# Memory and Beliefs in Financial Markets: A Machine Learning Approach [†]

Zhongtian Chen[‡] Jiyuan Huang[§]

November 12, 2023

**Abstract**

We develop a machine learning (ML) approach to establish new insights into how memory affects financial market participants' belief formation processes in the field. Using analyst forecasts as proxies for market beliefs, we extract analysts' mental contexts and recalls that shape forecasts by training an ML memory model. First, we find that long-term memories are salient in analysts' recalls. However, compared to an ML benchmark trained to fit realized earnings, analysts pay more attention to distant episodes in regular times but less during crisis times, leading to recall distortions and therefore forecast errors. Second, we decompose analysts' mental contexts and show that they are mainly shaped by past earnings and forecasting decisions instead of current firm fundamentals as indicated by the ML benchmark. This difference in contexts further explains the recall distortion. Third, our comprehensive memory model reveals the significance of specific memory features and channels in analysts' belief formation, including the temporal contiguity effect and selective forgetting.

**Keywords:** Memory, Belief Formation, Machine Learning, Experience Effects, Analysts

[†]We are grateful to Jules van Binsbergen, Marius Guenzel, Alex Imas, Michael Kahana, Max Miller, Sean Myers, Marina Niessner, James Paron, Nikolai Roussanov, Jessica Wachter, Hongjun Yan and seminar participants at Wharton Finance, the Memory/Belief/Choice Group at the University of Pennsylvania, TADC 2023 and CFRC 2023 for helpful and valuable comments.

[‡]Department of Finance, The Wharton School, University of Pennsylvania; Email: chenzt@wharton.upenn.edu

[§]Department of Banking and Finance, University of Zurich; Swiss Finance Institute; Email: jiyuan.huang@bf.uzh.ch

# 1 Introduction

Financial agents make investment and economic decisions based on their beliefs about the future state of the economy. Thus, how financial agents' form their beliefs is crucial to the understanding of asset prices and economic decisions. The vast majority of economic models assume that agents hold rational expectations in the sense that agents have the full knowledge of the underlying dynamics of the economy and behave rationally, or at least, they exploit all relevant information to form expectations (Brunnermeier et al., 2021; Jiang et al., 2022). However, ample evidence in behavioral finance has shown that many psychological principles deviate agents' beliefs from the classical rational expectations. One of the burgeoning studies in behavioral finance literature is investigating the importance of memory in belief formation process (Bordalo, Gennaioli, and Shleifer, 2020; Wachter and Kahana, 2022). This area of research builds off a wealth of evidence from the psychology literature which suggests that an event that happens today might trigger the retrieval of certain similar past experiences from the memory database. The process of retrieving relevant experiences from the memory is called recall. Then the recalled episode serves as a reference point guiding the agents' belief formation today (Kahana, 2020). Until now, this research has mostly been theoretical and there have been few empirical studies, leaving a gap in our understanding of how significant the memory mechanisms are and how these mechanisms affect agents' economic decisions in the field. Therefore, we aim to fill this gap.

In this paper, we develop a new approach to show the impact of memory on financial market participants' belief formation processes by illustrating two essential but deep underwater concepts in memory models - recall and context, and the role of memory features and channels including the temporal contiguity effect and selective forgetting, with field evidence. We study the sell-side analysts, whose forecasting decisions are usually taken as benchmarks when analyzing financial market participants' belief formation processes[1].

We model analysts' belief formation processes in the following way. First, we follow van Binsbergen, Han, and Lopez-Lira (2020) and model the inputs that analysts exploit when making forecasts are the high-dimensional public signals including firm-specific fundamentals, macroeconomic variables, historical earnings decisions and stock market returns. Second, analysts process the public signals through a memory system, and the memory system outputs analysts' current mental context which is the analysts' perceived state of the

[1]For example, Bordalo et al. (2019), De La O and Myers (2021), and Brunnermeier et al. (2021).

covered firm. Third, analysts make forecasting decisions based on a linear function of the mental context.

We employ a machine learning memory model called Long Short-Term Memory model (LSTM, hereafter) by Hochreiter and Schmidhuber (1997) to model the analysts' memory system. By adapting LSTM, we are able to leverage the indispensable properties of both neuroscience and machine learning research in this study. First, LSTM is supported by evidence from neuroscientific research and contains multiple memory features and channels. Specifically, LSTM is a valid memory model because it can produce the three major laws of human memory: recency, temporal contiguity, and semantic similarity[2]. The three laws are well-documented in numerous empirical and laboratory settings, also in financial markets (Charles, 2022). Second, in contrast to traditional memory experiments conducted in laboratory settings, the memory systems of agents in financial markets need to cope with high-dimensional and non-stationary features that exhibit complicated functional forms. As machine learning methods have been designed specifically to tackle such challenges, they are better suited to our empirical studies than other memory models introduced in finance and economic literature, which make over-simplified assumptions, for example, Wachter and Kahana (2022) and Bordalo, Gennaioli, and Shleifer (2020).

LSTM is a type of neural network architecture that comprises one dynamic memory cell and three control gates. The memory cell represents the agent's memory state, while the three gates control how much information should be allowed to pass through to update the memory cell (forget and input) and output the mental context. With its recurrent structure, in each period, LSTM processes input features with the agent's existing memory and mental context, updating the memory and mental context accordingly, and using them for processing the next set of information. The context vector in LSTM is dynamic and evolves endogenously in response to the feature and memory stimuli, which reflects the evolving internal mental state. This dynamic context structure is advocated by recent memory literature (Howard and Kahana, 2002; Polyn, Norman, and Kahana, 2009) emphasizing the importance of context as an evolving internal mental state, as opposed to a static context setting in Bordalo, Gennaioli, and Shleifer (2020).

Using the data on public signals and analysts' forecasting decisions from 1986 to 2020, we train our memory model to fit the analysts' consensus forecast revisions, and extract

[2]Recency means people refer to recently experienced events while retrieving memory, temporal contiguity indicates people tend to recall an event that occurred contiguously in time to presently-recalled event, and semantic similarity means people are more likely to access the events that are most similar to that they are experiencing.

the analysts' latent memory and mental context vectors. The training process follows the standard approach in the machine learning literature (Gu, Kelly, and Xiu, 2020). Our memory model outperforms the baseline logistic regression in fitting the analysts' forecasts both in-sample and out-of-sample. The comparison provides strong support for incorporating rich memory mechanisms in understanding financial market participants' behaviors.

