emotion dimension indexes are included. We also find that both the realistic valence emotion-dimension index and realistic dominance emotion-dimension index can predict market-level return reversals, highlighting the importance of considering image characteristics and their influence on stock markets. Our evidence suggests that imagery content, particularly realistic images, has an important role in shaping investor emotions, in turn influencing stock returns and market dynamics.

The remainder of this paper is organized as follows. Section 2 delves into the theoretical framework of emotion dimensions, visual stimuli and their effects on the stock market. Section 3 delineates our StockTwits sample and the utilization of machine learning algorithms in creating models for image filters and image identification algorithms for assigning valence, arousal, and dominance scores. Section 4 describes the image and text preprocessing methods, the construction of emotion dimension indexes for images and textual data, and descriptive statistics of these indexes. Section 5 presents the econometric models and examines the effects of emotion dimensions, based on imagery and textual content, on stock market performance. The robustness tests, including using a different method to capture GIF VAD scores, controlling for Fama-French five factors, and out-of-sample tests, confirm our findings. Finally, Section 6 presents the conclusions.

2. Emotion dimensions, visual stimuli, and stock market returns

Investment decisions are stimulus-based judgments since investors assess and respond to stimuli, including risk perception, potential returns, and market trends (Lerner, Small, and Loewenstein, 2004). Our application of the VAD model in visual studies allows for a nuanced understanding of the emotional impact of visual stimuli and their influences on individuals' attitudes, behaviors, and decision-making processes (Lang, Bradley, and Cuthbert, 1997).

2.1. Valence, arousal, and dominance emotion dimensions

The Valence-Arousal-Dominance model developed by Russell and Mehrabian (1977) and Mehrabian (1996) is a well-recognized emotion dimensional model in psychology that describes affective states using three independent emotion dimensions: valence, arousal, and dominance.[3] Lang and Bradley (2007) illustrate the link between each emotion dimension and imagery visual content. Wadlinger and Isaacowitz (2006) find that high emotional valence in images enhances the processing and perception of positive stimuli, increasing the scope of attention. Coleman (2010) documents that the intensity of valence in images can profoundly affect the decision-making process. Individuals experiencing high (low) sentiments are prone to formulate optimistic (pessimistic) judgments (Han, Lerner, and Keltner, 2007; and Keltner and Lerner, 2010). Thus, we expect high valence levels positively influence stock returns.

Arousal, the second emotion dimension, is crucial for motivating specific behaviors when individuals encounter stimuli. Riemer and Viswanathan (2013) show that when individuals experience high-arousal emotions, they are more likely to be motivated to make accurate judgments and hence, will perform substantive processing and extensive information searches (Storbeck and Clore, 2008). Javela, Mercadillo, and Ramírez (2008) find that the arousal levels of pictures influence stimulus-based judgments where increases in arousal levels

[3] Russell (1980) develop the circumplex model that contains two fundamental dimensions of emotional experiences: valence and arousal. Bradley and Lang (1994) identify the third emotion dimension—dominance. Russell (1997) further validates the VAD model by adding the dominance dimension to the circumplex model.

enhance an individual's alertness and capacity to focus, leading to efficient processing of information and improved attention to tasks. Sheth and Pham (2008) find that images with higher arousal levels are more likely to capture and retain individuals' attention for extended periods. Cookson et al. (2024) differentiae social media sentiment and attention. Barber and Odean (2008) find that investors frequently consider purchasing stocks that initially catch their attention.

Importantly, arousal also has a profound impact on risk appetite. High arousal levels are associated with individuals' exhibiting risk-seeking behavior (Mano, 1994; Wilson and Brekke, 1994). Porcelli and Delgado (2009) and Galentino, Bonini, and Savadori (2017) find that participants in high-arousal groups are prone to make riskier choices, highlighting the impact of arousal on decision-making processes. Jahedi, Deck, and Ariely (2017) show that exposures to visual stimuli with high arousal levels increase risk-taking, hence affecting investment decisions. For example, Anderson and Brown (1984) document that, in the gambling environment of casinos, with flashing lights, loud sounds, and constant actions of games, gamblers are kept at a heightened state of arousal, prompted to bet large amounts of money or choose games with low odds of winning but high payouts.[4] Wegner and Giuliano (1980) demonstrate that high-arousal emotions tend to induce self-focused attention, resulting in individuals insisting on their opinions and being more willing to act. As the daily arousal index in our study indicates the arousal level experienced on a given day, we postulate that investors experiencing a higher level of arousal tend to increase their risk exposure to stocks and hence, positively impact stock market returns.

[4] Toma et al. (2023) describe the phenomenon during the dot-com bubble in the late 1990s and early 2000s, where the euphoria in the market led many to invest heavily in internet-related stocks without due consideration of the underlying risks. When the stock market experiences a significant upturn, the excitement and positive mood generated among investors can increase their arousal levels. This heightened state can lead investors to engage in riskier financial behaviors, such as making speculative investments or trading more frequently, driven by the thrill of potential high returns, despite the increased risk of substantial losses.

Russell (1980) defines that the dominance dimension of emotion as relating to the perception of freedom or limitations in one's behavior and representing a sense of control over a stimulus or situation. When individuals feel a sense of dominance, they perceive themselves as being unrestricted and free to act in various ways (Johnson, Leedom, and Muhtadie, 2012). High dominance corresponds to emotions associated with feelings of control, or assertiveness, whereas low dominance is associated with emotions characterized by vulnerability, submissiveness, or helplessness. Jerram et al. (2014) find that a high dominance level comprehensively boosts cognitive processes since dominance requires the individual to assess internal resources and states, as well as the external environment, and to make predictions about likely outcomes.

Mäntylä et al. (2016) and Dai, Han, Dai, and Xu (2015) demonstrate that individuals with high dominance emotions make quick decisions and share opinions with others to boost their creditability on social media. Shen, Hentschel, and Hideg (2023) document a real-life example that high-dominance leaders take bold and swift actions when facing the outbreak of the COVID-19 pandemic by imposing lockdowns and mobilizing national resources rapidly. We posit that investors with high dominance emotions are more inclined to make purchase decisions than sales decisions because of short-sale constraints that restrict sales decisions when investors do not possess stocks (e.g., Jones and Lamont, 2002). Thus, we expect dominance to positively affect stock market returns.

2.2. Visual stimuli

The cognitive load theory (Sweller, Van Merrienboer, and Paas, 1998) demonstrates that visual stimuli speed up information perception procedure and that the process for learning novel information is more effective when instructional materials or tasks are designed in a way that minimizes cognitive load, allowing individuals to allocate their cognitive resources more efficiently. The use of visual stimuli such as images, graphics, and videos can evoke emotional

responses, influence cognitive processes, and impact decision-making (Glimcher, 2003). Schnotz and Kürschner (2007) and Hemmig (2009) document the power of visual cues in capturing attention, eliciting emotional responses, and shaping cognitive processes, ultimately influencing individuals' attitudes, beliefs, and behaviors. Moreover, visual elements, such as photographs or videos, can contribute to the interpretation and understanding of news events.

In this study, we adopt two types of visual stimuli for analysis—images and emojis. Images are the most common targets for examining how the brain processes and interprets visual stimuli, such as color, form, depth, and motion. Psychological studies have explored the effects of images on emotional expressions. For example, Alpers and Gerdes (2007), Frischen, Eastwood, and Smilek (2008), and Kragel et al. (2019) show that images can evoke emotional experiences as visual stimuli impact cognitive processing even prior to conscious awareness of the stimulus, thus, influencing the decision-making process. Wong et al. (2012) show that visual stimuli applied to broadcast information can minimize cognitive load, thereby allowing individuals to allocate cognitive resources efficiently.

The second form of visual stimuli we examine is emojis. Emojis are graphic symbols standardized by Unicode, which are employed as a concise means of expressing notions and ideas to express and amplify individuals' opinions. Riordan (2017) demonstrates that both facial and non-facial emojis enable users to perform emotional labor that fosters and enhances their social connections. Fox, Kelley, and Paolucci (2023) score the valence dimension of emotions— sentiment—of emojis on Twitter posts, without classifying images.

3. Applying machine learning for image identification on StockTwits

3.1. The StockTwits sample

Using the StockTwits API, we obtain a comprehensive collection of 9,456,627 posts related to the S&P 500 index, specifically those containing the cash tag $SPY, from 1 January 2012 to

31 December 2021. Each post contains a timestamp, a unique user identifier, textual content, and, where applicable, imagery content associated with the post. Because StockTwits users can insert one image per post in either PNG, JPG, or GIF format, we verify the presence of image URLs ending with ".png" or ".jpg" for static images and ".gif" for the animated images. Our initial StockTwits dataset contains 6,980,000 images, including 6,225,000 PNG and JPG images and 755,000 GIF images. Each StockTwits post is accompanied by an accurate time stamp, enabling us to precisely define the time span of post messages for a trading day from 16:00 on day *t-1* to 16:00 on day *t*, rather than relying on a calendar day. This approach can encompass a comprehensive range of information and market-relevant events that occur outside regular trading hours.

Using data from StockTwits to extract investors' emotions offers numerous advantages over newspaper or journal content. First, with the platform's user base growing from approximately 170,000 in 2011 to over six million by December 2022, StockTwits has become the largest community for investors and traders. Second, StockTwits posts represent individuals' personal opinions, making them a straightforward proxy for capturing investor emotions (see, also, Renault, 2017; Cookson and Niessner, 2020; and Cookson et al., 2024). Moreover, StockTwits posts exhibit greater synchronization with the financial market than news content from traditional media sources since time lags are involved in publishing articles in the press (Renault, 2017). Hence, using data from StockTwits provides an opportunity to tap into the vast amount of information available on social media platforms, presenting an unprecedented opportunity for real-time and cost-effective aggregation and monitoring of emotional states, beliefs, and perspectives of a significant proportion of the investor population.

With the prevalence of social media, imagery memes represent an influential form of communication and expression that are used to share ideas, emotions, and commentaries on current events, popular culture, and internet phenomena (Sterelny, 2006; Highfield and Leaver,

2016). Images can spread rapidly across social media platforms, forums, and websites, reflecting and shaping zeitgeist internet culture. According to Heath, Bell, and Sternberg (2001), using imagery meme content has become part of social media's culture as a supplement or even a substitute to express individuals' opinions, and they have also emerged as a means of visually conveying emotions and amplifying emotional expressions (Coleman, 2010). Bikhchandani et al. (2021) and Nekrasov, Teoh, and Wu (2022) suggest the need for additional scholarly and empirical investigations of the connection between meme sentiment and market performance.

3.2. The machine learning method for assigning VAD scores to images

We implement a three-stage machine learning method, illustrated in Figure 1, for assigning VAD scores to images in the StockTwits sample. We adopt the ResNet-50 model, developed by He et al. (2016) at Microsoft and validated as an effective tool for image identification tasks (Zhou et al., 2022). The ResNet-50 model belongs to the class of Convolutional Neural Networks (CNNs), which is robust for financial applications (Gu, Kelly, and Xiu, 2020; and Jiang, Kelly, and Xiu, 2023).

<Insert Figure 1 here>

In the first stage, we pre-train the ResNet-50 model using the ImageNet database developed by Deng et al. (2009) and updated by Russakovsky et al. (2015). The ImageNet database contains over 14 million labeled images across approximately 21,000 categories and is widely used to train and evaluate algorithms for image classification, object detection, and other visual-recognition tasks. Panel A of Figure 2 shows a selected sample from the ImageNet database. During the pre-training stage, the ResNet-50 model learns to extract relevant image features and develops a general understanding of visual patterns and concepts. This process allows the ResNet-50 model to adapt these features to the distinct attributes inherent in the training datasets in the later stages.

<Insert Figure 2 here>

We then use image datasets to train the pre-trained ResNet-50 model to build the Filter ResNet-50 model and the Categorization ResNet-50 model. We obtain the training set of emotion images from four image databases, including the International Affective Picture System (IAPS) (Lang and Bradley, 2007), the Geneva Affective Picture Database (GAPED) introduced by Dan-Glauser and Scherer (2011), the Military Affective Picture System (MAPS) (Goodman, Katz, and Dretsch, 2016), and the Open Affective Standardized Image Set (OASIS) (Kurdi, Lozano, and Banaji, 2017). These databases have undergone rigorous validation processes and are widely recognized in the fields of emotion recognition, affective computing, and psychology, enabling us to achieve high accuracy in assigning valence, arousal, and dominance scores to the StockTwits sample. Panel B of Figure 2 shows a sample of pictures randomly selected from our training sets.[5] Internet Appendix 1 gives details about these four image training datasets.

In total, our training dataset contains 3,062 images, each of which is assigned valence and arousal scores. Among these 3,062 images, 1,432 are assigned scores for valence, arousal, and dominance because the score ratings for pictures on the IAPS and MAPS are based on valence, arousal, and dominance, while the ratings for pictures on the GAPED and OASIS are based on valence and arousal without dominance. Additionally, we apply the normalization method in Peng et al. (2004) to preprocess 3,062 images in the training datasets. Specifically, we resize the image to 224 × 224 pixels and adjust the RGB dimensions of the pictures such that both the mean and standard deviation are equal to 0.5. We augment our training dataset by randomly selecting 3,000 emotion-related images in PNG, JPG, and GIF formats from the StockTwits image pool to ensure a balanced distribution of samples across different classes in the training set.

[5] As per the contract agreements with the database providers of IAPS and MAPS, we are not permitted to display any images from these sources because they are proprietary and are not publicly accessible.

After the training process, in the second stage we use the Filter ResNet-50 model to exclude finance-related images from our initial StockTwits sample of 6,980,000 images. These finance-related images include price and candlestick charts that convey no emotions and may introduce noise when assigning VAD scores to images. Panel A of Figure 3 provides a sample of finance-related images. The selection of images intended to stimulate emotional responses is governed by certain attributes, including color and facial expressions (Mayer, DiPaolo, and Salovey, 1990). Notably, these elements do not resemble the characteristics inherent in financial figures and charts. In addition, the analysis of price charts and financial figures necessitates professional technical interpretation, a process outside the scope of machine learning algorithms' capabilities due to the lack of relative training sources in the image databases.

<Insert Figure 3 here>

The ResNet-50 model has a key innovation—residual or skip connections—which allows information to flow directly from one layer to another, thereby addressing the problem of vanishing gradients and enabling the training of deep neural networks using image databases. Therefore, we can train the Filter ResNet-50 model based on the pre-trained ResNet-50 model from the first stage by connecting it to three fully connected layers (FC layers), a building block of the neural network. Within each FC layer, we introduce nonlinearity into the neural network by applying the Rectified Linear Unit (ReLU) activation developed by Goodfellow, Bengio, and Courville (2016). Specifically, we input 2,048 dimensions into the first FC layer, as in the method of He et al. (2016). This is then reduced to an output of 64 dimensions, directed toward the subsequent FC layer. The second layer further processes the data and delivers 16 dimensions to the third FC layer. This successive reduction in dimensions aids in the refinement of the model. Finally, the third FC layer outputs two dimensions. We construct an image filter using the third FC layer by setting emotional images to 0 and finance-related images to 1. Our

final StockTwits sample contains 2,046,190 images, including 1,319,545 with PNG and JPG formats and 726,645 with GIF format. Figure 4 shows a collection of randomly chosen images from the final image sample.

<Insert Figure 4 here>

The other task in the second stage is to select realistic images from among those that have excluded finance-related images because the images in the StockTwits sample include both realistic and cartoon images. Compared with cartoons, realistic images in virtual reality (VR) environments can elicit stronger cortical activity and evoke greater emotional responses in viewers (Tian and Zhang, 2021). Moreover, applying the ResNet-50 model to realistic images yields more accurate results than using cartoons because the images in the four training databases (IAPS, GAPED, MAPS, and OASIS) primarily consist of realistic images depicting people, objects, landscapes, and animals. Thus, we train the Categorization ResNet-50 model to differentiate between realistic and cartoon images. We follow a procedure similar to that of the Filter ResNet-50 model and assign a value of 0 to realistic images and 1 to cartoon images. The training sample for realistic images includes the images in the four VAD image datasets and randomly selected 3,000 realistic images from the StockTwits sample to balance the sample distribution. As a result, the Categorization ResNet-50 model identifies 879,695 realistic images out of 1,319,545 PNG and JPG images and 484,430 realistic images out of 726,645 GIF images. Panels B and C in Figure 3 show selected samples of realistic and cartoon images, respectively.