We present the analysts' recalled episodes based on the extracted analysts' mental contexts. We adopt the standard approach in memory literature and define the process of contextual-cued recall as searching for the most similar context vector in history (Kahana, 2020). We denote the recalled episodes as pairs of firm and time. For example, when thinking of firm $A$ today, analysts may recall firm $B$ at time $t$ since the context for today's firm $A$ is similar to the context for firm $B$ at time $t$. In this case, $(B, t)$ is considered the recalled episode for today's firm $A$. To demonstrate the impact of memory on the aggregate market, we focus on the consensus forecasts and assume a representative analyst who knows the entire market history and can recall any firms in the same industry. By examining the representative analyst's recalled episodes, our approach provides detailed evidence of how to quantify the relative impact of past experiences under different market conditions. We find that the well-established rules of human memory retrieval, such as recency effect and temporal contiguity, are significant. Analysts typically focus on recent episodes and recall those that occurred contiguously in time with recently-recalled episodes. Moreover, the impact of past experiences is, on average, decreasing in year gaps, which is consistent with the assumption made in the experience effect literature (Malmendier and Nagel, 2011). However, this assumption is challenged by the evidence that long-term experiences and memories are salient and even dominate the recent episodes in certain periods. For example, during the COVID pandemic in 2020, the market focused more on the crash and recovery period of the 2008 global financial crisis than the recent quarters[3]. Our findings thus present new disciplines of modeling the impact of past experiences in different time periods.

The question of whether analysts recall the appropriate historical episodes in response to current events remains unanswered. But this is of utmost importance since although forming beliefs or making decisions based on these recalled episodes may be rational, deviation from full rationality may occur if the recalled episodes are misleading or distorted. To answer this question, we follow van Binsbergen, Han, and Lopez-Lira (2020) and provide a machine learning benchmark. The benchmark recalls are defined as the econometrician's recalled

[3]Such results are also reconciled with the survey evidence shown in Jiang et al. (2022) that investors tend to recall both recent episodes and dramatic episodes.

episodes, who just replicates the true data-generating process of the realized earnings revision decisions (In what follows, we will use the terminology "machine learning benchmark" and "econometrician" interchangeably). The process of searching for the benchmark recalls is analogous to the process of searching for the representative analyst's recalls, but we apply the memory model to fit the realized earnings revisions instead of the analysts' consensus forecast revisions. Our results indicate that, compared to the benchmark recalls, analysts tend to pay more attention to distant episodes in regular times but less during crisis periods. This suggests that analysts may underreact to changes in external conditions, which can be rationalized by the confirmation bias. We further find robust evidence that analysts may recall the wrong episodes. For example, during the 2010-2014 period, analysts recalled episodes more related to the end of the boom period (before 2008), but the benchmark recalls rarely fell into that period. A paired $t$-test presents that such recall distortions are significant. This comparison reinforces the importance of incorporating long-term memory and distant experiences into the modelling of financial market participants' belief formation processes and highlights the deviation of agents' belief formation from the full rationality. To study the economic impact of these wrong recalls, we develop a test and show that analysts' misleading recalls could explain their forecast errors. The positive relation between the recall distortion and forecast errors suggests that when analysts recall over-optimistic (over-pessimistic) episodes, they tend to make over-optimistic (over-pessimistic) forecasts. Specifically, a 10% increase in recall distortion leads to a 7.88% increase in the probability of jumping to a more optimistic level of forecast revisions.

To gain a better understanding of the analysts' contextual-cued recalls and recall distortions, we need to investigate the formation of mental contexts. Mental context is essential in memory models, as it links the encoding and retrieval of information in the memory system (Wachter and Kahana, 2022). However, mental context as agents' internal mental state, is latent in many memory models, including those introduced by Kahana (1996) and Wachter and Kahana (2022). In this paper, we are the first to systematically explore the black-box of mental context and memory cell in the domain of financial markets, providing insights into how the variable importance of contexts and memories changes over time. We divide all the features into four groups: firm fundamentals, macroeconomic variables, stock market returns, and historical earnings-related variables. Then we evaluate the importance of each group by calculating the reduction in the R-squared when the covariates in that group were set to zero one at a time[4]. The regression is performed based on the 1-year

[4]The method follows Gu, Kelly, and Xiu (2020).

rolling window. By examining the time-series plot of the group variable importance of the analysts' and the machine learning benchmark context and memory vectors, we make the following findings. First, we show that during the recession period, macroeconomic variables are particularly salient, confirming the prediction of the limited attention theory by Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) that agents pay more attention to aggregate news during economic downturns. Second, we document that the analysts and the machine learning benchmark differ in the composition of their contexts and memories. Analysts over-weight the importance of historical earnings and forecasting decisions, whereas the machine learning benchmark mainly focuses on current firm fundamentals. These differences in context composition explain recall distortions among analysts and provide support for encoding errors (Woodford, 2020), self-herding bias (Hirshleifer et al., 2019) and limited attention (Hirshleifer and Teoh, 2003) in the field. These results shed new light on how future research on memory in financial markets should conceptualize mental contexts.

Mental context also serves as a powerful representation of an agent's past experiences and the information they have processed. To demonstrate this, we conduct a test that reveal how the different mental contexts formed through analysts' different covering experiences can lead to variations in their forecasting decisions. Conversely, when we attempt to explain the analysts' forecast dispersion using the lengths of their respective covering experiences, we found no significant correlation. These results provide compelling evidence that memory models and mental contexts can help us better study the financial market participants' experience effect (Malmendier and Nagel, 2011) when navigating complex and high-dimensional problems.

The human memory system is complex, with multiple features and channels. However, our approach which employs the a comprehensive memory model - LSTM, allows us to examine the role of specific memory features. First, we quantify the significance of the temporal contiguity effect in the field, in addition to the theoretical argument made by Wachter and Kahana (2022). Through a simulation study using the model trained to fit the representative analyst's forecasting decisions, we demonstrate that the results align with the prediction of the temporal contiguity effect, emphasizing the need for a valid memory model like LSTM to capture analysts' belief formation processes[5]. Second, we perform a counterfactual analysis to show how selective forgetting works. We systematically break

[5]Charles (2022) documents the temporal contiguity effect in the setting of market responses to earnings announcements. We study temporal contiguity effect in the setting of analyst forecasts and belief formation, and further indicate the significance of forward asymmetry which is an important property of the temporal contiguity effect (Howard and Kahana, 2002). For details, see Section 6.1.

down the forget gate in LSTM and assess which experiences the representative analyst and the machine learning benchmark selectively forget. Without the forget gate, the recency effect for the machine learning benchmark is attenuated, while for the analysts, the recency effect is reinforced. The results suggest that analysts tend to ignore more recent experiences than the distant experiences, and indicate that for analysts, the distant experiences have a stronger long-lasting effect than for the machine learning benchmark. Such differences can contribute to explaining the recall distortion between analysts and the benchmark. The pattern of selective forgetting is also reconciled with the evidence that analysts perceive the world as more stationary than it is (De La O and Myers, 2022). This counterfactual analysis is not feasible in other empirical memory studies so far.

This paper contributes to the growing literature on application of human memory in economics and finance. We provide strong empirical evidence in support of the theoretical memory literature in this field (Bordalo, Gennaioli, and Shleifer, 2020; Wachter and Kahana, 2022; Nagel and Xu, 2022), and introduce a novel structural approach to studying the impact of memory with field evidence. This approach offers a distinct advantage over survey-based methods (Jiang et al., 2022) and reduced-form analyses (Charles, 2022; Goetzmann, Watanabe, and Watanabe, 2022) by allowing us to extract agents' memories, recalls and contexts at any specific historical time, an otherwise infeasible task with other methods. By illustrating how financial market participants' memories look and how they influence belief formation, we establish the new disciplines of memory modelling in financial markets that are guided by the extracted recalls and mental contexts, as well as the role of specific memory channels. Our approach is unparalleled in its ability to derive these new disciplines, and is easily translatable to other settings to study the impact of memory in the field.