Finally, in the third stage, we use the VAD image datasets to train the pre-trained ResNet-50 model to develop the Score ResNet-50 model and assign VAD scores to each image in the final StockTwits sample, in which all finance-related images have been excluded. As images in the IAPS, GAPED, MAPS, and OASIS are rated with VAD scores ranging from 0 to 10, we use these 3,062 images as the training set. We begin the training process by

connecting the model to three FC layers. In each FC layer, we use ReLU activation to implement nonlinearity of the neural network. Figure 5 illustrates the method used to assign valence, arousal, and dominance scores to each image in the StockTwits sample, employing the pre-trained Resnet-50 model. Specifically, we input 2,048 dimensions into the first FC layer and then output 64 dimensions into the second FC layer. Finally, the output layer containing 16 dimensions exported from the second layer is converted into one-dimensional scalars, which are the VAD scores for each image in the StockTwits sample.

<Insert Figure 5 here>

To evaluate the performance of the Score ResNet-50 model in assigning VAD scores, we divide the StockTwits sample into a training set consisting of 70% of the images and a validation set consisting of the remaining 30%. The purpose of this division is to ensure that the model genuinely learn to generalize its predictions to new, unseen data. For training, we utilize the Adam optimizer, which was enhanced by Kingma and Ba (2014), and complete 30 iterations with a learning rate of 10^{-3}. We employ Mean Absolute Error (MAE) loss as a metric to measure performance.[6] Figure 6 illustrates the performance of the scoring ResNet-50 model by presenting the MAE losses across all 30 epochs. In the final epoch, the MAE loss is recorded as 0.88, indicating an average loss error between the predicted and actual scores of only 0.293 (0.88 divided by 3).[7] The low average loss error demonstrates the excellent performance of the scoring ResNet-50 model. This also confirms that our machine learning approach, which utilizes a deep learning model and convolutional neural network, effectively captures emotions from big data in StockTwits posts (Sohangir et al., 2018).

[6] The MAE loss calculates the average absolute difference between the paired observations, providing a straightforward and intuitive measure of the average magnitude of errors without considering their directions. By disregarding the sign of the differences, MAE focuses solely on the magnitude of the errors and clearly indicates how far, on average, the paired observations deviate from each other.

[7] As the MAE loss constitutes the cumulative loss of three distinct components—valence, arousal, and dominance—it is necessary to divide the resulting figure by 3. The range of VAD scores is from 0 to 10. Measuring the MAE loss individually for each component separately is not feasible, as there is a process known as Backpropagation in machine learning. This process learns complex mappings from inputs to outputs and only generates a single MAE loss for a machine-learning model.

<Insert Figure 6 here>

4. Construction of emotion indicators from images and textual data

4.1. Daily imagery valence, arousal, and dominance indexes

The Score ResNet-50 model can straightforwardly output VAD scores for static PNG and JPG images. However, GIF format images are animated and contain multiple frames of 10–200 frames. Obaid and Pukthuanthong (2022) adopt static news pictures in the JPG or PNG format and distinguish only between news pictures with high and low sentiment. We transfer a GIF image into a series of static images, as in Jou, Bhattacharya, and Chang (2014), and then compute the image-level VAD scores by averaging the VAD scores of all frames in a GIF image, as in Equations (1a), (1b), and (1c):

$$ImageValence\ for\ GIF_{mt} = \frac{\sum_{n=1}^{N_m} valence\ score\ for\ frame_n}{N_m} \quad (1a)$$

$$ImageArousal\ for\ GIF_{mt} = \frac{\sum_{n=1}^{N_m} arousal\ score\ for\ frame_n}{N_m} \quad (1b)$$

$$ImageDominance\ for\ GIF_{mt} = \frac{\sum_{n=1}^{N_m} dominance\ score\ for\ frame_n}{N_m} \quad (1c)$$

where GIF_{mt} represents GIF image m on day t, and $frame_n$ is frame n in the GIF image, and N_m represents the total number of frames in GIF image m.

Figure 7 presents a collection of StockTwits post images in our sample that exhibit high and low levels of arousal and dominance, respectively. For the purpose of constructing VAD indexes, we adjust the range of VAD scores to between 0 and 1. We classify an image as having a low level of valence, arousal, or dominance if its value is below 0.33, and as having a high level if its value exceeds 0.67 by following Mohammad (2020).

<Insert Figure 7 here>

We then construct our daily imagery indexes for valence, arousal, and dominance, respectively, by dividing the sum of the image-level values of valence, arousal, and dominance across GIF, PNG, and JPG images by the total number of posted images each day. Thus, the imagery valence (arousal and dominance) index values on day t are the average values of

valence (arousal and dominance), for all posts that contain imagery content on day t as shown in Equations (2a), (2b), and (2c):

$$ImgValence_t = \frac{\sum ImageValence_{lt}}{L_t} \quad (2a)$$

$$ImgArousal_t = \frac{\sum ImageArousal_{lt}}{L_t} \quad (2b)$$

$$ImgDominance_t = \frac{\sum ImageDominance_{lt}}{L_t} \quad (2c)$$

where $ImageValence_{lt}$, $ImageArousal_{lt}$, and $ImageDominance_{lt}$, respectively, are the image-level VAD scores for imagery content on day *t*, and L_t is the number of images posted on day *t*.

4.2. Daily textual valence, arousal, and dominance indexes

We also construct the daily textual emotion dimension indexes using conventional linguistic words and emojis from StockTwits posts via machine learning. Garcia, Hu, and Rohrer (2023) utilize machine learning techniques to develop a finance-specific lexicon by labeling words sourced from *Wall Street Journal* articles, while Renault (2017) employs Stocktwits posts for the same purpose. As the raw textual content is formatted in sentences and includes marks that cannot be recognized by natural language processing, we apply Python Libraries, including Pandas and the DateTime module, to read the raw textual data from the metadata Pickle file with a defined time span. We then break every post into sentences and words after removing all marks, stop words, and website links in the second step. Finally, we apply the Natural Language Toolkit (NLTK) developed by Bird, Loper, and Klein (2009) before using lexicons to extract investor emotions from the media text. Specifically, we apply the Porter Stemming algorithm from the NLTK to obtain word prototypes and output lower-case word prototypes stored in a generic structure. Emojis are also considered textual content in our study, so they are processed together with words. We treat an emoji as a word in a post when determining post-level VAD values.

We then input the preprocessed textual data into the lexicon-based textual analysis and apply our LM-Renault-augmented word and emoji VAD lexicon to compute the VAD values for each word and emoji of a post. We apply the pre-trained word embedding model GloVe to construct our LM-Renault-augmented word and emoji VAD lexicon. Specifically, we assign VAD scores to the words in the Loughran and McDonald (2011) word list and Renault (2017) sentiment word list. We then add these words into the National Research Council Valence-Arousal-Dominance lexicon developed by Mohammad (2018). In addition to textual characters, we also include emojis with VAD scores in this lexicon. Internet Appendix 2 presents the details on the construction of LM-Renault-augmented word and emoji VAD lexicon, and Table 1 in Internet Appendix 2 presents a selected sample from this lexicon.

Next, we construct, for each day, the valence, arousal, and dominance emotion dimension indexes. To this end, we first determine valence, arousal, and dominance values at the post level. We employ the SentiStrength method in Thelwall et al. (2010) and Mäntylä et al. (2016) as we use textual data extracted from social media.[8] Specifically, for a post p that contains a list of K words, this method first measures the values of valence V_k, arousal A_k, and dominance D_k, respectively, of each word $k = 1, 2, \ldots K$ in a post. We map each word in a post onto our LM-Renault-augmented word and emoji VAD lexicon to obtain the word-level valence, arousal, and dominance values, respectively.

In order to compute the post-level valence, arousal, and dominance values, we first calculate the average value of valence $\bar{V}$, and those of arousal $\bar{A}$ and dominance $\bar{D}$, across all

[8] The SentiStrength method, an industrial-strength tool first developed by Thelwall et al. (2010), is specifically designed to estimate the degree of high or low sentiment in brief text pieces extracted from Twitter. Its unique advantage lies in its ability to analyze sentiments accurately, even in informal language. Since it is designed with an emphasis on social media conversations, the SentiStrength method demonstrates precision at the human level when applied to short English social web texts. Mäntylä et al. (2016) further enhance the SentiStrength method and enable it to quantify arousal and dominance scores for a piece of post text. When applied to brief English texts from social media, this methodology demonstrates an accuracy comparable to that of human-based analyses (Thelwall, Buckley, and Paltoglou, 2012; and Saif et al., 2016).

words and emojis in the 23,226-unigram LM-Renault-augmented word and emoji VAD lexicon. Next, for each post in a day, we identify a word in the post that has the maximum valence value and a word in the post that has the minimum valence value. Likewise, we identify the maximum and minimum values of arousal and dominance, respectively, in a post. Equation (3a) shows the calculation for the post-level textual valence value. We give a valence value to post *p* in day *t* as the difference between the maximum $Max\ (V_k)$ and the minimum $Min\ (V_k)$ values in the post if the average value $\bar{V}$ of the VAD lexicon is between $Max\ (V_k)$ and $Min\ (V_k)$; when the maximum value in the post is less than $\bar{V}$, the value of post *p* in day *t* is the difference between $\bar{V}$ and $Max\ (V_k)$; when the minimum value in the post is larger than than $\bar{V}$, the value of post *p* in day *t* is the difference between $Min\ (V_k)$ and $\bar{V}$. Equations (3b) and (3c) show, respectively, the calculations for the post-level textual arousal and dominance values:

$$TextValence_{pt}=$$

$$Max\ (V_k) - Min\ (V_k), if\ Min\ (V_k) < \bar{V} < Max(V_k);$$

$$\bar{V} - Max\ (V_k), if\ Max\ (V_k) < \bar{V};$$

$$Min\ (V_k) - \bar{V}\ , if\ \bar{A} < Min\ (V_k). \quad (3a)$$

$$TextArousal_{pt}=$$

$$Max\ (A_k) - Min\ (A_k), if\ Min\ (A_k) < \bar{A} < Max(A_k);$$

$$\bar{A} - Max\ (A_k), if\ Max\ (A_k) < \bar{A};$$

$$Min\ (A_k) - \bar{A}\ , if\ \bar{A} < Min\ (A_k). \quad (3b)$$

$$TextDominance_{pt}=$$

$$Max\ (D_k) - Min\ (D_k), if\ Min\ (D_k) < \bar{D} < Max(D_k);$$

$$\bar{D} - Max\ (D_k), if\ Max\ (D_k) < \bar{D};$$

$$Min\ (D_k) - \bar{D}, if\ \bar{D} < Min\ (D_k). \quad (3c)$$

For illustration, Table 1 lists all words that can be mapped to the LM-Renault-augmented word and emoji VAD lexicon from a StockTwits post illustrated in Figure 8.

According to Equations (3a), (3b), and (3c), the textual valence, arousal, and dominance values of this post are 0.520, 0.452, and 0.373, respectively.

<Insert Table 1 here>

<Insert Figure 8 here>

Having measured the emotion dimension value of textual content in a single StockTwits post, we then construct the daily textual valence, arousal, and dominance indexes by dividing the cumulative post-level textual values of valence, arousal, and dominance, respectively, by the total number of posts per day. Thus, the valence (arousal, dominance) index values on day t are the average values of the valence (arousal, dominance) of all posts in the day, as shown in Equations (4a), (4b), and (4c):

$$TextValence_t = \frac{\sum TextValence_{pt}}{P_t} \tag{4a}$$

$$TextArousal_t = \frac{\sum TextArousal_{pt}}{P_t} \tag{4b}$$

$$TextDominance_t = \frac{\sum TextDominance_{pt}}{P_t} \tag{4c}$$

where $TextValence_{pt}$, $TextArousal_{pt}$, and $TextDominance_{pt}$ are the post-level valence, arousal, and dominance values of the textual content, respectively, on day t. P_t is the total number of StockTwits posts on day t.

Figure 9 presents the time series of both daily imagery and textual indexes. All three series exhibit spikes during market turmoil over the 10-year sample period.[9] The imagery indexes appear less volatile than the textual indexes but have more extreme spikes. This finding occurs because, unlike textual content, the StockTwits sample contain numerous repetitive images (see also, Gu, Teoh, and Wu, 2023). The imagery and textual valence indexes, displayed in Panel A, reached a peak in March 2012, indicating a surge in market-level

[9] We also plot three imagery indexes every year together with daily S&P 500 index returns. These detailed plots show clear time-series variation and are available upon request.

optimism. This followed a nearly one-trillion-euro boost in the Eurozone bailout fund, which strongly supported the crisis-hit Eurozone nations, such as Germany and Greece. In December 2017, the valence indexes experienced positive spikes following Congress's approval of a Republican tax bill that reduced the U.S. corporate tax rates from 35% to 21%. The indexes then reached low levels in February 2018 as concerns mounted over high inflation. In June 2019, the valence indexes remained higher as U.S. President Trump promised to negotiate trade with China at the G-20 summit. The index decreased in 2020 as the COVID-19 pandemic hit but trended higher in December 2020 as the U.S. stock market ended the year at an all-time high, with a 16% annual increase, indicating a strong recovery from the initial decline caused by the COVID-19 pandemic.

<Insert Figure 9 here>

Furthermore, the arousal index has demonstrated an upward trajectory in recent years, indicating an increase in investors' inclination to express their opinions. This index exhibited intermittent spikes, notably between 2018 and 2019, which coincided with the strongest quarterly performance of the S&P 500 index in the past five years, with a remarkable 7.2% increase. The arousal index remained elevated when the stock market returns surpassed 28% in 2019.

Similarly, the dominance index displayed sporadic sharp spikes throughout the analyzed period, where positive spikes reflected investors' confidence and sense of control, and negative spikes indicated their uncertainty towards prevailing circumstances. For instance, the dominance index displayed an upward trend in 2014, coinciding with the solid growth of the U.S. economy, which recorded a GDP growth rate of 10.5%, the highest in over a decade. The index peaked in late November 2017 after the passage of the Tax Cuts and Jobs Act by the House of Representatives. It reached another peak in December 2019 following the Federal Reserve's expenditure of $428 million to acquire debt from individual firms. However, the

dominance index declined in March 2020, triggered by the implementation of COVID-19-related shutdowns by the U.S. government. It subsequently rebounded in March 2021, when the government began easing COVID-19 restrictions, and certain states, such as Texas and Mississippi, announced plans for full reopening. Finally, we observe that imagery and textual indexes become more closely aligned during market turmoil than during normal times.

4.3. Descriptive statistics

Panel A in Table 2 reports the summary statistics of the daily imagery emotion dimension indexes, namely, imagery valence, arousal, and dominance, quantified based on all images in our StockTwits sample. On average, the imagery valence index is 0.287, lower than 0.5, indicating that StockTwits users are more likely to insert images containing low sentiments.

<Insert Table 2 here>

The mean value of the textual valence index, reported in Panel B of Table 2, is 0.393, confirming that textual sentiment in the stock market tends to be low (e.g., Garcia, 2013), as it is lower than 0.5. The mean imagery arousal and dominance values in Panel A are 0.574 and 0.515, respectively. The mean values of the textual arousal and dominance in Panel B are 0.518 and 0.524, respectively. Both the imagery and textual indexes of arousal and dominance have mean values higher than 0.5, indicating that StockTwits users are more likely to express opinions when they exhibit emotions with high levels of arousal or dominance. Our findings are consistent with the theory that arousal is crucial for motivating certain behaviors, such as writing posts when individuals perceive stimuli. The mean value of imagery dominance also shows that individuals who experience high dominance emotions believe that they can control the current situation, making them feel comfortable expressing comments or thoughts (Graziotin, Wang, and Abrahamsson, 2015).

Importantly, as shown in Panel C, the three emotion dimensions of both imagery and textual content have relatively low correlations in general, consistent with the findings of

Mohammad (2018). The relatively low correlations suggest that some of the information in the images is distinct from that in the text of posts on StockTwits. As we further test the influence of the imagery emotion dimension indexes constructed based on realistic images, we also report the descriptive statistics in Panel D of Table 2.

5. Econometric analyses

This section examines the impact of investors' emotion dimensions on stock returns. We first use imagery emotion dimension indexes based on all images and then include textual emotion dimension indexes. Next, we focus on the impact of imagery emotion dimension indexes based on realistic images, together with textual emotion dimension indexes, on stock returns.

5.1. The influences of all-image imagery content and textual content on stock market returns

Our analysis begins by examining the impact of daily imagery emotion dimension indexes constructed based on all images. Following Tetlock (2007) and Obaid and Pukthuanthong (2022), we run a time-series regression using the lagged values of the emotion dimension indexes for imagery content, as shown in Equation (5):

$$R_t = \beta_1 L5(ImgEdim_{j,t}) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t \quad (5)$$

The daily returns on the S&P 500 index, denoted as R_t, are obtained from the Center for Research in Security Prices (CRSP). The emotion indicator for imagery content, $ImgEdim_{j,t}$, consists of three imagery emotion dimension indexes, $ImgValence_t$, $ImgArousal_t$, and $ImgDominance_t$, representing the daily values of valence, arousal, and dominance of the imagery contents, respectively. The coefficients on $ImgEdim_{j,t}$ represent the dependence of the S&P 500 Index on the image emotion dimension indexes. We account for lagged effects using the lag operator $L5$, where $L5(Variable)_t$ represents the variable at five lags, as per Tetlock (2007). Following Garcia (2013) and Obaid and Pukthuanthong (2022), we incorporate five lags of the S&P 500 daily returns and detrended squared return residuals.

The set of exogenous variables X_t comprises a constant term, the daily CBOE volatility index (VIX) obtained from the FED St. Louis, and dummy variables for the day-of-the-week and January effects, as employed by Tetlock (2007), to account for potential return anomalies. We use Newey and West's (1987) heteroskedasticity-and-autocorrelation robust standard errors.

We start by analyzing the influence of the daily imagery valence index on the S&P 500 index return. Panel A of Table 3 reports the OLS estimates of the coefficients β_1 in Equation (5). Each coefficient measures the effect of a one standard deviation increase in the imagery valence index on the S&P 500 index returns in basis points, with one basis point equal to a daily return of 0.01%. We find that the imagery valence index on day *t-1* positively influences stock market returns on day *t* with statistical significance at the 1% level. On average, a one standard deviation change in the imagery valence index in *t-1* exerts a positive impact of 2.37 basis points on the S&P 500 return, indicating that high valence levels in messages posted on StockTwits are followed by increased returns on subsequent trading days. Our finding is in line with that of Garcia (2013), who shows that investor optimism leads to a positive increase in stock prices. The initial increase in returns is consistent with the finding of Obaid and Pukthuanthong (2022). Additionally, the effect of the fifth trading day on day *t* is 3.17 basis points, which is statistically significant at the 5% level.

<Insert Table 3 here>

We then analyze the effects of the imagery arousal emotion dimension on stock market returns. To this end, we re-estimate Equation (5) using $ImgArousal_t$, the imagery arousal index, as the main explanatory variable. Panel B of Table 3 shows that the imagery arousal index influences market-level returns, with the coefficients on the first, third, and fifth lags being statistically significant at the 10% level. These results are consistent with our hypothesis that arousal predicts positive returns over the trading week and support the conjecture that individual investors, influenced by high-arousal emotions, make purchase decisions.

Next, we focus on the effect of dominance on daily stock market returns. We re-estimate Equation (5) using $ImgDominance_t$, the imagery dominance index, as the main explanatory variable. The results reported in Panel C of Table 3 show that the daily dominance index predicts upward increases in market prices, where a one standard deviation increase in the daily imagery dominance index on day *t-1* predicts an increase of 3.37 basis points in the S&P 500 index return on day *t*. This finding is consistent with our hypothesis that dominance positively predicts stock market returns. The initial return increase is followed by continuous increases in the second and third lags, with magnitudes of 3.74 and 3.73 basis points, respectively; both are significant at the 5% level. This finding suggests that the influence of dominance on stock market returns are long-lasting, consistent with Liu and Sourina (2012), who find that individuals need more time to move away from the influence of dominance emotions than from the effects of valence and arousal emotions. Our finding that the imagery emotion dimension indexes impact market returns for several days suggests enduring effects of emotions, in line with Fox et al. (2004), who contend that individuals maintain a sustained attentional focus on images appearing in television news stories.

We further introduce the textual emotion dimension indexes, together with the three imagery emotion dimension indexes, respectively. First, we use both imagery and textual valence indexes to estimate Equation (6):

$$R_t = \beta_1 L5(ImgEdim_{j,t}) + \gamma_1 L5(TextEdim_{j,t}) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t \quad (6)$$

where $TextEdim_{j,t}$ is the series of emotion indicators for the textual content, including $TextValence_t$ for the daily textual valence index, $TextArousal_t$ for the daily arousal index, and $TextDominance_t$ for the daily dominance index.

Panel A of Table 4 reports the regression results. We find that the coefficients of the imagery valence index remain positive and statistically significant after controlling for the textual valence index. We also find that the textual valence index can predict market-level

return in which a one standard deviation increase in the daily textual valence index on day *t-1* predicts an increase of 10.51 basis points in the S&P 500 index return on day *t*, significant at the 1% level. The return increase is immediately followed by a decrease of 6.02 basis points on the next day, as reflected by the coefficient on the second lag, which shows a reversal from the initial positive return impact of valence.

<Insert Table 4 here>

Similarly, Panels B and C in Table 4 show that both daily imagery arousal and dominance indexes impact market returns. The coefficients on the imagery arousal and dominance indexes remain positive and statistically significant with the inclusion of textual arousal and dominance indexes. In addition, both textual arousal and dominance indexes predict initial return increases and subsequent return reversals. Specifically, a one standard deviation increase in the daily textual arousal index on day *t-1* impacts the S&P 500 index return on day *t* with an increase of 6.93 basis points. This is immediately followed by a return decrease of 4.56 basis points the next day, as reflected by the coefficient on the second lag. For dominance, a one standard deviation increase in the daily textual dominance index on day *t-1* predicts an increase of 6.52 basis points in the S&P 500 index return in day *t*, which is then immediately followed by a decrease of 8.87 basis points and 5.62 basis points in the second and third lags, respectively.

Overall, our findings show that the valence, arousal, and dominance indexes extracted from imagery data on StockTwits positively predict stock market returns, whereas textual emotion dimension indexes positively predict initial market-level stock returns and subsequent return reversals. Our evidence suggests that images inserted into StockTwits posts express investors' opinions, making them an ideal source for capturing investor emotions. Our findings also show the influence of visual content in social media on financial markets, consistent with Obaid and Pukthuanthong (2022).

5.2. The influences of realistic-only imagery content and textual content on stock market returns

We distinguish between realistic and cartoon images and then construct emotion dimension indexes based on only realistic images. We first estimate the impact of realistic imagery valence index on S&P 500 index returns. Panel A of Table 5 reports the OLS estimates of coefficients β_1 in Equation (5). Each coefficient measures the impact of a one standard deviation increase in the realistic imagery valence index on the returns in basis points (one basis point equals a daily return of 0.01%). We find that the effect of the realistic imagery valence index on market-level stock returns is positive from the first to the third lags. Notably, a one standard deviation increase in the realistic imagery valence index on day *t-3* predicts an increase of 3.08 basis points, on average, in the S&P 500 return on day *t* with statistical significance at the 5% level. This increase in returns is mostly reversed later in the trading week. The scale of the reversal in lag four is 3.70 basis points, which significantly differs from zero at the 1% level. Our results indicate that the realistic imagery valence index, assembled using realistic images on StockTwits, can predict stock market return reversals. In addition, the effect of emotions expressed in realistic images differs from that when all images are used. The *p*-value is 0.0052 for the test of the sum of the coefficients on lags two to five, showing that the lag effects of image valence on stock market returns are statistically significant.

<Insert Table 5 here>

We then analyze the effects of the realistic imagery arousal index on stock market returns. We re-estimate Equation (5) using the realistic imagery arousal index as the main explanatory variable. Panel B of Table 5 shows that the realistic imagery arousal index on day *t-1* predicts positive and statistically significant market returns on day *t*, indicating that days with high-arousal emotional levels extracted from realistic images are followed by high returns the next day. On average, a one standard deviation increase in the realistic imagery arousal

index in a day predicts an increase of 1.38 basis points in the S&P 500 index returns the next day. Furthermore, the impacts of the third, fourth, and fifth lags on returns on day *t* are all positive with high statistical significance. These results indicate stronger impacts of the arousal index specific to realistic images on market returns than using all images (Panel B in Table 3). This finding is consistent with Tian and Zhang's (2021) finding that realistic images can instigate heightened cortical activity, thereby eliciting more intense emotion arousal in viewers.

Panel C of Table 5 reports the regression results obtained by employing $ImgDominance_t$ based on only realistic images. We find that the daily realistic imagery dominance index predicts increases in market prices, as indicated by the positive coefficients from day *t-1* to day *t-3*. A one standard deviation augmentation in the daily dominance index on day *t-3* predicts an average increase of 2.85 basis points in the S&P 500 index return on day *t*, significant at the 1% level. This return escalation is followed by a reversal in the fourth lag, with a magnitude of 3.46 basis points and statistical significance at the 1% level.

We further estimate Equation (6) to include both realistic imagery indexes and textual emotion dimension indexes. Panel A of Table 6 reports the results obtained using the realistic imagery valence index. We find that both the second and third lags of the realistic imagery valence index have strong and positive impacts on market returns. For example, a one standard deviation increase in the realistic imagery valence index on day *t-2* predicts an increase of 2.96 basis points on the S&P 500 index return on day *t*, which is significant at the 5% level. The patterns of the impacts of the third and fourth lags are similar to those without the inclusion of the textual valence dimension index (Panel A, Table 5). Panel B of Table 6 shows that the realistic imagery arousal index has strong effect on market returns. Specifically, a one standard deviation augmentation in the realistic imagery arousal index on day *t-1* predicts an increase of 1.32 basis points on the S&P 500 index return on day *t*, which is statistically significant at the 5% level. The coefficients on lags 3 through 5 are all highly statistically significant at the 1%

level, with magnitudes of 2.98 basis points, 4.38 basis points, and 2.91 basis points, respectively. Panel C in Table 6 shows particularly stronger effects of the realistic imagery dominance index than those without including the textual dominance dimension index (Panel C in Table 5). Specifically, a one standard deviation increase in the realistic imagery dominance index on days *t-2* and *t-3* predicts an increase of 4.17 basis points and 6.56 basis points, respectively, in the S&P 500 index return on day *t*. In addition, a reversal emerges on day *t*-4 with a scale of 6.74 basis points that is statistically significant at the 1% level.

<Insert Table 6 here>

Overall, we find that all three realistic imagery indexes exert stronger impacts, both economically and statistically, on market returns than those without the textual emotion dimension indexes. On the other hand, the pattern of the impacts of the three textual emotion dimension indexes is similar to that reported in Table 4. Our findings suggest that realistic images in posts evoke strong emotional responses, which is in line with Gu and Han (2007), who show that the activation of neural substrates is elevated in response to stimuli generated by realistic images, whereas neural engagement appears to diminish when subjects are presented with cartoon images, which degraded the stimulus.

5.3. Robustness tests

We conduct robustness checks by first modifying the imagery emotion dimension indexes following Jou, Bhattacharya, and Chang's (2014) method. Specifically, we calculate the VAD scores of the GIF images by considering only the last frame. The results reported in Tables 1 to 4 in Internet Appendix 3 are quantitatively similar to those reported in Tables 3 to 6 and do not change our conclusions. Further, we control for Fama and French (2015) five factors in all our analyses, and our conclusions remain unchanged. The results are presented in Internet Appendix 4.

We further evaluate the out-of-sample return predictability of the imagery emotion dimension indexes. Our previous analysis was based on the entire sample (in-sample) to estimate the return predictive ability of the VAD scores of imagery content. To mitigate the potential issues of in-sample overfitting, we assess the predictive power of our model out-of-sample. Following Welch and Goyal (2008) and Campbell and Thompson (2008), we evaluate the out-of-sample predictive performance using Equation (7):

$$R_{oos}^2 = 1 - \frac{\sum_{t=1}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=1}^{T}(r_t - \bar{r}_t)^2} \quad (7)$$

where $\hat{r}_t$ represents the fitted value for returns obtained from a predictive regression model that predicts future stock returns based on the imagery emotion dimension indexes with information available at time *t*-1. $\bar{r}$ denotes the historical average benchmark, which is estimated up to period *t*-1 using the constant expected return model. We set the estimation period from 1 January 2012 to 31 December 2019, which allows us to assess the model's predictive performance from 1 January 2020 to 31 December 2021.

<Insert Table 7 here>

The out-of-sample *R*-squared, R_{oos}^2, measures the relative decrease in the mean-squared prediction error (MSPE) for a given predictive regression compared with a benchmark based on historical average returns. The R_{oos}^2 statistic takes values in the range (-∞, 1]. If $R_{oos}^2 > 0$, it indicates that the predicted $\hat{r}$ (estimated return) performs better than the historical average $\bar{r}$ in terms of mean squared forecast error. In our analysis, we estimate R_{oos}^2 for the full image sample and the sample of realistic images. The realistic imagery valence and dominance indexes exert different influence patterns compared to imagery indexes with all images, while the imagery arousal index has a stronger effect on returns. The results, presented in Table 7, show the out-of-sample R_{oos}^2 for future returns in different image samples. The R_{oos}^2values range from 0.963% to 1.016%, consistently exceeding the threshold of 0. These findings demonstrate the robustness of our results when evaluated out-of-sample.

6. Conclusions

In this study, we extract investor emotions expressed in both imagery and textual content in posts on the social media platform StockTwits and examine whether emotion dimensions predict daily stock market returns. By integrating imagery and textual data, we aim to gain a comprehensive understanding of the relationship between investors' emotions and stock market returns. Our study addresses the importance of investor psychology on asset prices by adopting well-established psychological theories of emotions, the VAD dimension model of Russell and Mehrabian (1977), and the cognitive load theory of Sweller, Van Merrienboer, and Paas (1998). We employ advanced machine-learning algorithms that identify and classify the images present in StockTwits posts to analyze the imagery content. These algorithms assign valence, arousal, and dominance scores to each image, thereby allowing the construction of daily imagery emotion dimension indexes. We further exclude cartoons to construct emotion dimension indexes based on realistic images to explore the influence of image genres.

We summarize our findings as follows. First, we find that the daily imagery indexes of valence, arousal, and dominance strongly and positively predict returns on the S&P 500 index. The inclusion of textual data strengthens the influence of imagery content on stock returns. Second, realistic imagery emotion indexes that exclude cartoon images have strong impacts on stock market returns. In particular, the arousal index derived from realistic images exhibits a more pronounced influence on market-level stock returns than the index created using all images. Our evidence suggests that imagery content, particularly that categorized as realistic images, plays a significant role in shaping investor emotions, which, in turn, influences stock returns. Our research sheds light on how investors' emotions, viewed through the lens of emotion dimensions, affect the stock market.

References

Anderson, G., and Brown, R. I. F. (1984). Real and laboratory gambling, sensation-seeking and arousal. *British Journal of Psychology*, *75*(3), 401-410.

Alpers, G. W., and Gerdes, A. (2007). Here is looking at you: emotional faces predominate in binocular rivalry. *Emotion*, 7(3), 495.

Arjaliès, D. L., and Bansal, P. (2018). Beyond numbers: How investment managers accommodate societal issues in financial decisions. *Organization Studies*, *39*(5-6), 691-719.

Bai, Q., Dan, Q., Mu, Z., and Yang, M. (2019). A systematic review of emoji: Current research and future perspectives. *Frontiers in Psychology*, *10*, 2221.

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

Barber, B. M., and Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, *21*(2), 785-818.

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of behavioral economics: applications and foundations 1* (Vol. 1, pp. 79-175). North-Holland.

Bazley, W. J., Cronqvist, H., and Mormann, M. (2021). Visual finance: The pervasive effects of red on investor behavior. *Management Scienc*e, 67(9), 5616-5641.

Bikhchandani, S., Hirshleifer, D., Tamuz, O., and Welch, I. (2021). Information cascades and social learning. *National Bureau of Economic Research*. WP. 28887.

Bird, S., Klein, E., and Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*, O'Reilly Media, Inc.

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59.

Campbell, J. Y., and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average?. *Review of Financial Studies*, *21*(4), 1509-1531.

Chen, H., De, P., Hu, Y., and Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367-1403.

Chen, H., and Hwang, B. H. (2022). Listening in on investors' thoughts and conversations. *Journal of Financial Economics*, *145*(2), 426-444.

Coleman, E. G. (2010). Ethnographic approaches to digital media. *Annual Review of Anthropology*, *39*, 487-505.

Coleman, R. (2010). Framing the pictures in our heads: Exploring the framing and agenda-setting effects of visual images. In *Doing News Framing Analysis*, 249-278, Routledge.

Cookson, J. A., Lu, R., Mullins, W., and Niessner, M. (2024). The Social Signal. *Journal of Financial Economics*, Forthcoming.

Cookson, J. A., and Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors, *Journal of Finance*, 75(1): 173-228.

Dai, W., Han, D., Dai, Y., and Xu, D. (2015). Emotion recognition and affective computing on vocal social media. *Information & Management*, *52*(7), 777-788.