The paper strengthens the literature on experience effects by developing a micro-founded and unified approach to studying the impact of various types of experiences on agents' behavior. While previous research, such as the work of Malmendier and Nagel (2011, 2016), has employed reduced-form identification strategies to examine specific types of past experiences, our approach incorporates neuroscientific underpinnings and memory mechanisms to avoid over-simplified assumptions. By leveraging the context derived from our memory model, we are able to simultaneously study the impact of and interaction between multiple types of experiences in a high-dimensional and realistic setting. This approach is more reflective of the complex nature of experiences in the real world, where multiple types of experiences can interact with each other to influence agents' beliefs and behavior.

This paper also adds to an emerging literature that applies machine learning methods

in finance and economics. Gu, Kelly, and Xiu (2020) and Chen, Pelger, and Zhu (2023) bring neural networks to this field and use them to predict the panel of individual U.S. stock returns and estimate the stochastic discount factor (SDF), respectively. Chen, Pelger, and Zhu (2023) also exploit LSTM to extract the low-dimensional hidden state from the non-stationary, cyclical, and high-dimensional macroeconomic variables. We jump out of the framework of predicting and explaining cross-section returns, and develop the neural network to uncover the underlying mechanisms that drive financial market participants' behaviors. This paper also builds on van Binsbergen, Han, and Lopez-Lira (2020) and Bianchi, Ludvigson, and Ma (2022). They both use machine learning approaches to provide a real-time unbiased and optimal benchmark for interested financial indexes, and then demonstrate how analysts' beliefs or behaviors are distorted compared to the benchmark. In contrast, we use machine learning methods to replicate the analysts' expectation formation processes directly; in such a way, we can show which attributes exactly lead to agents' biased beliefs, and specifically, we suggest that memory can explain the biased beliefs.

In general, this paper inspires a new research avenue of applying machine learning methods to behavioral finance, as neural networks and reinforcement learning are designed to replicate the human brain and mimic how humans make decisions. Similar concepts are just beginning to emerge in the behavioral literature. In their theoretical work, Barberis and Jin (2021) analyze the framework with model-free and model-based learning in asset pricing, which is consistent with the main principles of reinforcement learning in the machine learning literature. We show that by exploiting machine learning's virtues of dealing with high-dimensional problems and neuroscientific foundations, we can empirically analyze the deep underwater but important mechanisms in the domain of behavioral finance.

The remainder of the paper is organized as follows. Section 2 illustrates the structure of the model. Section 3 describes the data and how we proceed the training, validation and testing of the machine learning memory model. Section 4 defines the process of contextual-cued recall, and find the recalled episodes. Section 5 shows how mental contexts vary over time and its role in belief formation process. Section 6 discusses the significance of specific memory channels - the temporal contiguity effect and selective forgetting. Section 7 concludes.

# 2 Model

This section introduces the long short-term model (LSTM) and how it works in analysts' belief formation process. One of key features in belief formation process is memory's long-term dependence. Past experiences is able to affect analysts' decision right now. LSTM model is designed to tackle down the vanishing gradient problem which is crucial when the model exists long-term dependence (Pascanu, Mikolov, and Bengio, 2013).

## 2.1 Whole Model Structure

Figure 1 presents the LSTM aggregate model structure of one analyst and one firm as time progresses from left to right. The basic block of a LSTM aggregate model is a LSTM cell. The input of a LSTM cell are mental context in the last period $h_{t-1}$, memory in the last period $c_{t-1}$ and input features $X_t$ while the output are mental context in the current period $h_t$ and memory in the current period $c_t$. After receiving new information, analysts update their long-term and short-term beliefs $c_t$ and $h_t$ and make new forecasts. In our model, short-term mental context $h_t$ is an efficient representation for an analyst to make decisions but long-term memory $c_t$ can affect $h_t$ in the future dates. LSTM model is a recurrent model, which means that the output of a LSTM cell at period $t-1$ are used as the input of the LSTM cell at period $t$. The output of the aggregate LSTM model $AF_t$ depends on a softmax function of a linear model of the mental context $h_t$.

To be specific, we suppose analyst $i$ started to cover firm $j$ from time $s_{i,j}$[6], after observing the history of input features $X_{j,s_{i,j}}, X_{j,s_{i,j}+1}, ..., X_{j,t}$, the analyst $i$ encodes these observed features and forms a latent context vector

$$h_{i,j,t} = LSTM\_cell(h_{i,j,t-1}, c_{i,j,t-1}, X_{i,j,t}) = LSTM(X_{j,s_{i,j}}, X_{j,s_{i,j}+1}, ..., X_{j,t})$$

from the memory model for firm $j$ in month $t$. Then following So (2013), the analysts decode the mental context vectors and makes final forecast decision $AF_{i,j,t}$ simply by a linear decision function of the context vector. Specifically, $AF_{i,j,t}$ is the classification decision of forecast revision (revise up, revise down, or remain the same), and $AF_{i,j,t}$ is obtained by a softmax

[6]We define the analyst's experiences on the covered firms from the time when the analyst started to cover the firm. Since we consider analysts as experts on the financial market, we are interested in their professional experiences, in contrast with Malmendier and Nagel (2011) which focus on the household's life-time experience.

function with a linear combination of $h_{i,j,t}$

$$Prob\left(AF_{i,j,t}=k\right)=\frac{e^{h_{i,j,t,k}^{T}\beta}}{\sum_{l=1}^{K}e^{h_{i,j,t,l}^{T}\beta}}.$$

The framework can be easily extended to more complicated models[7] if one is interested in other specific memory or neuroscientific channels.

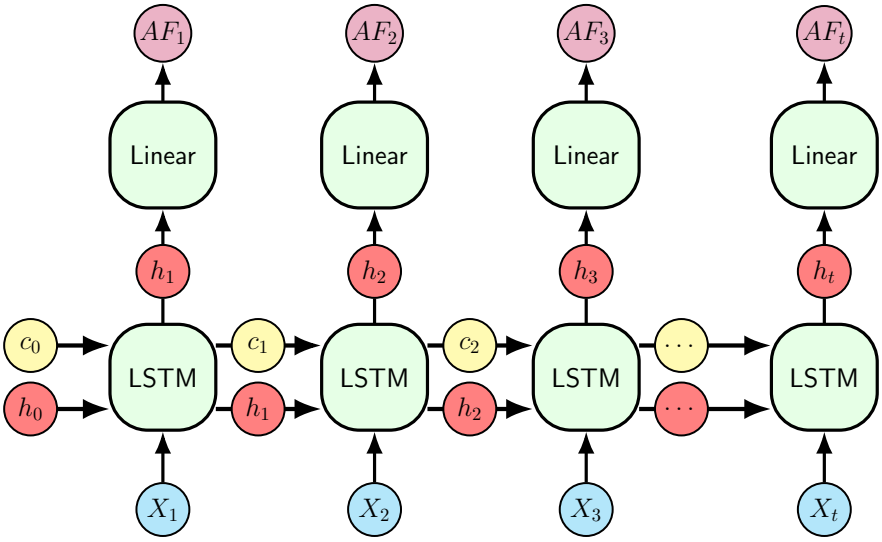



Figure 1: Whole Structure

Next, we introduce the variables in the model in detail, before demonstrating the memory model.