Dan-Glauser, E. S., and Scherer, K. R. (2011). The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, *43*, 468-477.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248-255.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*(5596), 1191-1194.

Edmans, A., Garcia, D., and Norli, Ø. (2007). Sports sentiment and stock returns. *Journal of Finance*, *62*(4), 1967-1998.

Fama, E. F., and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1-22.

Frischen, A., Eastwood, J. D., and Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological Bulletin*, *134*(5), 662.

Fox, C., Kelley, E., Chung, Y., Lee, S., Schwartz, N., and Potter, D. (2004). Picture this: Effects of graphics on the processing of television news. *Journal of Broadcasting & Electronic Media*, *48*(4), 646-674.

Fox, J. R., Lang, A., and Paolucci, R. (2023). How Many Words is a Picture Worth? Using Emojis from Social Media to Predict Future Stock Returns. *SSRN Electronic Journal 4427860*.

Galentino, A., Bonini, N., and Savadori, L. (2017). Positive arousal increases individuals' preferences for risk. *Frontiers in Psychology*, *8*, 2142.

Garcia, D. (2013). Sentiment during recessions. *Journal of Finance*, *68*(3), 1267-1300.

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

Glimcher, P. W. (2003). The neurobiology of visual-saccadic decision making. *Annual Review of Neuroscience*, *26*(1), 133-179.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*, MIT press.

Goodman, A. M., Katz, J. S., and Dretsch, M. N. (2016). Military Affective Picture System (MAPS): A new emotion-based stimuli set for assessing emotional processing in military populations. *Journal of Behavior Therapy and Experimental Psychiatry*, *50*, 152-161.

Graziotin, D., Wang, X., and Abrahamsson, P. (2015). Do feelings matter? On the correlation of affects and the self-assessed productivity in software engineering. *Journal of Software: Evolution and Process*, *27*(7), 467-487.

Gu, X., and Han, S. (2007). Attention and reality constraints on the neural processes of empathy for pain. *Neuroimage*, *36*(1), 256-267.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223-2273.

Gu, M., Teoh, S. H., and Wu, S. (2023). GIF Sentiment and Stock Returns. *SSRN Electronic Journal 4110191*.

Hakokõngäs, E., Halmesvaara, O., and Sakki, I. (2020). Persuasion through bitter humor: Multimodal discourse analysis of rhetoric in internet memes of two far-right groups in Finland. *Social Media+ Society*, *6*(2), 2056305120921575.

Han, S., Lerner, J. S., and Keltner, D. (2007). Feelings and consumer decision making: The appraisal-tendency framework. *Journal of Consumer Psychology*, *17*(3), 158-168.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.

Heath, C., Bell, C., and Sternberg, E. (2001). Emotional selection in memes: the case of urban legends. *Journal of Personality and Social Psychology*, *81*(6), 1028.

Hemmig, W. (2009). An empirical study of the information-seeking behavior of practicing visual artists. *Journal of Documentation*, *65*(4), 682-703.

Highfield, T., and Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice*, *2*(1), 47-62.

Hirshleifer, D. (2001). Investor psychology and asset pricing. *Journal of Finance*, *56*(4), 1533-1597.

Hirshleifer, D., and Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *Journal of Finance*, *58*(3), 1009-1032.

Jahedi, S., Deck, C., and Ariely, D. (2017). Arousal and economic decision making. *Journal of Economic Behavior & Organization*, *134*, 165-189.

Jerram, M., Lee, A., Negreira, A., and Gansler, D. (2014). The neural correlates of the dominance dimension of emotion. *Psychiatry Research: Neuroimaging*, *221*(2), 135-141.

Javela, J. J., Mercadillo, R. E., and Ramírez, J. M. (2008). Anger and associated experiences of sadness, fear, valence, arousal, and dominance evoked by visual scenes. *Psychological Reports*, *103*(3), 663-681.

Jiang, J., Kelly, B., and Xiu, D. (2023). (Re-) Imag (in) ing Price Trends. *Journal of Finance*, *78*(6), 3193-3249.

Johnson, S. L., Leedom, L. J., and Muhtadie, L. (2012). The dominance behavioral system and psychopathology: evidence from self-report, observational, and biological studies. *Psychological Bulletin*, *138*(4), 692.

Jones, C. M., and Lamont, O. A. (2002). Short-sale constraints and stock returns. *Journal of Financial Economics*, *66*(2-3), 207-239.

Jou, B., Bhattacharya, S., and Chang, S. F. (2014, November). Predicting viewer perceived emotions in animated GIFs. In *Proceedings of The 22nd ACM International Conference on Multimedia*, 213-216.

Keltner, D., and Lerner, J. S. (2010). Emotion, In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Ed.), Handbook of Social Psychology, 317-352. New York, Wiley.

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

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

Kragel, P. A., Reddan, M. C., LaBar, K. S., and Wager, T. D. (2019). Emotion schemas are embedded in the human visual system. *Science Advances*, *5*(7), eaaw4358.

Kurdi, B., Lozano, S., and Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, *49*, 457-470.

Lang, P., and Bradley, M. M. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. *Handbook of Emotion Elicitation and Assessment*, 29–46. Oxford University Press.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*(39-58), 3.

Lerner, J. S., and Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, *81*(1), 146.

Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*, 799-823.

Lerner, J. S., Small, D. A., and Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science*, *15*(5), 337-341.

Liu, Y., and Sourina, O. (2012, July). EEG-based dominance level recognition for emotion-enabled interaction. In *2012 IEEE International Conference on Multimedia and Expo*, 1039-1044.

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

Mahmoudi, N., Docherty, P., and Moscato, P. (2018). Deep neural networks understand investors better. *Decision Support Systems*, 112, 23-34.

Mano, H. (1994). Risk-taking, framing effects, and affect. *Organizational Behavior and Human Decision Processes*, *57*(1), 38-58.

Mäntylä, M., Adams, B., Destefanis, G., Graziotin, D., and Ortu, M. (2016, May). Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity?. In *Proceedings of The 13th International Conference on Mining Software Repositories*, 247-258.

Mayer, J. D., DiPaolo, M., and Salovey, P. (1990). Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence. *Journal of Personality Assessment*, *54*(3-4), 772-781.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, *14*, 261-292.

Mohammad, S. (2018, July). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 174-184).

Mohammad, S. M. (2020). Practical and ethical considerations in the effective use of emotion and sentiment lexicons. *arXiv preprint arXiv:2011.03492*.

Nekrasov, A., Teoh, S. H., and Wu, S. (2022). Visuals and attention to earnings news on Twitter. *Review of Accounting Studies*, *27*(4), 1233-1275.

Newey, W. K., and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777-787.

Obaid, K., and Pukthuanthong, K. (2022). A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics*, *144*(1), 273-297.

Pedersen, L. H. (2022). Game on: Social networks and markets. *Journal of Financial Economics*, *146*(3), 1097-1119.

Peng, X. Q., Cao, J., Chen, J., Xue, P., Lussier, D. S., and Liu, L. (2004). Experimental and numerical analysis on normalization of picture frame tests for composite materials. *Composites Science and Technology*, *64*(1), 11-21.

Pennington, J., Socher, R., and Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of The 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543.

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness?. *Current Opinion in Neurobiology*, *15*(2), 188-196.

Porcelli, A. J., and Delgado, M. R. (2009). Acute stress modulates risk taking in financial decision making. *Psychological Science*, *20*(3), 278-283.

Qi, J., Du, J., Siniscalchi, S. M., Ma, X., and Lee, C. H. (2020). On mean absolute error for deep neural network based vector-to-vector regression. *IEEE Signal Processing Letters*, *27*, 1485-1489.

Renault, T. (2017). Intraday Internet investor sentiment and return patterns in the US stock market. *Journal of Banking and Finance*, 84, 25-40.

Riemer, H., and Viswanathan, M. (2013). Higher motivation-greater control? The effect of arousal on judgement. *Cognition and Emotion*, *27*(4), 723-742.

Riordan, M. A. (2017). Emojis as tools for emotion work: Communicating affect in text messages. *Journal of Language and Social Psychology*, *36*(5), 549-567.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... and Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*, 211-252.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161.

Russell, J. A. (1997). 13-reading emotion from and into faces: Resurrecting a dimensional-contextual perspective. *The Psychology of Facial Expression*, 295-320.

Russell, J. A., and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*(3), 273-294.

Saif, H., He, Y., Fernandez, M., and Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management*, *52*(1), 5-19.

Schnotz, W., and Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*, 469-508.

Shen, W., Hentschel, T., and Hideg, I. (2023). Leading through the uncertainty of COVID-19: The joint influence of leader emotions and gender on abusive and family-supportive supervisory behaviours. *Journal of Occupational and Organizational Psychology*.

Sheth, B. R., and Pham, T. (2008). How emotional arousal and valence influence access to awareness. *Vision Research*, *48*(23-24), 2415-2424.

Sohangir, S., Wang, D., Pomeranets, A., and Khoshgoftaar, T. M. (2018). Big Data: Deep Learning for financial sentiment analysis. *Journal of Big Data*, *5*(1), 1-25.

Sterelny, K. (2006). Memes revisited. *British Journal for the Philosophy of Science*, 57, 145–165.

Storbeck, J., and Clore, G. L. (2008). Affective arousal as information: How affective arousal influences judgments, learning, and memory. *Social and Personality Psychology Compass*, *2*(5), 1824-1843.

Sul, H. K., Dennis, A. R., and Yuan, L. (2017). Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, *48*(3), 454-488.

Sweller, J., Van Merrienboer, J. J., and Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 251-296.

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

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, *63*(1), 163-173.

Tian, F., and Zhang, W. (2021, November). Difference of Emotional Arousal Between Real Shooting and Cartoon Pictures in VR Environment-A Comparative Study Based on EEG Signals. In *2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)*, 547-550.

Toma, F. M., Cepoi, C. O., Kubinschi, M. N., and Miyakoshi, M. (2023). Gazing through the bubble: an experimental investigation into financial risk-taking using eye-tracking. *Financial Innovation*, *9*(1), 28.

Wadlinger, H. A., and Isaacowitz, D. M. (2006). Positive mood broadens visual attention to positive stimuli. *Motivation and Emotion*, *30*, 87-99.

Wegner, D. M., and Giuliano, T. (1980). Arousal-induced attention to self. *Journal of Personality and Social Psychology*, *38*(5), 719.

Welch, I., and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455-1508.

Wilson, T. D., and Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117.

Wong, A., Leahy, W., Marcus, N., and Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction*, *22*(6), 449-457.

Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation, *Journal of Comparative Neurology & Psychology*, *18*, 459-482.

Zhao, J., Meng, Q., An, L., and Wang, Y. (2019). An event-related potential comparison of facial expression processing between cartoon and real faces. *PLoS One*, *14*(1), e0198868.

Zhou, K., Yang, J., Loy, C. C., and Liu, Z. (2022). Learning to prompt for vision-language models. *International Journal of Computer Vision*, *130*(9), 2337-2348.

Figure 1. The three-stage machine learning process for image identification

The figure shows the machine-learning process for image identification in each of the three stages. In the first stage, we train the ResNet-50 model using images from ImageNet to develop a pre-trained ResNet-50 model. Based on the pre-trained ResNet-50 model, we construct Filter ResNet-50 models, with one for excluding finance-related images and then Categorization ResNet-50 model for distinguishing between realistic images and cartoon images among emotion images in Stage 2. We develop Score ResNet-50 model to compute valence, arousal, and dominance scores for images from the StockTwits database in Stage 3.

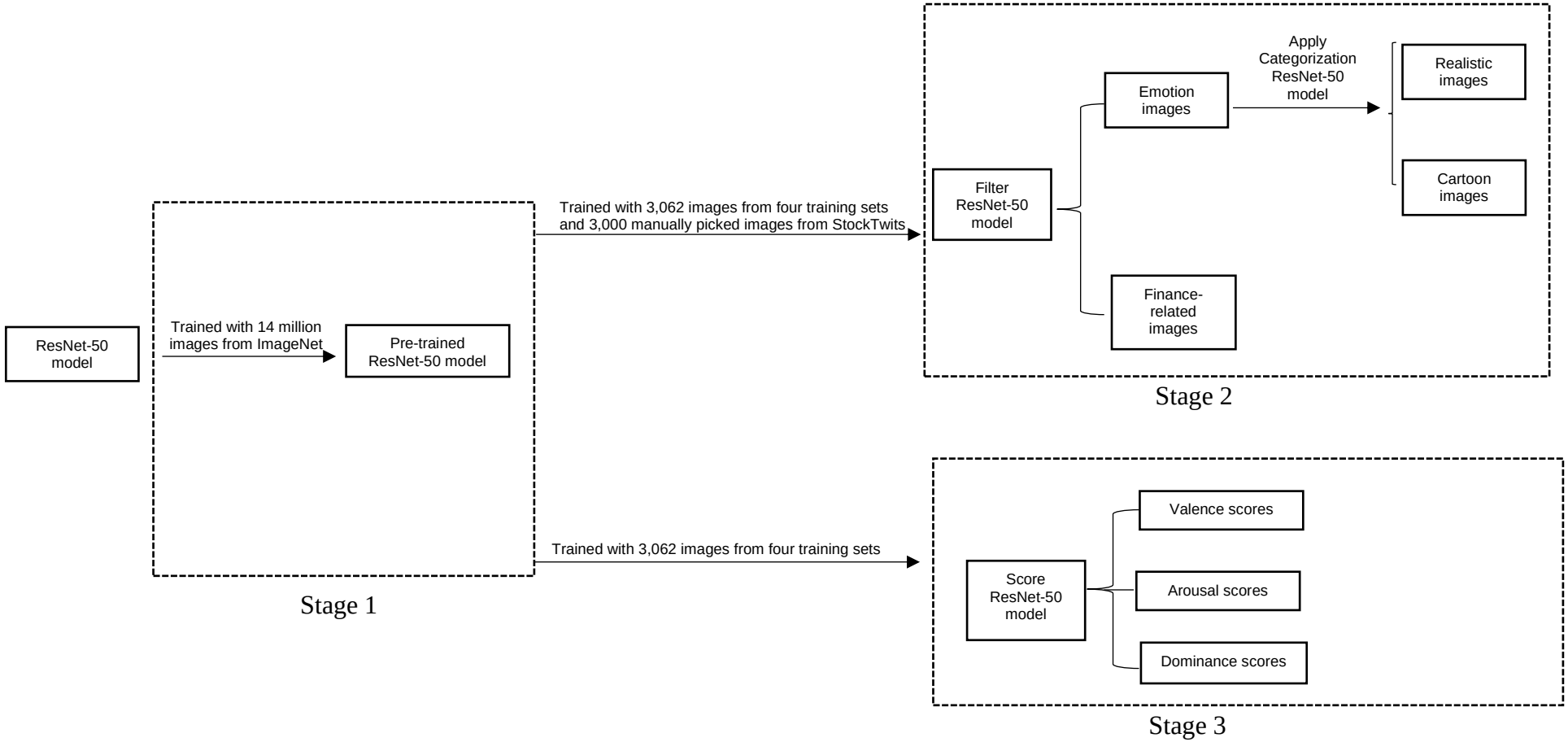


Figure 2. Selected samples of images from the training image databases

The figure shows selected image samples that we apply to train the ResNet-50 model in three stages. In the first stage, we use the ImageNet database, containing over 14 million images, to pretrain the ResNet-50 model. Panel A shows a sample from the ImageNet database. We further train the ResNet-50 model with four image databases to be the Filter and Categorization ResNet-50 models in the second stage and the Score ResNet-50 model in the third stage. The four image databases include 1,192 images from the International Affective Picture System, 730 images from the Geneva Affective Picture Database, 240 images from the Military Affective Picture System, and 900 images from the Open Affective Standardized Image Set. Panel B shows a selected sample of images from the Geneva Affective Picture Database and the Open Affective Standardized Image Set. The International Affective Picture System and the Military Affective Picture System are not open to the public.

Panel A: Selected sample from ImageNet



Panel B: Selected sample from GAPED and OASIS

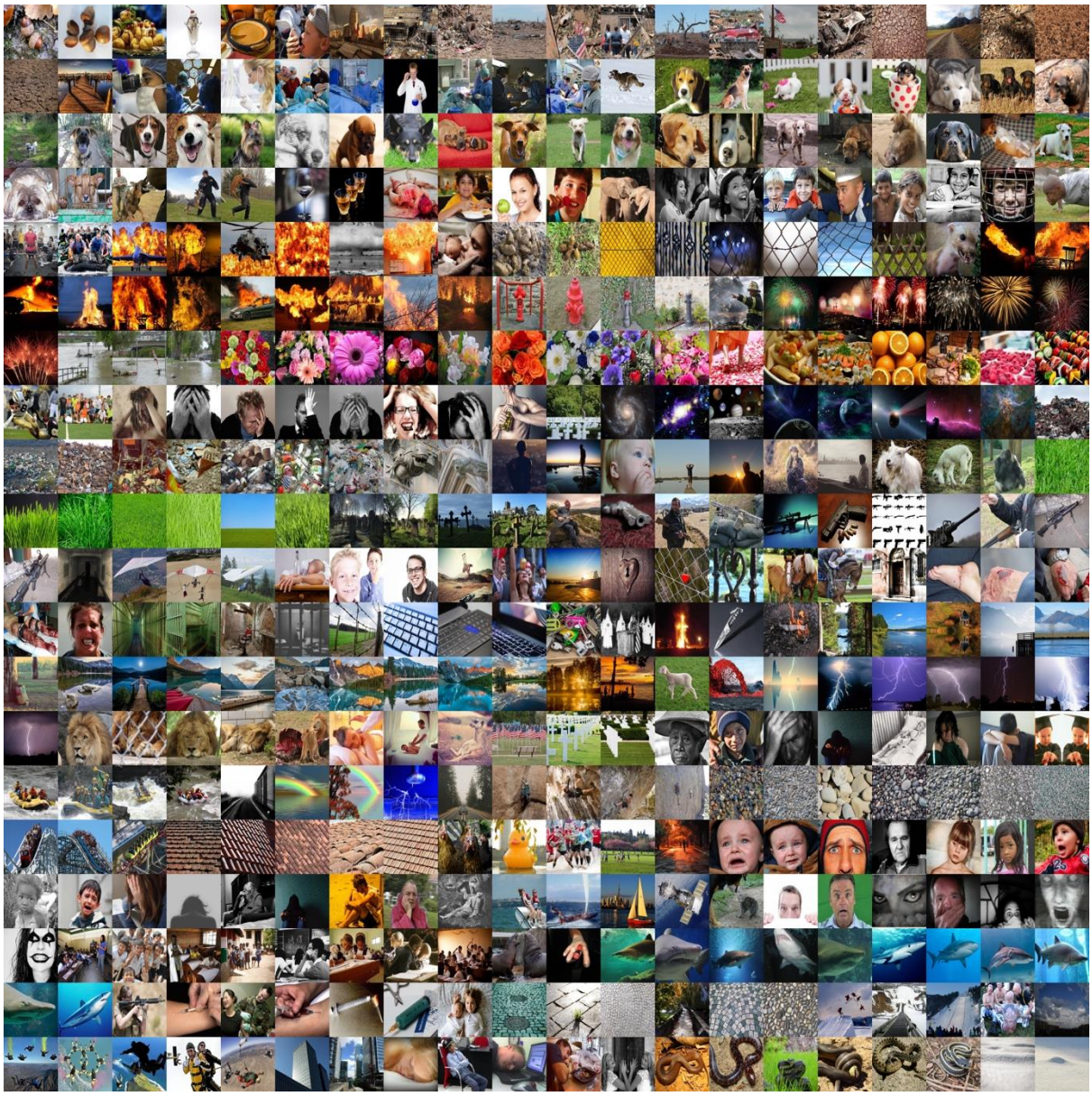


Figure 3. Image types

This figure displays examples of finance-related, realistic, and cartographic images selected from the output of the ResNet-50 model trained in the second stage for filtering images. All images are presented in static format. The static format of the GIF images is in the middle frame.

Panel A: Examples of images containing financial information



Panel B: Examples of realistic images



Panel C: Examples of cartoon images



Figure 4. Selected sample from the final StockTwits database

Figure 4 presents a curated subset of images from the final StockTwits sample, which encompasses 2,046,190 images in PNG, JPG, and GIF formats, spanning the sample period from 1 January 2012 to 31 December 2021.

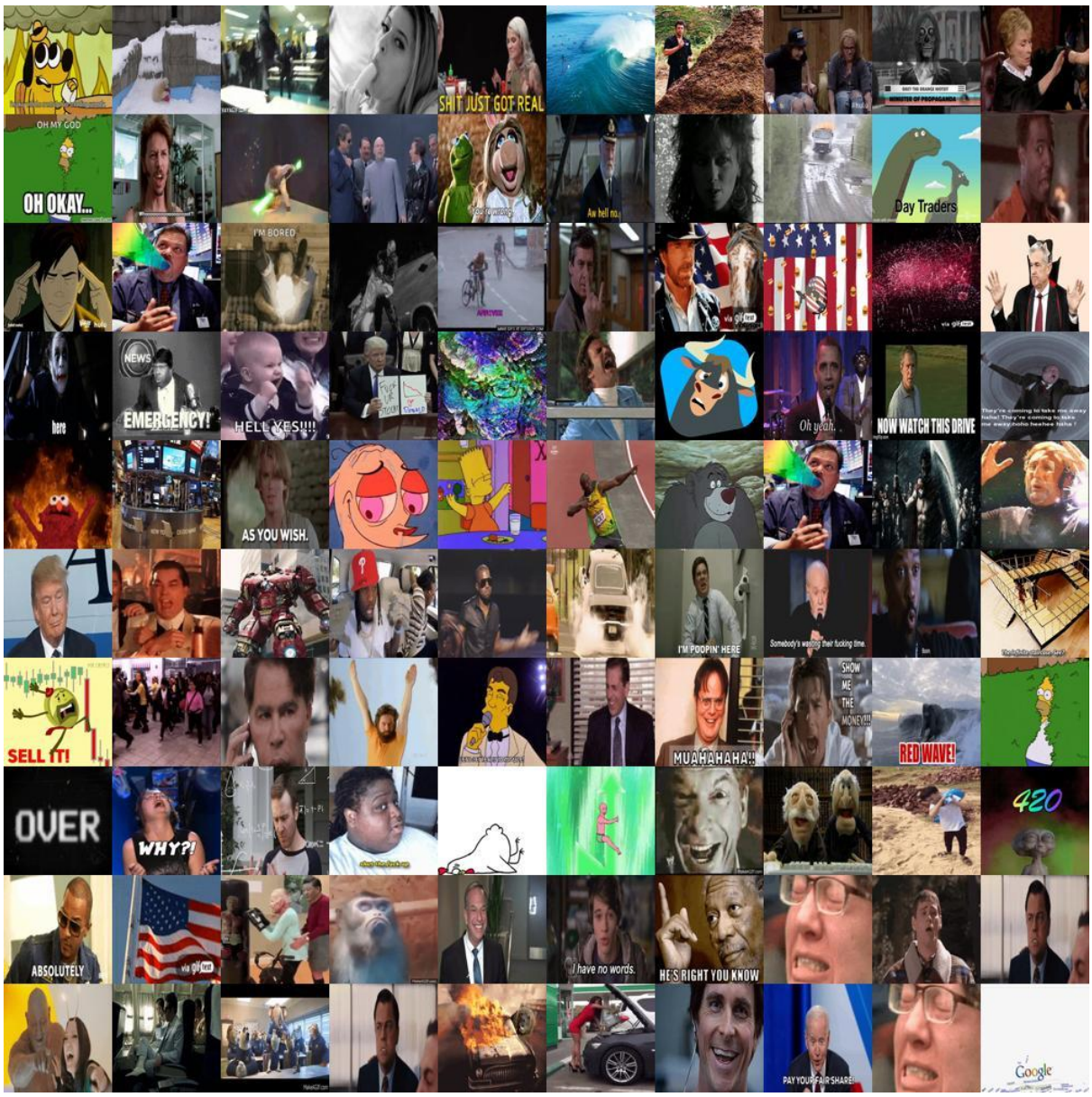
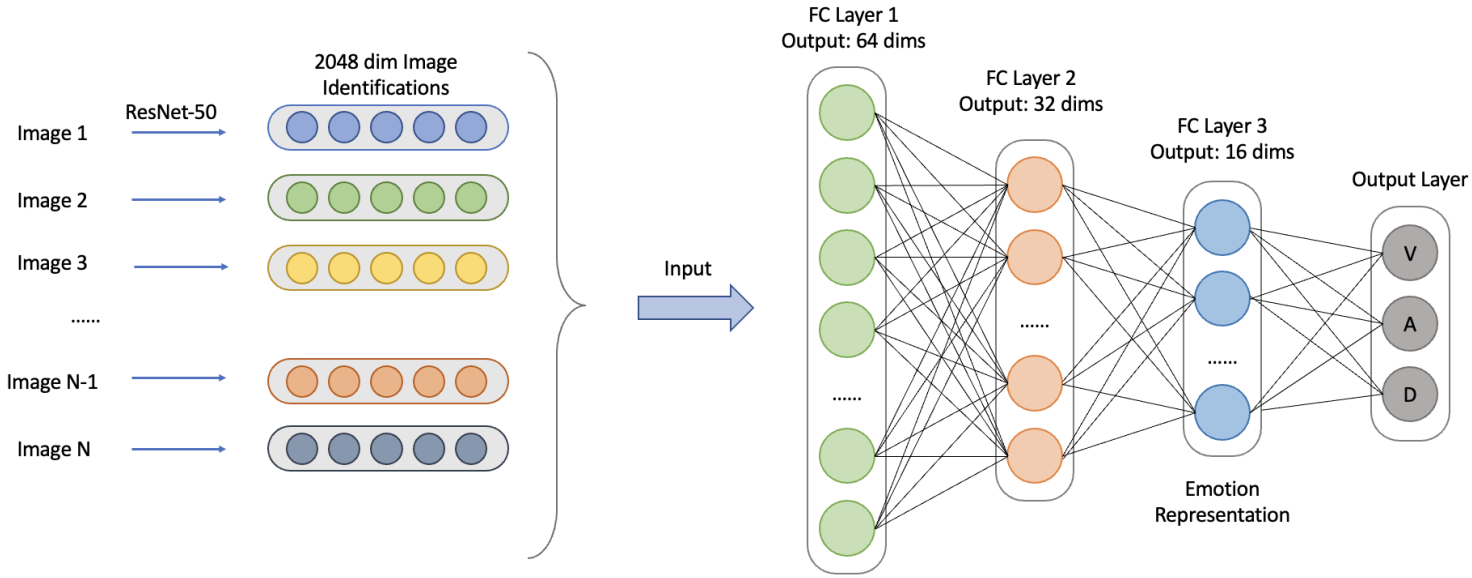


Figure 5. Neural mapping in assigning VAD scores to images

This figure presents the procedure for applying the pre-trained ResNet-50 model and the Neural Space Mapping Method to assign valence, arousal, and dominance scores to each image in the StockTwits sample.



Stage 1: Pre-trained ResNet-50 model

Stage 2: Neural Mapping

Figure 6. The MAE loss for measuring the performance of the ResNet-50 model

This figure shows the performance of the ResNet-50 model in assigning the valence, arousal, and dominance scores. We randomly select 70% of the images in the StockTwits sample as the training set and the remaining 30% as the validation set. The figure shows the Mean Absolute Error (MAE) loss over 30 training epochs in the third stage of the machine-learning process illustrated in Figures 1 and 5. The dark green line indicates the performance of the training set, and the light green line indicates the performance of the validation set. The model performance is measured by the MAE loss, which calculates the average absolute difference between the paired observations as $MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n}$, where y_i and x_i, respectively, are the prediction and the true values of observations, and n is the sample size.

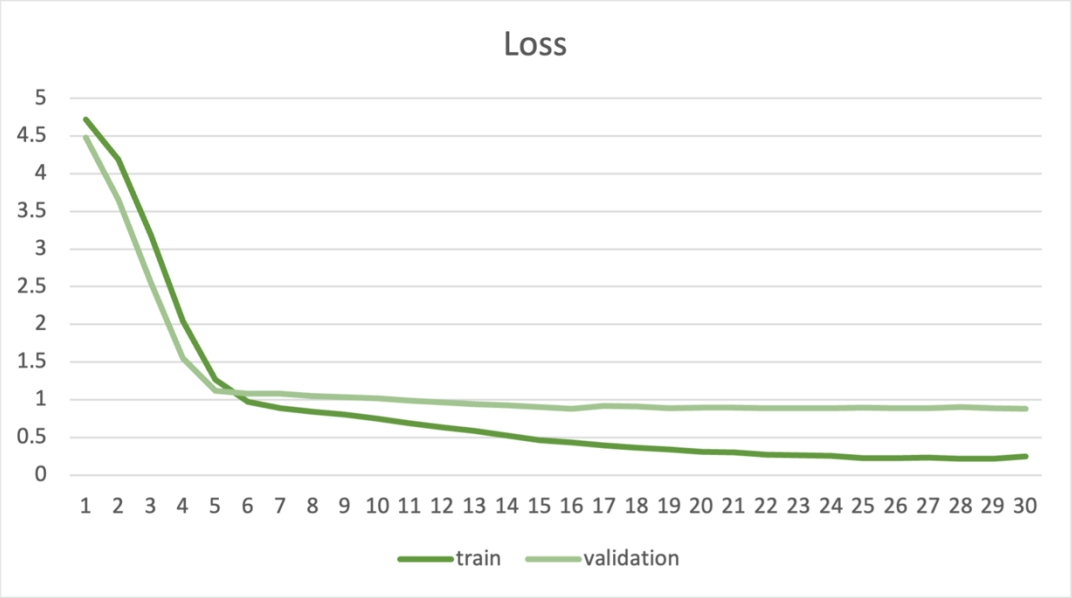


Figure 7. Examples of StockTwits Post Images with High versus Low Valence, Arousal, and Dominance

High Valence



Low Valence



High Arousal



Low Arousal

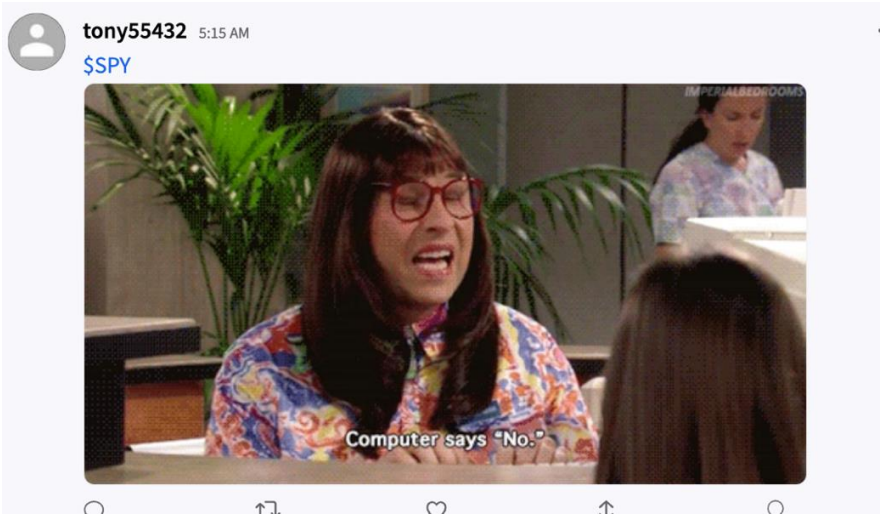


Figure 7 continues from above

High Dominance



Low Dominance

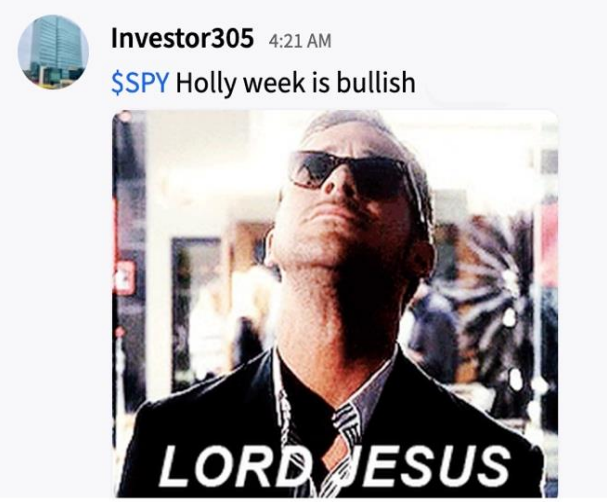


Figure 8. An example of a StockTwits post

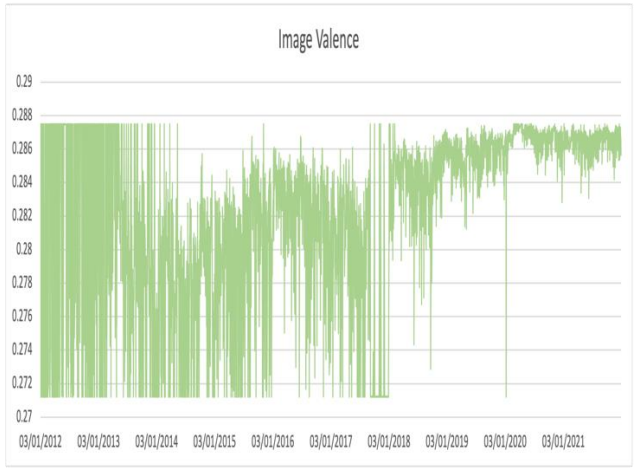
This figure shows an example of a Stocktwit post.