## 2.2 Variables

First, let $X_{j,t} \in \mathbb{R}^M$ denote the public signals that analysts rely on to make forecasts in month $t$ for firm $j$, where $M$ is the dimension of the input features. $X_{j,t}$ include macroeconomics variables, historical earnings-related variables, stock market return, and firm fundamentals, the details of these input features will be demonstrated in next section. The model aims to replicate and describe analysts' forecasting process, hence the input features are chosen to be available to analysts before month $t$ (before the first day of each month) to avoid the look-ahead bias. That is to say, although that $X_{j,t}$ are the input features that analysts may use to make forecasts in month $t$, all of them are actually publicly announced before month $t$ and most of them are published during month $t-1$.

[7]Beyond LSTM, for example, Vaswani et al. (2017), Weston, Chopra, and Bordes (2014) and Graves et al. (2016) incorporate more comprehensive memory or neuroscientific channels into their models to mimic human decision-making processes.

Second, let $AF_{i,j,t}$ denote the earnings forecasting decision made by analyst $i$ in month $t$ for firm $j$. We set $AF_{i,j,t}$ as the discrete forecast revision as:

$$AF_{i,j,t} = \begin{cases} -1 & \text{if } AFlevel_{i,j,t} < \overline{AFlevel}_{j,t-1}, \\ 0 & \text{if } AFlevel_{i,j,t} = \overline{AFlevel}_{j,t-1}, \\ 1 & \text{if } AFlevel_{i,j,t} > \overline{AFlevel}_{j,t-1}, \end{cases}$$

where $AFlevel_{i,j,t}$ denote analyst $i$'s EPS level forecast for firm $j$ in month $t$, and $\overline{AFlevel}_{j,t-1}$ denote analysts' mean forecast (consensus forecast from last period). In other words, $AF_{i,j,t} = -1$, if analysts revise down the forecast; $AF_{i,j,t} = 1$, if analysts revise up the forecast; and $AF_{i,j,t} = 0$, if the forecast is remained the same. Since individual analysts may not issue forecasts every month, in order to keep most of the data, we compare individual analyst's forecasts available today with the consensus forecasts obtained from last period. Similarly, we define consensus analyst's forecasting decision as:

$$AF_{j,t} = \begin{cases} -1 & \text{if } \overline{AFlevel}_{j,t} < \overline{AFlevel}_{j,t-1}, \\ 0 & \text{if } \overline{AFlevel}_{j,t} = \overline{AFlevel}_{j,t-1}, \\ 1 & \text{if } \overline{AFlevel}_{j,t} > \overline{AFlevel}_{j,t-1}. \end{cases} \tag{2.1}$$

We also follow van Binsbergen, Han, and Lopez-Lira (2020) and provide an unbiased benchmark for firms' earnings expectations in real-time conditional on the same information set that analysts hold in month $t$, this can be taken as the view of an econometrician or a machine learner who replicates the data-generating process of realized earnings decisions. The realized earnings revision decisions are analogously defined as shown in Equation (2.1):

$$RE_{j,t} = \begin{cases} -1 & \text{if } EPS_{j,t} < \overline{AFlevel}_{j,t-1}, \\ 0 & \text{if } EPS_{j,t} = \overline{AFlevel}_{j,t-1}, \\ 1 & \text{if } EPS_{j,t} > \overline{AFlevel}_{j,t-1}, \end{cases} \tag{2.2}$$

where $EPS_{j,t}$ denote the realized annual earnings per share of firm $j$ at the same fiscal end date as that for $\overline{AFlevel}_{j,t-1}$. Thus, by comparing $EPS_{j,t}$ with $\overline{AFlevel}_{j,t-1}$ (instead of $EPS_{j,t-1}$), $RE_{j,t}$ is interpreted as the correct forecast revision decision that the analysts should have made conditional on the consensus forecast in the preceding period.

Third, let $h_{i,j,t}$ denote the $K$-dimensional latent vector that analyst $i$ has in mind and uses to make forecasts for firm $j$ at time $t$ according to their past experiences of input features and the memory process. $K$ is a model hyperparameter, which will be tuned according to the model performance in the validation sample (See Section 3.1). Vector $h$ stands for

the short-term highly generalized information from the past experiences and current inputs, and is the analysts' mental context vector. One example of the short-term generalized information is the default rate of one-year corporate bond. Bordalo, Gennaioli, and Shleifer (2020) first introduce the importance of context on economic decisions, they consider the static context which represents for the external physical state, such as the locations or the weather. However, as pointed out by Wachter and Kahana (2022), the notion of context should go beyond the physical environment and be more abstract. Hence in this paper, we follow Wachter and Kahana (2022) and denote the context vector $h$ as the analysts' mental account, or the agent's perceived internal state. As will be shown in the details of LSTM, the context is dynamic and evolving endogenously according to the feature and memory stimuli, such properties are consistent with the recent memory literature (Glenberg and Swanson, 1986; Howard and Kahana, 2002; Polyn, Norman, and Kahana, 2009; Kahana, 2020) which addresses the importance of taking context as an evolving internal mental state, so that the model could incorporate the time dimension of memory-driven decisions and support several major memory evidences (e.g., recency, contiguity effects and effects of early-life experience).

Lastly, let $c_{i,j,t}$ denote the $K$-dimensional memory cell. It records the information people store in their memory. In contrast to the context vector $h$, the memory cell $c$ stores the relatively long-term generalized information, such as the firm's position in the product-life cycle and whether the firm is growth firm or value firm.

## 2.3 Long Short Term Memory Networks

### 2.3.1 Recurrent Neural Network

LSTM is a modified version of Recurrent Neural Network (RNN), which overcomes the shortcoming of RNN that can not learn the long-term dependency. To better understand the advantage of LSTM, we first introduce RNN.