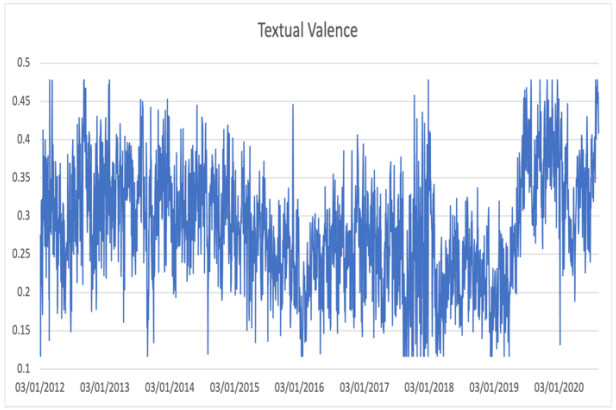


Figure 9. The time series of daily valence, arousal, and dominance indexes

Panels A, B, and C displaythe time series of the daily imagery valence, arousal, and dominance imagery indexes, respectively, (upper frames) and textual indexes (lower frames) between 1 January 2012 and 31 December 2021.

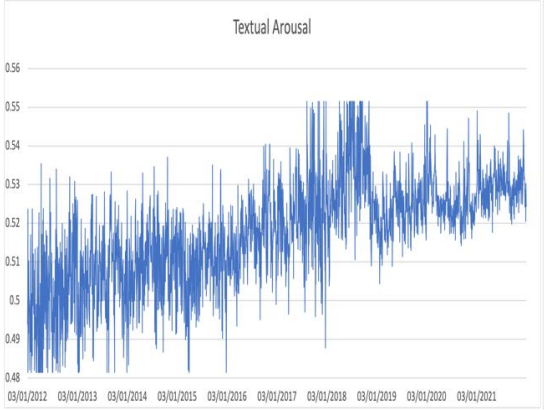
Panel A: Daily valence indexes





Panel B: Daily arousal indexes





Panel C: Daily dominance indexes



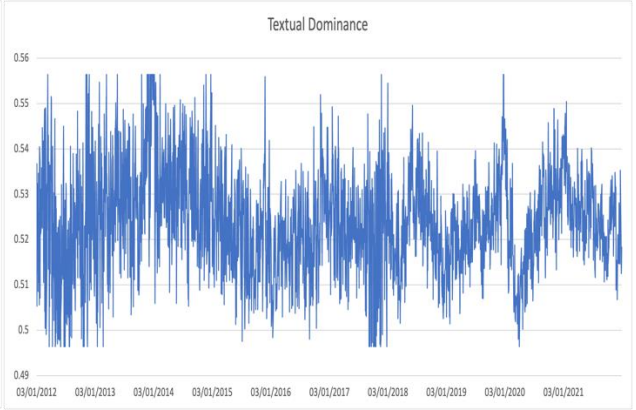


Table 1. Mapped words

Table 1 shows the valence, arousal and dominance values of the words and emojis in the StockTwits post example shown in Figure 8 by mapping to the LM-Renault-augmented word and emoji Valence-Arousal-Dominance (VAD) lexicon. The average values of arousal ($\bar{A}$) and dominance ($\bar{D}$) across all words in the LM-Renault-augmented word and emoji VAD lexicon are presented in the bottom row.

	Valence	Arousal	Dominance
buying	0.665	0.784	0.589
calls	0.608	0.431	0.551
go	0.512	0.441	0.444
down	0.208	0.332	0.264
	0.623	0.409	0.474
	0.145	0.668	0.216
$\bar{V}$ ($\bar{A}$ or $\bar{D}$)	0.501	0.495	0.489

Table 2. Descriptive statistics of the daily emotion dimension indexes

Panel A presents the descriptive statistics of the daily valence, arousal, and dominance indexes, calculated in Equations (1) and (2), based on all imagery content. Panel B reports the descriptive statistics of the daily arousal and dominance indexes calculated using Equations (3) and (4) based on textual content. Panel C shows the correlations between the three imagery-emotion dimension indexes and three textual-emotion dimension indexes. Panel D reports the descriptive statistics of imagery emotion dimension indexes calculated in Equations (1) and (2) based on realistic images. The sample period is over 2,512 trading days from 1 January 2012 to 31 December 2021.

Panel A: Descriptive statistics of image valence, arousal, and dominance

Variable	N	Min	Mean	Max	Std Dev
$ImgValence$	2517	0.087	0.287	0.549	0.011
$ImgArousal$	2517	0.103	0.574	0.577	0.011
$ImgDominance$	2517	0.273	0.515	0.524	0.014

Panel B: Descriptive statistics of text valence, arousal, and dominance

Variable	N	Min	Mean	Max	Std Dev
$TextValence$	2517	0.217	0.393	0.578	0.080
$TextArousal$	2517	0.458	0.518	0.571	0.014
$TextDominance$	2517	0.476	0.524	0.582	0.012

Panel C: Correlations between emotion dimension imagery indexes and textual indexes

	$I_Valence$	$I_Arousal$	$I_Dominance$	$T_Valence$	$T_Arousal$	$T_Dominance$
$I_Valence$	1					
$I_Arousal$	-0.1451	1				
$I_Dominance$	0.3371	0.3493	1			
$T_Valence$	0.0242	-0.0362	0.0967	1		
$T_Arousal$	0.0621	0.079	0.1223	-0.0977	1	
$T_Dominance$	0.009	-0.0656	-0.0028	0.5116	0.073	1

Panel D: Descriptive statistics of realistic images

Variable	N	Min	Mean	Max	Std Dev
$RealValence$	2517	0.185	0.391	0.392	0.008
$RealArousal$	2517	0.184	0.619	0.929	0.014
$RealDominance$	2517	0.370	0.592	0.593	0.008

Table 3. The influence of imagery emotion dimension indexes on stock market returns: all images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the following model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the following model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the following model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index, $ImgArousal_t$ is the daily imagery arousal, and $ImgDominance_t$ is the daily imagery dominance index. All imagery indexes were constructed based on all the images. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, and January effect dummy. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery valence index			Panel B: Predicting S&P 500 returns with daily imagery arousal index			Panel C: Predicting S&P 500 returns with daily imagery dominance index		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	2.37***	(3.20)	$ImgArousal_{t-1}$	1.85*	(1.80)	$ImgDominance_{t-1}$	3.37*	(1.92)
$ImgValence_{t-2}$	2.04	(1.48)	$ImgArousal_{t-2}$	0.39	(0.24)	$ImgDominance_{t-2}$	3.74**	(2.37)
$ImgValence_{t-3}$	1.71	(1.43)	$ImgArousal_{t-3}$	2.73*	(1.68)	$ImgDominance_{t-3}$	3.73**	(2.25)
$ImgValence_{t-4}$	1.71	(0.85)	$ImgArousal_{t-4}$	0.28	(0.23)	$ImgDominance_{t-4}$	-1.60	(-0.87)
$ImgValence_{t-5}$	3.17**	(2.05)	$ImgArousal_{t-5}$	2.93*	(1.71)	$ImgDominance_{t-5}$	1.59	(0.84)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	3.04***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	3.47***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	4.76***	
p-value	0.0097		*p-value*	0.0040		*p-value*	0.0002	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	1.71		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	2.35*		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	3.97***	
p-value	0.1443		*p-value*	0.0525		*p-value*	0.0033	
$adj.R^2$	0.0924		$adj.R^2$	0.0895		$adj.R^2$	0.0900	

Table 4. The influence of imagery and textual emotion dimension indexes on stock market returns: all images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(TextValence_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(TextArousal_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(TextDominance_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes, as well as relevant textual indexes, on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index, $ImgArousal_t$ is the daily imagery arousal, and $ImgDominance_t$ is the daily imagery dominance index. All imagery indexes were constructed based on all the images. $TextValence_t$ is a daily textual valence index. $TextArousal_t$ is the daily textual arousal index. $TextDominance_t$ is a daily textual dominance index. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, and January effect dummy. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery and textual valence indexes			Panel B: Predicting S&P 500 returns with daily imagery and textual arousal indexes			Panel C: Predicting S&P 500 returns with daily imagery and textual dominance indexes		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	2.52***	(3.43)	$ImgArousal_{t-1}$	2.18*	(1.83)	$ImgDominance_{t-1}$	3.66**	(1.98)
$ImgValence_{t-2}$	1.69	(1.17)	$ImgArousal_{t-2}$	0.47	(0.23)	$ImgDominance_{t-2}$	3.95**	(2.48)
$ImgValence_{t-3}$	1.74	(1.44)	$ImgArousal_{t-3}$	3.85*	(1.74)	$ImgDominance_{t-3}$	3.85**	(2.19)
$ImgValence_{t-4}$	1.96	(1.02)	$ImgArousal_{t-4}$	-0.31	(-0.22)	$ImgDominance_{t-4}$	-2.02	(-1.09)
$ImgValence_{t-5}$	3.58**	(2.27)	$ImgArousal_{t-5}$	3.66*	(1.67)	$ImgDominance_{t-5}$	2.36	(1.25)
$TextValence_{t-1}$	10.51***	(3.25)	$TextArousal_{t-1}$	6.93***	(2.77)	$TextDominance_{t-1}$	6.52***	(2.91)
$TextValence_{t-2}$	-6.02*	(-1.86)	$TextArousal_{t-2}$	-4.56*	(-1.78)	$TextDominance_{t-2}$	-8.87***	(-3.47)
$TextValence_{t-3}$	-0.67	(-0.23)	$TextArousal_{t-3}$	-1.13	(-0.45)	$TextDominance_{t-3}$	-5.62**	(-2.46)
$TextValence_{t-4}$	-1.41	(-0.53)	$TextArousal_{t-4}$	3.97	(1.54)	$TextDominance_{t-4}$	0.05	(0.03)
$TextValence_{t-5}$	-0.16	(-0.06)	$TextArousal_{t-5}$	-1.45	(-0.67)	$TextDominance_{t-5}$	-3.31	(-1.55)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	3.37***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	3.76**		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	6.03***	
p-value	0.0049		*p-value*	0.0022		*p-value*	0.0000	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	1.71		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	2.64**		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	4.86***	
p-value	0.1453		*p-value*	0.0322		*p-value*	0.0007	
$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextValence=0]$	2.55**		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextArousal=0]$	1.91*		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextDominance=0]$	4.07***	
p-value	0.0263		*p-value*	0.0897		*p-value*	0.0011	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextValence=0]$	1.10		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextArousal=0]$	1.28		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextDominance=0]$	4.75***	
p-value	0.3560		*p-value*	0.2741		*p-value*	0.0008	
$adj.R^2$	0.0953		$adj.R^2$	0.0910		$adj.R^2$	0.0995	

Table 5. The influence of imagery emotion dimension indexes on stock market returns: realistic images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index $ImgArousal_t$ is the daily imagery arousal index. $ImgDominance_t$ is a daily imagery dominance index. All imagery indexes were constructed based on realistic images only. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, and January effect dummy. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery valence index			Panel B: Predicting S&P 500 returns with daily imagery arousal index			Panel C: Predicting S&P 500 returns with daily imagery dominance index		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	1.98	(1.47)	$ImgArousal_{t-1}$	1.38**	(2.25)	$ImgDominance_{t-1}$	1.78	(1.47)
$ImgValence_{t-2}$	2.45	(1.61)	$ImgArousal_{t-2}$	0.20	(0.12)	$ImgDominance_{t-2}$	2.15	(1.54)
$ImgValence_{t-3}$	3.08**	(2.44)	$ImgArousal_{t-3}$	2.90***	(3.11)	$ImgDominance_{t-3}$	2.85***	(2.56)
$ImgValence_{t-4}$	-3.70***	(-2.64)	$ImgArousal_{t-4}$	4.45***	(4.99)	$ImgDominance_{t-4}$	-3.46***	(-2.80)
$ImgValence_{t-5}$	1.13	(0.69)	$ImgArousal_{t-5}$	3.01**	(2.04)	$ImgDominance_{t-5}$	1.10	(0.75)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	3.90***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	8.90***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	3.88***	
p-value	0.0016		*p-value*	0.0000		*p-value*	0.0017	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	3.70**		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	11.01***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	3.82***	
p-value	0.0052		*p-value*	0.0000		*p-value*	0.0042	
$adj.R^2$	0.0894		$adj.R^2$	0.0892		$adj.R^2$	0.0894	

Table 6. The influence of imagery and textual emotion dimension indexes on stock market returns: realistic images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the =model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(TextValence_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(TextArousal_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the =model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(TextDominance_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes, as well as relevant textual indexes, on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index $ImgArousal_t$ is the daily imagery arousal index. $ImgDominance_t$ is a daily imagery dominance index. All imagery indexes were constructed based on realistic images only. $TextValence_t$ is a daily textual valence index. $TextArousal_t$ is the daily textual arousal index. $TextDominance_t$ is a daily textual dominance index. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, and January effect dummy. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery and textual valence indexes			Panel B: Predicting S&P 500 returns with daily imagery and textual arousal indexes			Panel C: Predicting S&P 500 returns with daily imagery and textual dominance indexes		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	1.04	(0.65)	$ImgArousal_{t-1}$	1.32**	(2.24)	$ImgDominance_{t-1}$	2.90	(1.05)
$ImgValence_{t-2}$	2.96**	(2.11)	$ImgArousal_{t-2}$	-0.22	(-0.13)	$ImgDominance_{t-2}$	4.17*	(1.79)
$ImgValence_{t-3}$	3.91**	(2.47)	$ImgArousal_{t-3}$	2.98***	(3.38)	$ImgDominance_{t-3}$	6.56***	(2.85)
$ImgValence_{t-4}$	-4.17***	(-2.87)	$ImgArousal_{t-4}$	4.38***	(5.02)	$ImgDominance_{t-4}$	-6.74***	(-2.82)
$ImgValence_{t-5}$	1.11	(0.69)	$ImgArousal_{t-5}$	2.91***	(2.05)	$ImgDominance_{t-5}$	3.02	(1.07)
$TextValence_{t-1}$	10.30***	(3.17)	$TextArousal_{t-1}$	6.92***	(2.80)	$TextDominance_{t-1}$	4.40***	(2.93)
$TextValence_{t-2}$	-6.51**	(-2.00)	$TextArousal_{t-2}$	-4.28*	(-1.67)	$TextDominance_{t-2}$	-5.93***	(-3.47)
$TextValence_{t-3}$	-0.48	(-0.16)	$TextArousal_{t-3}$	-1.75	(-0.71)	$TextDominance_{t-3}$	-3.66**	(-2.39)
$TextValence_{t-4}$	-0.53	(-0.20)	$TextArousal_{t-4}$	3.98	(1.56)	$TextDominance_{t-4}$	0.13	(0.10)
$TextValence_{t-5}$	-0.34	(-0.13)	$TextArousal_{t-5}$	-1.41	(-0.65)	$TextDominance_{t-5}$	-2.21	(-1.55)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	5.02***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	8.68***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	5.39***	
p-value	0.0001		*p-value*	0.0000		*p-value*	0.0001	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	5.29***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	10.77***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	5.60***	
p-value	0.0003		*p-value*	0.0000		*p-value*	0.0002	
$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextValence=0]$	2.45**		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextArousal=0]$	1.86*		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextDominance=0]$	3.94***	
p-value	0.0320		*p-value*	0.0979		*p-value*	0.0014	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextValence=0]$	1.10		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextArousal=0]$	1.32		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextDominance=0]$	4.57***	
p-value	0.3557		*p-value*	0.2583		*p-value*	0.0011	
$adj.R^2$	0.0925		$adj.R^2$	0.0910		$adj.R^2$	0.0996	

Table 7. The Out-of-sample test

This table reports R^2_{oos} (out-of-sample R^2) as a percentage and the associated *p*-values using the MSPE-adjusted statistic in Clark and West (2007) for the recursively estimated predictive regression of the S&P 500 index returns on a one-period lag of $ImgValence_t$, $ImgArousal_t$, and $ImgDominance_t$, respectively. Panel A reports the results by applying the full sample of images, and Panel B presents the results using realistic images only. We used data for the initial estimation period between 1 January 2012 and 31 December 2019. The out-of-sample period ranged from 1 January 2020 to 31 December 2021.

Re_{t+1}	R^2_{oos} (%)
Panel A: Full sample	
$ImgValence_t$	1.009
$ImgArousal_t$	0.991
$ImgDominance_t$	0.976
Panel B: Realistic images	
$ImgValence_t$	1.016
$ImgArousal_t$	0.989
$ImgDominance_t$	0.963

Internet Appendices

Internet Appendix 1. The four image training databases

The IAPS is the most widely used database in behavioral and neuroimaging studies for analyzing emotional processing, cognitive neuroscience, and affective disorders. Numerous cross-validation studies have consistently demonstrated the effectiveness of these stimuli in reliably eliciting expressive and physiological emotional responses (Weinberg and Hajcak, 2011; Marchewka et al., 2014). Following subsequent updates by Lang, Bradley, and Cuthbert (2005) based on the original norms, the IAPS includes 1,192 high-quality pictures containing natural pictures designed to evoke emotional responses and encompass a variety of subjects, including people, animals, landscapes, and objects. Moreover, the discrete-category theory of emotions has been employed to partially characterize the dataset (Mikels et al., 2005).

GAPED, introduced by Dan-Glauser and Scherer in 2011, contains 730 distinct pictures from the IAPS and is developed to provide a wider range of visual stimuli for studying emotions, encompassing six distinct categories. Among the negative pictures, the four specific content categories include spiders, snakes, and scenes that elicit emotions associated with violations of moral or legal norms, such as human rights violations or animal mistreatment. Positive pictures predominantly feature human and animal babies as well as natural landscapes, whereas neutral pictures primarily depict inanimate objects. The GAPED database can be especially valuable for phobic reaction studies (Aue, Hoeppli, and Piguet, 2012) or research requiring multiple presentations of stimuli of the same type. Such categorization allows researchers to select images relevant to their specific areas of investigation, enhancing the database's applicability across different research contexts.

MAPS and OASIS are two recently developed picture databases that serve as supplementary resources for IAPS. Goodman, Katz, and Dretsch (2016) develop the MAPS stimuli set, consisting of 240 carefully selected images intended to evoke a range of commonly

experienced emotional responses, including stress, fear, pride, and other emotions prevalent in military settings. While military research is the primary focus of MAPS, civilians are also involved in rating the pictures, making it an appropriate research instrument for eliciting context-dependent emotional responses in other fields (Goodman, Katz, and Dretsch, 2016; Dolcos et al., 2022). The OASIS database includes 900 images that cover a broad spectrum of themes and is designed to provide a collection of standardized images for evoking and assessing emotional responses across different studies. The aim of OASIS is to ensure consistency and comparability in emotional research by offering a diverse range of visual stimuli that are carefully selected to represent various emotional states and experiences. The images in OASIS underwent rigorous validation processes, including ratings for emotion dimensions, such as valence and arousal (Kurdi, Lozano, and Banaji, 2017). Researchers in the field of affective science can utilize OASIS as a resource for enhancing their studies' consistency and replicability, contributing to a deeper understanding of the complexities of human emotions (Kragel et al., 2019).

Internet Appendix 2. Construction of the LM-Renault augmented word and emoji VAD lexicon

We describe the process of applying machine learning algorithms to augment the LM wordlist and Renault (2017) with additional words in the NRC VAD lexicon and then label emojis with VAD scores.

We utilized two distinct machine learning techniques to handle this task: the GloVe model and the Neural Space Mapping Method (NSMM). Each technique was employed during the different phases of the process. The GloVe model has demonstrated excellent performance in various natural language processing tasks, such as word analogy, word similarity, and named entity recognition. Its effectiveness and versatility make it a powerful tool for various language-processing applications. Additionally, the NSMM is well suited for capturing and modeling complex nInternetar relationships between inputs and outputs. The NSMM can learn and improve over time as it encounters more data, making it highly adaptable. Since we utilized the VAD model, which is a multidimensional emotion model, the NSMM is effective in handling high-dimensional data. This enables effective management of numerous inputs and variables, which is crucial for our analysis. By leveraging the strengths of the GloVe model and the NSMM, we aim to enhance the NRC VAD lexicon by incorporating additional words from the LM Wordlist and Renault (2017). This process enriched the lexicon and improved its coverage of words and their associated valence, arousal, and dominance scores, ultimately enhancing the accuracy and effectiveness of our analyses.

The LM wordlist contained 2,710 words, the Renault wordlist contained 1,055 unigram words, and 453 words were common between the two lists. Among these words, 1,141 were already included in the NRC VAD lexicon, whereas the remaining 2,171 were not. To handle the additional 2,171 words that were not present in the NRC VAD lexicon, we employed a supervised Neural Space Mapping Method (NSMM) in conjunction with the pre-trained GloVe model. This process involves mapping words using fully connected layers within a neural network in two stages. We utilized the supervised GloVe model to generate word embeddings for 2,171 additional words in the first stage. The pre-trained GloVe model, as described by Pennington, Socher, and Manning (2014), demonstrated superior performance in similarity tasks and named entity recognition, efficiently leveraging statistical information and producing a vector space with a meaningful substructure.

We used a dataset comprising three million news articles published in the Wall Street Journal to train the GloVe model. We selected the top 200,000 most frequent words from this

dataset in financial news. These words served as the training dataset for the GloVe model, which constructed word embeddings with 200 dimensions to define the semantic representation for each word. By leveraging the strengths of the pre-trained GloVe model and supervised NSMM, we can effectively map an additional 2,171 words to their corresponding word embeddings. This process enhances the coverage and accuracy of the NRC VAD lexicon, allowing us to assign valence, arousal, and dominance scores to a broader range of words and further improve our analysis of emotion dimensions in the text data.

In the second stage, we constructed an emotion neural convertor by applying a fully connected layer to gradually convert multidimensional word embeddings into emotion representations with fewer dimensions so that we can output valence, arousal, and dominance scores for each word, as each word is represented as a vector. According to Sainath et al. (2015), constructing a neural converter based on a neural network is a novel procedure that does not require troublesome parameter extraction for predicting the next point when mapping words. In addition, Jaderberg et al. (2014) stated that the neural network algorithm does not require any human-labeled data and holistically performs word recognition on the entire image, departing from the character-based recognition systems of the past. Specifically, the emotion neural converter has three hidden layers and one output layer, and all the layers are fully connected (FC layers). For the three hidden layers, we input 200 dimensions into the first C layer and 64 dimensions into the second FC layer. Finally, we output the emotional representations with 16 dimensions in the third fc layer. Among each FC layer, we use the Rectified Linear Unit[10] (ReLU) activation developed by Goodfellow, Bengio, and Courville (2016) to implement the nInternetarity of the neural network. For the output layer, emotion representations with 16 dimensions were separately converted into one-dimensional scalars, valence, arousal, and dominance scores.

Following a similar approach for assigning VAD scores to words, we trained a neural converter to process emotions. We trained an emotional neural converter with three FC layers and one output FC layer, using data procured from Twitter and StockTwits. Specifically, our training database comprised 901,028 posts from StockTwits from January to September 2022 and 691,377 posts from Twitter dated 1 September 2022. It is crucial to note that all the posts incorporate emojis. The first hidden layer accepts 200 dimensions, followed by the second layer, which accepts 64 dimensions—the third layer outputs emotional representations within

[10] The rectified linear activation function (ReLU) is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero.

a 16-dimensional space. We utilized the ReLU activation function as proposed by Goodfellow, Bengio, and Courville (2016) in each FC layer to facilitate nonlinearity within the neural network. Within the output layer, 16-dimensional emotional representations were transmuted into single-dimensional scalars corresponding to separate valence, arousal, and dominance scores. Consequently, we could allocate the VAD scores to each of the 1,114 emojis.

To measure the performance of the emotion neural convertor, we randomly selected 70% of the NRC VAD lexicon as the training set and the remaining 30% as the validation set. On the training set, we apply the Adam optimizer enhanced by Kingma and Ba (2014) and train 200 iterations with a learning rate of 10^{-3}. We then used the Mean Squared Error (MSE), as in Goodfellow, Bengio, and Courville (2016), to measure the training loss, which is the sum of MSE distances among the scores for valence, arousal, and dominance. In the validation set, the MSE value was 0.036, indicating satisfactory performance for the NSMM. After combining additional words with the original NRC VAD lexicon, the final version of the LM-Renault augmented word and emoji VAD lexicon included 23,225 unigram words in total, and each word was assigned a score for valence, arousal, and dominance. Moreover, the distributions of scores for valence, arousal, and dominance are similar between the NRC VAD lexicon and the LM-Renault augmented word and emoji VAD lexicons, providing additional solid evidence for the satisfactory performance of the NSMM.

Internet Appendix References

Aue, T., Hoeppli, M. E., and Piguet, C. (2012). The sensitivity of physiological measures to phobic and nonphobic fear intensity. *Journal of Psychophysiology*, 26, 154–167.

Dolcos, F., Bogdan, P. C., O'Brien, M., Iordan, A. D., Madison, A., Buetti, S., ... and Dolcos, S. (2022). The impact of focused attention on emotional evaluation: An eye-tracking investigation. *Emotion*, 22(5), 1088.

Jaderberg, M., Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Synthetic data and artificial neural networks for natural scene text recognition. *arXiv preprint arXiv:1406.2227*.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (pp. A-8). Gainesville, FL: NIMH, Center for the Study of Emotion & Attention.

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., and Reuter-Lorenz, P. A. (2005). Emotional category data on images from the International Affective Picture System. *Behavior Research Methods, 37*, 626-630.

Marchewka, A., Żurawski, Ł., Jednoróg, K., and Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods, 46*, 596-610.

Sainath, T. N., Vinyals, O., Senior, A., and Sak, H. (2015, April). Convolutional, long short-term memory, fully connected deep neural networks. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASS&P)*, 4580-4584.

Weinberg, A., and Hajcak, G. (2011). The late positive potential predicts subsequent interference with target processing. *Journal of Cognitive Neuroscience*, *23*(10), 2994-3007.

Internet Appendix 2, Table 1. Sample words and emojis from the LM-Renault-augmented word and emoji VAD lexicon

Table 1 presents the selection of words and emojis from the LM-Renault-augmented word and emoji VAD lexicons. The first column lists selected words from the LM and Renault wordlists along with a few emojis. The subsequent columns provide the corresponding values for valence, arousal, and dominance associated with each word and emoji.

Word & emoji	Valence	Arousal	Dominance
protestor	0.1175	0.7459	0.5001
demolishes	0.1027	0.5193	0.5752
opportunities	0.9273	0.7337	0.8203
riskiest	0.1735	0.7139	0.5421
undocumented	0.2071	0.3015	0.3007
burdens	0.3211	0.5100	0.6098
abusively	0.0504	0.6202	0.3332
deviated	0.1287	0.4181	0.2001
win	0.8232	0.6945	0.7597
impressively	0.7194	0.5995	0.7210
😂	0.4695	0.0906	0.1222
🤣	0.3742	0.1290	0.0525
🔥	0.7657	0.6723	0.8238
🤦‍♀️	0.2449	0.3882	0.4129
🐻	0.1764	0.5199	0.6776
💰	0.7053	0.6406	0.5104
📈	0.8062	0.6095	0.5160
🤑	0.2122	0.3875	0.3867
🤡	0.4458	0.4881	0.3231
😭	0.2886	0.7090	0.2993

Internet Appendix 3

Internet Appendix 3, Table 1. Influence of imagery emotion dimension indexes on stock market returns: All images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the following model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index $ImgArousal_t$ is the daily imagery arousal index. $ImgDominance_t$ is a daily imagery dominance index. All imagery indexes were constructed based on all the images. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, and January effect dummy. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery valence index			Panel B: Predicting S&P 500 returns with daily imagery arousal index			Panel C: Predicting S&P 500 returns with daily imagery dominance index		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	3.57**	(2.27)	$ImgArousal_{t-1}$	2.98*	(1.88)	$ImgDominance_{t-1}$	3.85**	(2.41)
$ImgValence_{t-2}$	1.97	(1.21)	$ImgArousal_{t-2}$	1.65	(0.80)	$ImgDominance_{t-2}$	2.68*	(1.78)
$ImgValence_{t-3}$	4.46***	(2.66)	$ImgArousal_{t-3}$	5.19**	(2.74)	$ImgDominance_{t-3}$	4.22**	(2.58)
$ImgValence_{t-4}$	1.16	(0.63)	$ImgArousal_{t-4}$	-0.70	(-0.34)	$ImgDominance_{t-4}$	0.36	(0.19)
$ImgValence_{t-5}$	1.68	(0.99)	$ImgArousal_{t-5}$	1.95	(1.13)	$ImgDominance_{t-5}$	1.58	(0.96)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	3.33***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	3.76***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	3.67***	
p-value	0.0053		*p-value*	0.0022		*p-value*	0.0026	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	2.64**		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	2.49**		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	2.86**	
p-value	0.0323		*p-value*	0.0413		*p-value*	0.0223	
$adj.R^2$	0.0948		$adj.R^2$	0.0946		$adj.R^2$	0.0951	

Internet Appendix 3, Table 2. Influence of imagery and textual emotion dimension indexes on stock market returns: All images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(TextValence_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(TextArousal_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(TextDominance_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes, as well as relevant textual indexes, on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index $ImgArousal_t$ is the daily imagery arousal index. $ImgDominance_t$ is a daily imagery dominance index. All imagery indexes are constructed based on all the images. $TextValence_t$ is a daily textual valence index. $TextArousal_t$ is the daily textual arousal index. $TextDominance_t$ is a daily textual dominance index. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, and January effect dummy. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery and textual valence indexes			Panel B: Predicting S&P 500 returns with daily imagery and textual arousal indexes			Panel C: Predicting S&P 500 returns with daily imagery and textual dominance indexes		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	4.32**	(2.57)	$ImgArousal_{t-1}$	2.80*	(1.80)	$ImgDominance_{t-1}$	4.30**	(2.45)
$ImgValence_{t-2}$	1.27	(0.76)	$ImgArousal_{t-2}$	1.61	(0.80)	$ImgDominance_{t-2}$	1.50	(0.98)
$ImgValence_{t-3}$	4.78***	(2.80)	$ImgArousal_{t-3}$	5.44***	(2.85)	$ImgDominance_{t-3}$	5.68***	(3.21)
$ImgValence_{t-4}$	0.57	(0.30)	$ImgArousal_{t-4}$	-1.24	(-0.59)	$ImgDominance_{t-4}$	0.01	(0.01)
$ImgValence_{t-5}$	1.86	(1.12)	$ImgArousal_{t-5}$	2.01	(1.15)	$ImgDominance_{t-5}$	1.86	(1.11)
$TextValence_{t-1}$	10.36***	(3.15)	$TextArousal_{t-1}$	6.37***	(2.57)	$TextDominance_{t-1}$	6.96***	(3.03)
$TextValence_{t-2}$	-6.20*	(-1.88)	$TextArousal_{t-2}$	-4.60*	(-1.83)	$TextDominance_{t-2}$	-9.13***	(-3.51)
$TextValence_{t-3}$	0.20	(0.07)	$TextArousal_{t-3}$	-1.93	(-0.78)	$TextDominance_{t-3}$	-5.10	(-2.19)
$TextValence_{t-4}$	-1.26	(-0.46)	$TextArousal_{t-4}$	4.01	(1.58)	$TextDominance_{t-4}$	-0.04	(-0.02)
$TextValence_{t-5}$	-0.20	(-0.07)	$TextArousal_{t-5}$	-1.77	(-0.82)	$TextDominance_{t-5}$	-3.46	(-1.60)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	3.79***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	3.88***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	4.62***	
p-value	0.0020		*p-value*	0.0016		*p-value*	0.0003	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	2.81**		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	2.65***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	3.90***	
p-value	0.0241		*p-value*	0.0318		*p-value*	0.0037	
$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextValence=0]$	2.52**		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextArousal=0]$	2.00*		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextDominance=0]$	4.07***	
p-value	0.0278		*p-value*	0.0763		*p-value*	0.0011	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextValence=0]$	0.99		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextArousal=0]$	1.48		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextDominance=0]$	4.71***	
p-value	0.4105		*p-value*	0.2066		*p-value*	0.0009	
$adj.R^2$	0.0979		$adj.R^2$	0.0894		$adj.R^2$	0.1054	