Building on the traditional neural network, RNN is capable of dealing with sequence (e.g. time-series) data, especially using its reasoning about previous state to predict current decisions. RNN has a chain-like structure and its components are RNN cells. Figure 2 shows an unrolled structure of RNN. The RNN cells (the green box in the figure) in each period remains the same, but it allows the information obtained and concluded from previous period to be sent to next period. Then, according to current inputs $X_t$ and messages from last period, the network will generate the new output and pass new messages $h_{t+1}$ to next

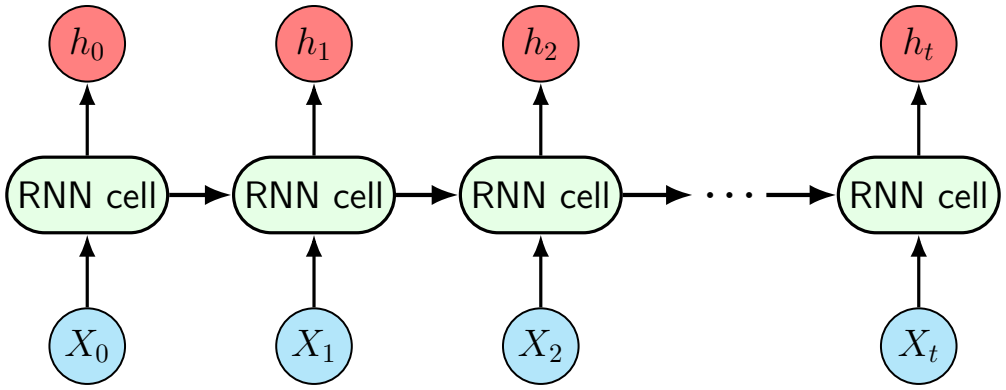



Figure 2: The Structure of RNN

cell. In a nutshell, RNN has the advantage of capturing time-series dependencies. Comparing to Hidden Markov Models (HMM), RNN shares the same intuition, but allows more flexible function forms and without imposing strong distribution assumptions. RNN has achieved success in the areas that need to deal with sequential data, for example, translation and speech recognition.

However, RNN's advantage is also its weakness: in practice, only short-term dependencies could be captured, and RNN is hard to handle long-term dependencies. That is the reason why researchers refer to RNN as the "short-term memory" model. This issue is well studied in theory, for example Bengio, Simard, and Frasconi (1994). Literature on behavioral finance has found solid evidence on the effect of very-long-term dependencies on economic decisions as well. For instance, Malmendier and Nagel (2011) show that life-time experiences on market returns affect agents' investment decisions, Wachter and Kahana (2022) also state that long-term memory does not disappear and might influence agents' decisions today once retrieved. Thus, modelling long-term dependencies is crucial in this paper's setting which studies financial market participants' expectation formation processes. More importantly, as argued in Howard and Kahana (2002), the RNNs of the kind developed by Elman (1990) are not capable of implementing one of the fundamental memory principals - temporal contiguity[8]. This further dampens the validity of RNN as a memory model. However, LSTM can solve these issues.

[8]Section 6.1 explains this issue with details on theory, and Section A.5 provides field evidences.

### 2.3.2 The Structure of LSTM

LSTM is introduced by Hochreiter and Schmidhuber (1997). LSTM in general, is a variant of RNN model. LSTM also has a chain-like structure but with different cells. A typical LSTM cell consists of one memory cell $c \in \mathbb{R}^K$, one hidden context $h \in \mathbb{R}^K$ and three gates: forget gate, input gate and output gate. The key innovation of LSTM is having a memory cell to store information for long periods of time, while the three gates control which information to be erased, stored and output from the memory cell.

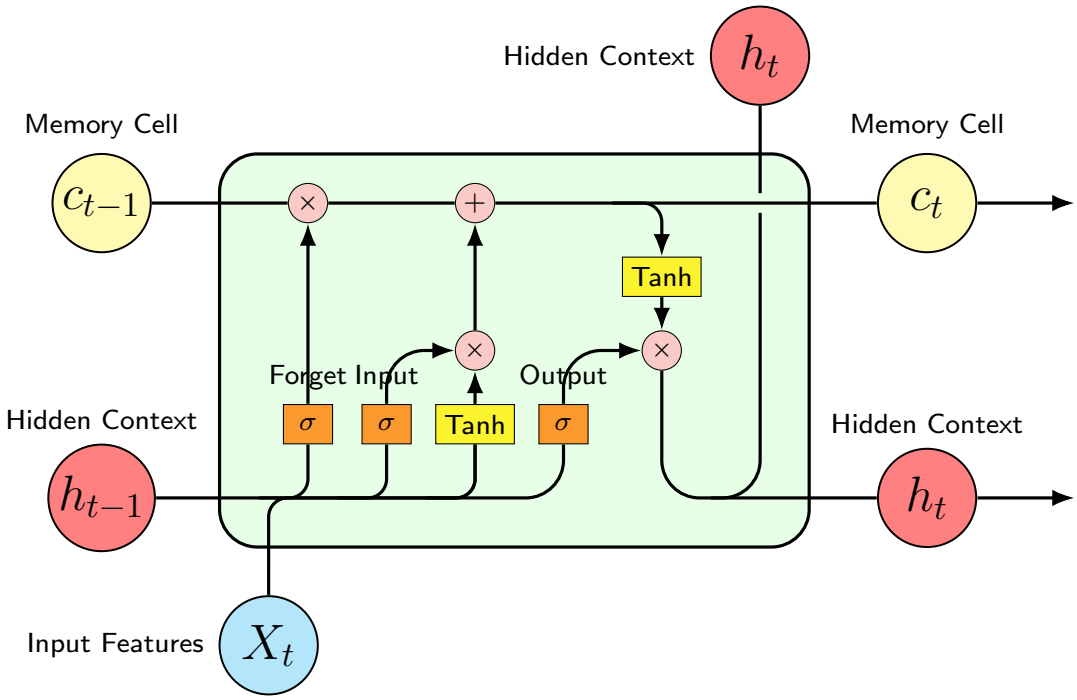



Figure 3: The Structure of LSTM

Figure 3 presents the structure of one LSTM cell. The three gates are shown in the orange boxes, which are all controlled by the sigmoid activation function displayed in orange boxes. The sigmoid activation function is an element-wise sigmoid function ($\sigma$), it outputs numbers between zero and one, and then controls the extent to which information should be passed through the gates. The operators "$\times$" and "$+$" shown in the pink circles are the element-wise multiplication operation and the element-wise addition operation, respectively. "Tanh" in the yellow boxes is the hyperbolic tangent activation function[9].

[9]The incorporation of these activation functions (tanh and sigmoid) has several advantages: first, they introduce the non-linearity to the structure which fits more closely to empirical economic and financial data (Teräsvirta, 2006); second, they further standardize the variables and reduce the impact of outliers; third, they make model estimation more accurate and efficient according to machine learning literature (Dubey, Singh, and Chaudhuri, 2022; Jagtap, Kawaguchi, and Karniadakis, 2020).

The memory process takes following steps. First, the forget gate controls the extent to which the information remains in the memory cell, an output of zero from the sigmoid function means the information should be completely erased while an output of one means the information should be completely retained in the memory cell.

$$forget_t = \sigma(W_h^f h_{t-1} + W_x^f X_t + w_0^f)$$

The output of the sigmoid function is determined by the linear combination of current input $X_t$ and previous hidden state $h_{t-1}$. Including both $X_t$ and $h_{t-1}$ is important, for example, in a natural language processing environment, the same word "apple" ($X_t$) may refer to the fruit or the technology company under different semantic context $h_{t-1}$.

Second, the input gate controls the extent to which the information should be stored into the memory cell, similarly, the information will be determined meaningless and will not be passed to the memory cell if the sigmoid function outputs zero, while it will be completely stored into the memory cell if the sigmoid function outputs one. Before the information goes through, it will be encoded by the *tanh* function (the values are now between -1 and 1), denoted as $\tilde{c}_t$:

$$\tilde{c}_t = tanh(W_h^c h_{t-1} + W_x^c X_t + w_0^c)$$
$$input_t = \sigma(W_h^i h_{t-1} + W_x^i X_t + w_0^i).$$

Third, according to the forget gate and input gate, we can update the memory cell

$$c_t = forget_t \times c_{t-1} + input_t \times \tilde{c}_t,$$

by erasing contents from the memory cell that should be forgotten and adding in new values that are considered valuable.

Finally, the output gate controls the extent to which the information should be elicited from the updated memory cell through a similar sigmoid function:

$$out_t = \sigma(W_h^o h_{t-1} + W_x^o X_t + w_0^o).$$

Then, the final output, current context vector $h_t$ will be chosen by the output gate and the information from the current memory cell through the *tanh* function:

$$h_t = out_t \times tanh(c_t).$$

### 2.3.3 Why LSTM?

The contribution of adapting LSTM in this paper's setting is twofold. First, LSTM is a valid model of human memory and provides a structural form to study the memory process. As suggested by the memory literature and ample experimental and field evidence (Howard and Kahana, 2002; Kahana, 1996; Wachter and Kahana, 2022), a valid model of human memory should incorporate the three basic laws of human memory system: recency, temporal contiguity, and semantic similarity. Recency means people refer to recently experienced events when accessing memory, temporal contiguity means people tend to recall an event that occurred contiguously in time to presently-recalled event, and semantic similarity means people are more likely to access the events that are most similar to that they are experiencing. LSTM adapts the retrieved-context model (Wachter and Kahana, 2022) in two ways: first, the context $h$ is evolving according to the association of external feature stimuli, inner memory process and previous mental context state, and $h$ generally has an autoregressive structure; second, incorporating a memory cell helps store past information and generate long-term dependency. In terms of producing the three major laws of human memory, according to Howard and Kahana (2002), theoretically, the autoregressive structure of mental context $h$ can help implement recency, the combination of autoregressive context and memory cell embeds the channel of temporal contiguity, and the semantic similarity arises naturally from the definition of contextual-cued recalls (see Section 4.1). Beyond Wachter and Kahana (2022), LSTM first has the advantage of admitting more flexible functional forms. For example, they only model the input features as basis vectors which is impractical for empirical analysis in a high-dimensional environment, but our framework is suitable for any types of inputs. Furthermore, LSTM contains more structural memory channels which allows for the comprehensive counterfactual analysis (see Section 6). For instance, forget gate controls explicitly the content to fade away from the memory bank, which is not embedded in the model of Wachter and Kahana (2022).

Second, analysts need to deal with non-stationary and high-dimensional financial variables that may exhibit complicated functional forms. LSTM inherits the advantages that most machine learning techniques possess in dealing with these difficulties. For example, machine learning methods are capable of feature selection and dimension reduction (Nagel, 2021). The three gates filter out the redundant information and features. This is more flexible and more realistic than the model shown in Bordalo, Gennaioli, and Shleifer (2020), they apply all kinds of external environmental variables to represent context and then get the cued recall for decision-making accordingly. However, it is obvious that not all the features,

especially external ones are used by agents for their decision-making and belief formation. LSTM presents a solution to this problem. We also assign a small number to the dimension of hidden context vector, specifically we have 79 input features, but for the analyses below we set $K = 10$ as the dimension of the final mental context vector. LSTM is also capable of capturing underlying dynamics from high-dimensional variables, as documented by Chen, Pelger, and Zhu (2023)[10].

In summary, we model the analysts' forecasting process as:

1. Current public signals $X_t$, previous context vector $h_{t-1}$ and memory state $c_{t-1}$ flow into the memory system LSTM;
2. The memory system LSTM update the memory state $c_t$, and output current context vector $h_t$;
3. Analysts make forecasts according to current context $h_t$ with a linear decision function.

# 3 Data and Model Training

## 3.1 Input Features and Analyst Forecasts

Following van Binsbergen, Han, and Lopez-Lira (2020), we exploit an extensive set of monthly public signals as input features, including financial ratios from WRDS[11], other firm-specific fundamentals from COMPUSTAT, Macroeconomic variables from the Federal Reserve Bank of Philadelphia, and earnings-related variables from I/B/E/S. The full list of input features is shown in Table A1, there are 79 variables in total. The sample period spans January 1986 to December 2020. In order to minimize the effect of extreme data points, all variables are winsorized at the 2.5% level in cross-section at each time point. For detailed explanations and data processing, one can refer to van Binsbergen, Han, and Lopez-Lira (2020).

The realized earnings and analysts' EPS forecasts are from the I/B/E/S database. We obtain both the consensus forecasts (mean forecasts) and individual forecasts, and focus on just the one-year-head forecasts for annual earnings (IBES $FPI$ of 1).

[10] Chen, Pelger, and Zhu (2023) present that LSTM could successfully extract the hidden states from the non-stationary and cyclical dynamic structure of macroeconomic variables.

[11] https://wrds-www.wharton.upenn.edu/pages/grid-items/financial-ratios-firm-level/

## 3.2 Training, Validation, and Test

We are interested in whether certain memory mechanisms work out at the aggregate level and due to the limit of computational power, we train the model by considering the view of a representative analyst. We assume that the representative analyst knows all the information in the market about all the firms and at any time. Thus, we pool all the firms from the entire market history together, then estimate the generalized model from the consensus forecasts $AF_{j,t}$.

We follow the setup in Gu, Kelly, and Xiu (2020) and standard approach for classification problem in machine learning literature to employ the sample splitting and performance evaluation schemes, as well as design the model training process. The base training sample is from January 1990 to December 2004. The validation sample spans January 2005 to December 2006[12]. The model training process is as following: we employ the Adam algorithm as the optimization algorithm for stochastic gradient descent (Kingma and Ba, 2014) with default hyperparameters; implement the early stopping scheme with patience equals to 5 to avoid over-fitting; set the batch size to 10,000; use the negative log-likelihood as the loss function since the final decision function is the logistic regression. Thus, the only hyperparameter needs to be tuned is the dimension of the latent context and memory vector- $K$. To elicit optimal $K$, we pick a set of candidates, then train the model with the training sample and evaluate the model performance with the candidate hyperparameter $K$ in the validation sample. Table 1 presents the model performance for fitting and predicting both the analysts' forecast revision ($AF$, see (2.1)) and realized earnings revision ($RE$, see (2.2)) in the base training and validation samples with different choices of $K$. According to the prediction accuracy in the validation sample, $K = 10$ is the best choice for analysts' forecast revision ($AF$) and $K = 5$ for realized earnings revision ($RE$), although the difference between models is marginal. Since we are mostly interested in interpreting analysts' behaviors and to keep comparison between analysts and the benchmark shown in later sections away from the impact of hyperparameters, we then pick $K = 10$ for both the models of analysts' forecast revision ($AF$) and realized earnings revision ($RE$) throughout the paper.