Internet Appendix 3, Table 3. Influence of imagery emotion dimension indexes on stock market returns: Realistic images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index $ImgArousal_t$ is the daily imagery arousal index. $ImgDominance_t$ is a daily imagery dominance index. All imagery indexes were constructed based on realistic images only. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, and January effect dummy. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery valence index			Panel B: Predicting S&P 500 returns with daily imagery arousal index			Panel C: Predicting S&P 500 returns with daily imagery dominance index		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	0.87	(0.62)	$ImgArousal_{t-1}$	1.50**	(2.26)	$ImgDominance_{t-1}$	0.68	(0.49)
$ImgValence_{t-2}$	3.12**	(2.14)	$ImgArousal_{t-2}$	0.68	(0.39)	$ImgDominance_{t-2}$	3.07**	(2.12)
$ImgValence_{t-3}$	2.29*	(1.76)	$ImgArousal_{t-3}$	2.88***	(2.90)	$ImgDominance_{t-3}$	2.16*	(1.74)
$ImgValence_{t-4}$	-4.11**	(-2.40)	$ImgArousal_{t-4}$	4.61***	(4.75)	$ImgDominance_{t-4}$	-3.92*	(-2.43)
$ImgValence_{t-5}$	2.80	(1.63)	$ImgArousal_{t-5}$	2.52	(1.56)	$ImgDominance_{t-5}$	2.92	(1.67)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	4.74***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	5.95***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	5.11***	
p-value	0.0003		*p-value*	0.0000		*p-value*	0.0001	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	4.22***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	7.42***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	4.63***	
p-value	0.0021		*p-value*	0.0000		*p-value*	0.0010	
$adj.R^2$	0.0895		$adj.R^2$	0.0912		$adj.R^2$	0.0893	

Internet Appendix 3, Table 4. Influence of imagery and textual emotion dimension indexes on stock market returns: realistic images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(TextValence_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(TextArousal_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(TextDominance_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes as well as relevant textual indexes, on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index $ImgArousal_t$ is the daily imagery arousal index. $ImgDominance_t$ is a daily imagery dominance index. All imagery indexes were constructed based on realistic images only. $TextValence_t$ is a daily textual valence index. $TextArousal_t$ is the daily textual arousal index. $TextDominance_t$ is a daily textual dominance index. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, and January effect dummy. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery and textual valence indexes			Panel B: Predicting S&P 500 returns with daily imagery and textual arousal indexes			Panel C: Predicting S&P 500 returns with daily imagery and textual dominance indexes		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	-0.30	(-0.20)	$ImgArousal_{t-1}$	1.41**	(2.28)	$ImgDominance_{t-1}$	0.40	(0.26)
$ImgValence_{t-2}$	4.01***	(2.80)	$ImgArousal_{t-2}$	0.26	(0.15)	$ImgDominance_{t-2}$	4.04***	(2.81)
$ImgValence_{t-3}$	3.09*	(1.81)	$ImgArousal_{t-3}$	2.88***	(3.05)	$ImgDominance_{t-3}$	3.12**	(2.03)
$ImgValence_{t-4}$	-4.84***	(-2.71)	$ImgArousal_{t-4}$	4.57***	(4.82)	$ImgDominance_{t-4}$	-4.31**	(-2.29)
$ImgValence_{t-5}$	2.90	(1.62)	$ImgArousal_{t-5}$	2.42	(1.55)	$ImgDominance_{t-5}$	2.89	(1.52)
$TextValence_{t-1}$	10.45***	(3.19)	$TextArousal_{t-1}$	6.80***	(2.76)	$TextDominance_{t-1}$	6.61***	(2.92)
$TextValence_{t-2}$	-6.63**	(-2.02)	$TextArousal_{t-2}$	-4.28*	(-1.68)	$TextDominance_{t-2}$	-8.85***	(-3.43)
$TextValence_{t-3}$	-0.55	(-0.18)	$TextArousal_{t-3}$	-1.80	(-0.73)	$TextDominance_{t-3}$	-5.71**	(-2.49)
$TextValence_{t-4}$	-0.56	(-0.21)	$TextArousal_{t-4}$	4.02	(1.58)	$TextDominance_{t-4}$	0.14	(0.07)
$TextValence_{t-5}$	-0.35	(-0.13)	$TextArousal_{t-5}$	-1.41	(-0.65)	$TextDominance_{t-5}$	-3.24	(-1.50)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	5.54***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	6.07***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	5.84***	
p-value	0.0000		*p-value*	0.0000		*p-value*	0.0000	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	5.43***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	7.49***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	5.73 ***	
p-value	0.0002		*p-value*	0.0000		*p-value*	0.0001	
$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextValence=0]$	2.41**		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextArousal=0]$	2.32**		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextDominance=0]$	4.06***	
p-value	0.0343		*p-value*	0.0408		*p-value*	0.0011	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextValence=0]$	1.12		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextArousal=0]$	1.21		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextDominance=0]$	4.74***	
p-value	0.3446		*p-value*	0.3062		*p-value*	0.0008	
$adj.R^2$	0.0926		$adj.R^2$	0.0971		$adj.R^2$	0.1036	

Internet Appendix 4.

Internet Appendix 4, Table 1. The influence of imagery emotion dimension indexes on stock market returns: all images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the following model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the following model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the following model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index, $ImgArousal_t$ is the daily imagery arousal, and $ImgDominance_t$ is the daily imagery dominance index. All imagery indexes were constructed based on all the images. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, January effect dummy, and Fama-French five factors. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery valence index			Panel B: Predicting S&P 500 returns with daily imagery arousal index			Panel C: Predicting S&P 500 returns with daily imagery dominance index		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	1.65**	(1.96)	$ImgArousal_{t-1}$	1.81*	(1.70)	$ImgDominance_{t-1}$	2.62*	(1.66)
$ImgValence_{t-2}$	1.82*	(1.76)	$ImgArousal_{t-2}$	0.80	(0.57)	$ImgDominance_{t-2}$	3.31**	(2.33)
$ImgValence_{t-3}$	1.43	(1.11)	$ImgArousal_{t-3}$	2.31*	(1.65)	$ImgDominance_{t-3}$	4.04**	(2.53)
$ImgValence_{t-4}$	2.34	(1.18)	$ImgArousal_{t-4}$	-0.71	(-0.68)	$ImgDominance_{t-4}$	-1.27	(-0.69)
$ImgValence_{t-5}$	2.62*	(1.85)	$ImgArousal_{t-5}$	2.19	(1.54)	$ImgDominance_{t-5}$	1.32	(0.71)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	2.10*		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	3.11***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	4.09***	
p-value	0.0628		*p-value*	0.0085		*p-value*	0.0011	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	1.78		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	2.95**		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	3.77***	
p-value	0.1303		*p-value*	0.0192		*p-value*	0.0046	
$adj.R^2$	0.1686		$adj.R^2$	0.1680		$adj.R^2$	0.1720	

Internet Appendix 4, Table 2. The influence of imagery and textual emotion dimension indexes on stock market returns: all images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(TextValence_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(TextArousal_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(TextDominance_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes, as well as relevant textual indexes, on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index, $ImgArousal_t$ is the daily imagery arousal, and $ImgDominance_t$ is the daily imagery dominance index. All imagery indexes were constructed based on all the images. $TextValence_t$ is a daily textual valence index. $TextArousal_t$ is the daily textual arousal index. $TextDominance_t$ is a daily textual dominance index. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, January effect dummy, and Fama-French five factors. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery and textual valence indexes			Panel B: Predicting S&P 500 returns with daily imagery and textual arousal indexes			Panel C: Predicting S&P 500 returns with daily imagery and textual dominance indexes		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	1.78**	(2.26)	$ImgArousal_{t-1}$	1.80*	(1.74)	$ImgDominance_{t-1}$	3.00*	(1.76)
$ImgValence_{t-2}$	1.48	(1.38)	$ImgArousal_{t-2}$	0.84	(0.60)	$ImgDominance_{t-2}$	3.65**	(2.45)
$ImgValence_{t-3}$	1.43	(1.10)	$ImgArousal_{t-3}$	2.65*	(1.71)	$ImgDominance_{t-3}$	4.16**	(2.55)
$ImgValence_{t-4}$	2.62	(1.38)	$ImgArousal_{t-4}$	-1.01	(-0.97)	$ImgDominance_{t-4}$	-1.68	(-0.89)
$ImgValence_{t-5}$	2.99**	(2.07)	$ImgArousal_{t-5}$	2.17	(1.43)	$ImgDominance_{t-5}$	2.03	(1.07)
$TextValence_{t-1}$	9.97***	(3.35)	$TextArousal_{t-1}$	6.23**	(2.53)	$TextDominance_{t-1}$	5.84***	(2.83)
$TextValence_{t-2}$	-5.42*	(-1.74)	$TextArousal_{t-2}$	-3.22	(-1.39)	$TextDominance_{t-2}$	-7.00***	(-3.00)
$TextValence_{t-3}$	-1.93	(-0.67)	$TextArousal_{t-3}$	-1.38	(-0.59)	$TextDominance_{t-3}$	-6.38***	(-2.87)
$TextValence_{t-4}$	-0.08	(-0.03)	$TextArousal_{t-4}$	3.49	(1.42)	$TextDominance_{t-4}$	0.42	(0.24)
$TextValence_{t-5}$	-0.58	(-0.24)	$TextArousal_{t-5}$	-1.76	(-0.85)	$TextDominance_{t-5}$	-3.60	(-1.76)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	3.37***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	3.49**		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	5.08***	
p-value	0.0049		*p-value*	0.0038		*p-value*	0.00011	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	1.71		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	3.32**		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	4.62***	
p-value	0.1453		*p-value*	0.0101		*p-value*	0.0010	
$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextValence=0]$	2.55**		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextArousal=0]$	1.74		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextDominance=0]$	4.01***	
p-value	0.0263		*p-value*	0.1224		*p-value*	0.0012	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextValence=0]$	1.10		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextArousal=0]$	1.08		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextDominance=0]$	4.76***	
p-value	0.3560		*p-value*	0.3672		*p-value*	0.0008	
$adj.R^2$	0.1713		$adj.R^2$	0.1686		$adj.R^2$	0.1809	

Internet Appendix 4, Table 3. The influence of imagery emotion dimension indexes on stock market returns: realistic images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(R_t) + \delta_1 L5(R_t^2) + \lambda_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index $ImgArousal_t$ is the daily imagery arousal index. $ImgDominance_t$ is a daily imagery dominance index. All imagery indexes were constructed based on realistic images only. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, January effect dummy, and Fama-French five factors. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery valence index			Panel B: Predicting S&P 500 returns with daily imagery arousal index			Panel C: Predicting S&P 500 returns with daily imagery dominance index		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	1.67	(1.43)	$ImgArousal_{t-1}$	1.13**	(2.15)	$ImgDominance_{t-1}$	1.72	(1.45)
$ImgValence_{t-2}$	1.82*	(1.82)	$ImgArousal_{t-2}$	-0.02	(-0.02)	$ImgDominance_{t-2}$	1.80*	(1.76)
$ImgValence_{t-3}$	2.52**	(2.19)	$ImgArousal_{t-3}$	2.05**	(2.43)	$ImgDominance_{t-3}$	2.50**	(2.23)
$ImgValence_{t-4}$	-3.01**	(-2.27)	$ImgArousal_{t-4}$	4.50***	(4.71)	$ImgDominance_{t-4}$	-3.15**	(-2.46)
$ImgValence_{t-5}$	0.70	(0.50)	$ImgArousal_{t-5}$	2.53**	(2.08)	$ImgDominance_{t-5}$	0.79	(0.57)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	3.10***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	7.54***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	3.14***	
p-value	0.0086		*p-value*	0.0000		*p-value*	0.0079	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	3.32**		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	8.71***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	3.46***	
p-value	0.0101		*p-value*	0.0000		*p-value*	0.0079	
$adj.R^2$	0.1679		$adj.R^2$	0.1697		$adj.R^2$	0.1679	

Internet Appendix 4, Table 4. The influence of imagery and textual emotion dimension indexes on stock market returns: realistic images

Panel A of this table presents OLS estimates with Newey and West (1987) standard errors of the =model: $R_t = \beta_1 L5(ImgValence_t) + \gamma_1 L5(TextValence_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel B shows OLS estimates of the coefficient β_1 of the model: $R_t = \beta_1 L5(ImgArousal_t) + \gamma_1 L5(TextArousal_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Panel C shows OLS estimates of the coefficient β_1 of the =model: $R_t = \beta_1 L5(ImgDominance_t) + \gamma_1 L5(TextDominance_t) + \delta_1 L5(R_t) + \lambda_1 L5(R_t^2) + \eta_1 X_t + \varepsilon_t$. Each coefficient measures the impact of a one standard deviation increase in the imagery valence, arousal, or dominance indexes, as well as relevant textual indexes, on returns in basis points (one basis point equals a daily return of 0.01%). R_t denotes daily returns on the S&P 500 index. $ImgValence_t$ is the daily imagery valence index $ImgArousal_t$ is the daily imagery arousal index. $ImgDominance_t$ is a daily imagery dominance index. All imagery indexes were constructed based on realistic images only. $TextValence_t$ is a daily textual valence index. $TextArousal_t$ is the daily textual arousal index. $TextDominance_t$ is a daily textual dominance index. X_t denotes a set of control variables, including the daily CBOE volatility index, day-of-the-week dummy, January effect dummy, and Fama-French five factors. All the variables are defined in Sections 4 and 5. The regressions are based on 2,512 observations from 1 January 2012 to 31 December 2021. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Regresseand: S&P 500 index daily returns								
Panel A: Predicting S&P 500 returns with daily imagery and textual valence indexes			Panel B: Predicting S&P 500 returns with daily imagery and textual arousal indexes			Panel C: Predicting S&P 500 returns with daily imagery and textual dominance indexes		
	β_1	*t-stat*		β_1	*t-stat*		β_1	*t-stat*
$ImgValence_{t-1}$	0.77	(0.55)	$ImgArousal_{t-1}$	1.16**	(1.98)	$ImgDominance_{t-1}$	1.68	(1.07)
$ImgValence_{t-2}$	2.26**	(2.18)	$ImgArousal_{t-2}$	-0.33	(-0.23)	$ImgDominance_{t-2}$	2.04*	(1.84)
$ImgValence_{t-3}$	3.38**	(2.53)	$ImgArousal_{t-3}$	2.17**	(2.55)	$ImgDominance_{t-3}$	3.39***	(2.76)
$ImgValence_{t-4}$	-3.50**	(-2.39)	$ImgArousal_{t-4}$	4.52***	(4.67)	$ImgDominance_{t-4}$	-3.50**	(-2.52)
$ImgValence_{t-5}$	0.64	(0.45)	$ImgArousal_{t-5}$	2.54**	(2.16)	$ImgDominance_{t-5}$	1.30	(0.85)
$TextValence_{t-1}$	9.79***	(3.27)	$TextArousal_{t-1}$	6.17**	(2.52)	$TextDominance_{t-1}$	5.94***	(2.87)
$TextValence_{t-2}$	-5.85*	(-1.86)	$TextArousal_{t-2}$	-2.81	(-1.22)	$TextDominance_{t-2}$	-7.00***	(-3.00)
$TextValence_{t-3}$	-1.75	(-0.60)	$TextArousal_{t-3}$	-2.04	(-0.88)	$TextDominance_{t-3}$	-6.21***	(-2.79)
$TextValence_{t-4}$	0.63	(0.24)	$TextArousal_{t-4}$	3.39	(1.39)	$TextDominance_{t-4}$	0.65	(0.37)
$TextValence_{t-5}$	-0.68	-0.27)	$TextArousal_{t-5}$	-1.77	(-0.85)	$TextDominance_{t-5}$	-3.57*	(-1.76)
$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgValence=0]$	4.44***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgArousal=0]$	7.35***		$\chi^2(1)[Sum\ t-1\ to\ t-5\ ImgDominance=0]$	4.60***	
p-value	0.0005		*p-value*	0.0000		*p-value*	0.0004	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgValence=0]$	4.78***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgArousal=0]$	8.45***		$\chi^2(1)[Sum\ t-2\ to\ t-5\ ImgDominance=0]$	5.25***	
p-value	0.0008		*p-value*	0.0000		*p-value*	0.0003	
$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextValence=0]$	2.40**		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextArousal=0]$	1.68		$\chi^2(1)[Sum\ t-1\ to\ t-5\ TextDominance=0]$	3.88***	
p-value	0.0354		*p-value*	0.1361		*p-value*	0.0017	
$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextValence=0]$	1.07		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextArousal=0]$	1.10		$\chi^2(1)[Sum\ t-2\ to\ t-5\ TextDominance=0]$	4.56***	
p-value	0.3704		*p-value*	0.3548		*p-value*	0.0011	
$adj.R^2$	0.1705		$adj.R^2$	0.1702		$adj.R^2$	0.1763	