Beyond the base training sample, we also follow Gu, Kelly, and Xiu (2020) and employ the recursive scheme to train the rest of the samples and evaluate the performance. Specifically,

[12]All variables are standardized following the recommended guidelines in the literature: each variable is subtracted off by the mean and divided by the standardized deviation that calculated using all data points in the base training sample. Then apply the mean and standardized deviation to the validation and test sample.

Table 1: Model performance with different hyperparameter $K$

| Part I. Analysts' Forecast Revision ($AF$) | | | | |
|---|---|---|---|---|
| $K$ | 5 | 10 | 15 | 20 |
| Training | 59.58% | 61.42% | 62.26% | 62.04% |
| Validation | 54.22% | **55.45%** | 55.15% | 54.72% |
| Part II. Realized Earnings Revision ($RE$) | | | | |
| $K$ | 5 | 10 | 15 | 20 |
| Training | 68.68% | 71.22% | 72.33% | 71.74% |
| Validation | **59.58%** | 59.95% | 57.73% | 57.58% |

This table presents the model performance for fitting and predicting both the analysts' forecast revision ($AF$, see (2.1)) and realized earnings revision ($RE$, see (2.2)) in the base training and validation samples with different choices of dimension of the latent context and memory vectors ($K$). The training sample is from January 1990 to December 2004. The validation sample is from January 2005 to December 2006.

after setting the best hyperparameters $K$, we first train the model using data from January 1990 to December 2004, and perform out-of-sample analysis over 2005, then we increase the training sample by the data points in 2005, next we re-train the model starting from previously trained model and perform out-of-sample analysis over 2006, repeat this process recursively until 2020. Such a recursive scheme has two advantages: first, recursively fitting the model adapts the changing economic environment and financial market, as well as delivers more accurate model estimates; second, analysts' cognition is also evolving, recursive scheme could simulate people's re-cognition process when they accept new information. Table 2 shows the model out-of-sample performance with the recursive scheme, this includes both the validation sample and test sample, hence it starts from 2005 to 2020. For comparison purpose, a baseline logistic regression is also included, which simply regresses $AF_{j,t}$ or $RE_{j,t}$ on $X_{j,t}$ directly with the same recursive scheme and the identical dataset used by LSTM. For replicating analysts' forecast revision ($AF$), LSTM is significantly performing better in prediction accuracy than the baseline logistic regression, the gap is above 8%. But for the realized earnings revision ($RE$), the performance of the LSTM and the logistic regression is close[13]. The models' different prediction performance between analysts' forecast revision and realized earnings revision is reconciled with our findings in later sections that the analysts are more deeply influenced by the long-term memory and experiences than the machine learning benchmark is. The results also buttresses the superiority of LSTM in describing

[13]For in-sample performance, LSTM outperforms the logistic regression with around 10% for both the analysts' forecast revision and the realized earnings revision.

analysts' expectation formation process over the simple logistic regression and importance of the memory channels in modeling analysts' belief formation processes.

Table 2: Out-of-sample prediction accuracy of LSTM and Logistic regression

| Model | Analysts' Forecast Revision ($AF$) | Realized Earnings Revision ($RE$) |
|---|---|---|
| LSTM | 56.68% | 59.52% |
| Logistic | 48.21% | 58.46% |

This table shows the average out-of-sample prediction accuracy of LSTM and Logistic regression for both the aanalysts' forecast revision ($AF$, see (2.1)) and realized earnings revision ($RE$, see (2.2)) over the year 2005 to 2020. We apply the recursive scheme to evaluate the out-of-sample performance.

# 4 Recall

*"In response to current events, people often reach for historical analogies, and this occasion was no exception. The trick is to choose the right analogy."* - Bernanke (2015)

In this section, we present the "historical analogies" that the analysts reach for when making forecasts, the "right analogies" that the analysts should have chosen, and the implications of choosing the "wrong analogies" by analysts.

## 4.1 Analysts' Contextual-Cued Recalls

To find the analysts' recalled episodes when making forecasts, we follow the memory literature and several applications in finance to define the process of contextual-cued recall. An associative recall is the process of searching for similar past experiences from the contents of memory when we are encountering previously experienced item. This is the process of how brain generates familiarity. In this paper, we extract the recalls based on the context cue. Context is commonly exploited as a retrieval cue in literature, for example the Search of Associative Memory (SAM) retrieval model by Gillund and Shiffrin (1984). Similarly, Bordalo, Gennaioli, and Shleifer (2020) define the context cue based on the external environmental context such as locations, as well as a full cue which adds the interested key features: price and quality. The process defined in this paper is similar to these approaches. We first use the analyst's mental context ($h$) defined in Section 2 as the cue. Second, we get the recalls by searching for the most similar historical episodes with the similarity is defined

as the negative Euclidean distance between the current context $h_t$ and historical context $h_\tau$ $(\tau < t)$[14].

To show how the analysts' recalled "historical analogies" look like, for simplicity, we present the representative analyst's recalls[15]. The representative analyst is defined in Section 3.2. The process of searching for the representative analyst's recalled episodes is formally defined as following: first, for each firm $j$ at each time $t$, we extract the representative analyst's context vector $h^A_{j,t}$ from the trained model; second, since we assume the representative analyst knows the entire market history, we then search for the pair $(l^A, \tau^A)$ and the model induced latent context vector $h^A_{l^A,\tau^A}$ which stands for firm $l^A$ at time $\tau^A$, that maximizes the similarity to $h^A_{j,t}$:

$(l^A, \tau^A) = \text{argmax}_{(m,s)}\ Similarity(h^A_{m,s}, h^A_{j,t})$ for all $s < t$ and $Industry(m) = Industry(j)$[16].

The similarity function $Similarity(\cdot,\cdot)$ is the negative Euclidean distance between the two context vectors

$$Similarity(h^A_{m,s}, h^A_{j,t}) = -\sqrt{\sum_{k=1}^{K}\left(h^{A,(k)}_{j,t} - h^{A,(k)}_{m,s}\right)^2}, \tag{4.1}$$

where $h^{A,(k)}_{j,t}$ is the $k$-th component of the vector $h^A_{j,t}$ and $h^{A,(k)}_{m,s}$ is the $k$-th component of the vector $h^A_{m,s}$. We also introduce another similarity measure, for example cosine similarity:

$$Similarity(h_{m,s}, h_{j,t}) = \frac{h_{m,s}\cdot h_{j,t}}{\|h_{m,s}\|\|h_{j,t}\|} = \frac{\sum_{k=1}^{K} h^{(k)}_{m,s}h^{(k)}_{j,t}}{\sqrt{\sum_{k=1}^{K}\left(h^{(k)}_{m,s}\right)^2}\sqrt{\sum_{k=1}^{K}\left(h^{(k)}_{j,t}\right)^2}}, \tag{4.2}$$

where $\cdot$ is the dot product operation. The following results are robust to both similarity measures. Without loss of generality, we then mainly report the results based on the similarity definition shown in (4.1). We present some robustness check based on the cosine similarity

[14]To illustrate the general process of contextual-cued recall, we suppress the subscript $(i, j)$ for the analysts and firms, then we have

$$Similarity(h_\tau, h_t) = -\sqrt{\sum_{k=1}^{K}\left(h^{(k)}_{\tau} - h^{(k)}_{t}\right)^2},$$

where $h^{(k)}_t$ is the $k$-th component of the vector $h_t$. As shown in Kahana (2020), the general definition of similarity is of the form: $Similarity(h_\tau, h_t) = \exp\left(-\xi\|h_\tau - h_t\|_\gamma\right)$, where $\gamma$ is the distance metric, with $\gamma = 2$ denotes the Euclidean norm, and $\xi \geqslant 0$ measures how quickly similarity decays with distance. In this paper, we just apply the Euclidean norm and set the distance decay $\xi = 1$ for simplicity, the exponential function then can be ignored as we focus on the relative distance in the following analyses.

[15]One can easily extend our framework to study the heterogeneity among analysts, e.g., showing potentially different recalls from different age groups. We explore the heterogeneity to some extent in Section 5 by showing different covering experiences may lead to forecast dispersion.

[16]The industries are defined as in Fama-French 49 industry portfolios.

measure in Appendix A.2.

We show the extracted representative analyst's recalls in Figure 4. The darker blue gradients indicate that from the view of the aggregate market, the episodes at time $y$ (corresponding to the rows) are being recalled more often when analyzing each firm at current time $x$ (corresponding to the columns). The findings are summarized as following. First, we observe the well-established rules of human memory retrieval - strong recency effect and temporal contiguity[17]. Most of the time, the recent episodes are being focused more frequently, usually the last quarter or the same quarter in the preceding year. And the agent recalls the episodes that occurred contiguously in time to recently-recalled episodes. Second, during the COVID in 2020, the GFC got recalled. Specifically, Figure 5 shows the details of the recalls in 2020. In the second quarter of 2020, the market crash period of the GFC was being recalled. But going to the third quarter of 2020, instead of focusing on the next quarter to the quarter just got recalled in the second quarter of 2020, the aggregate market shifted quickly to look at the recovery period of the GFC, which is intuitive since the economic stimulus policies were enforced starting from the third quarter of 2020. Turning to the fourth quarter of 2020, the GFC was seldom recalled while the market considered the COVID should be unprecedented and concentrated on what happened in the most recent periods. Third, being consistent with the theoretical memory literature (Wachter and Kahana, 2022) and the experience effect literature (Malmendier and Nagel, 2011), we notice that long-term experiences and memories are salient most of the time in some periods, for instance the GFC happened in 2008 was non-trivial to analysts in 2020. The overall pattern of the analysts' recalls is also reconciled with (Jiang et al., 2022) which present that the investors are more likely to recall both the recent and dramatic episodes in a survey.

[17]We observe the pattern that is consistent with the temporal contiguity effect, but we can not conclude the importance of the temporal contiguity effect here. We examine the significance of temporal contiguity effect with a simulation study in Section 6.1.

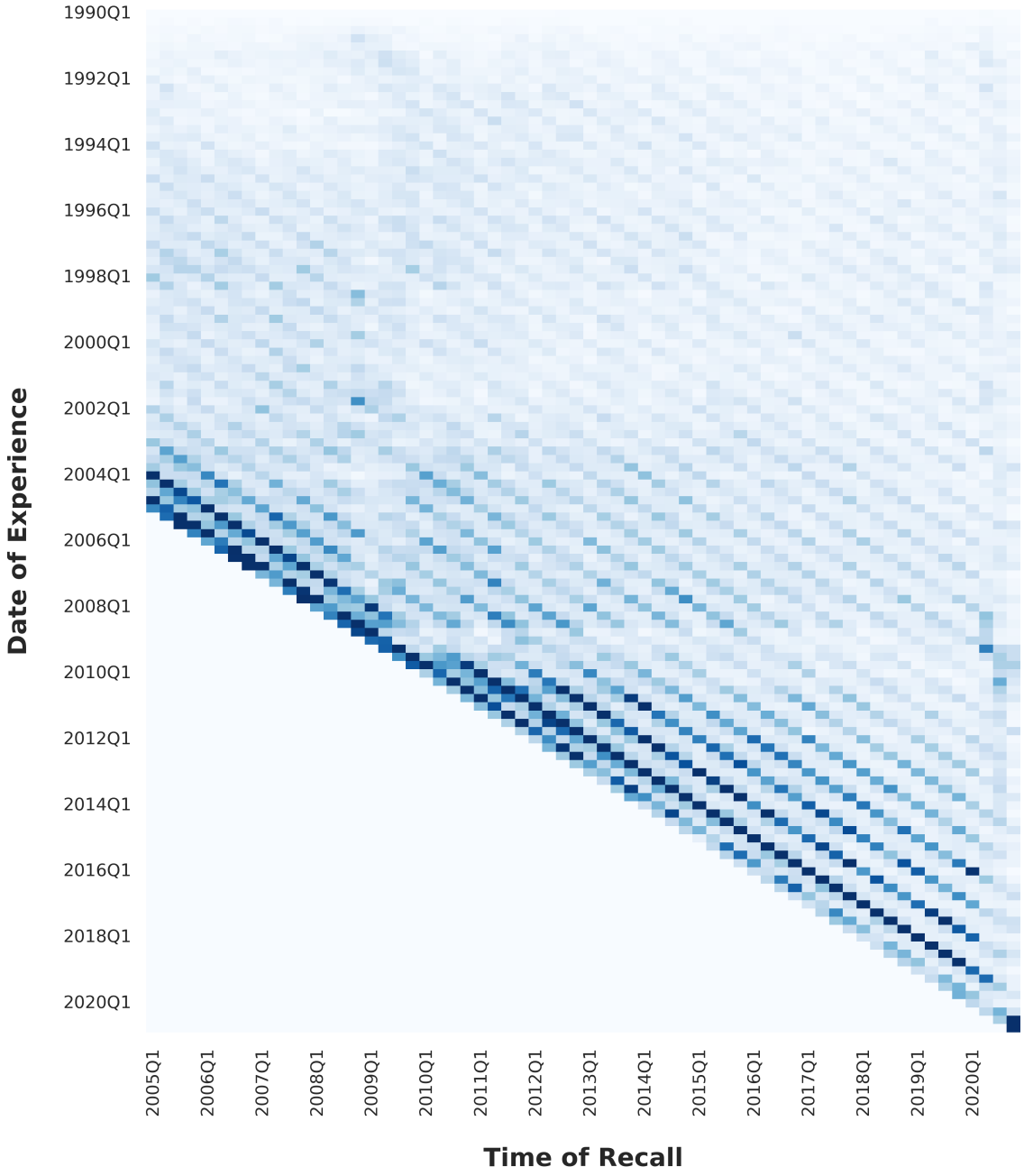



Figure 4: The representative analyst's recalled episodes

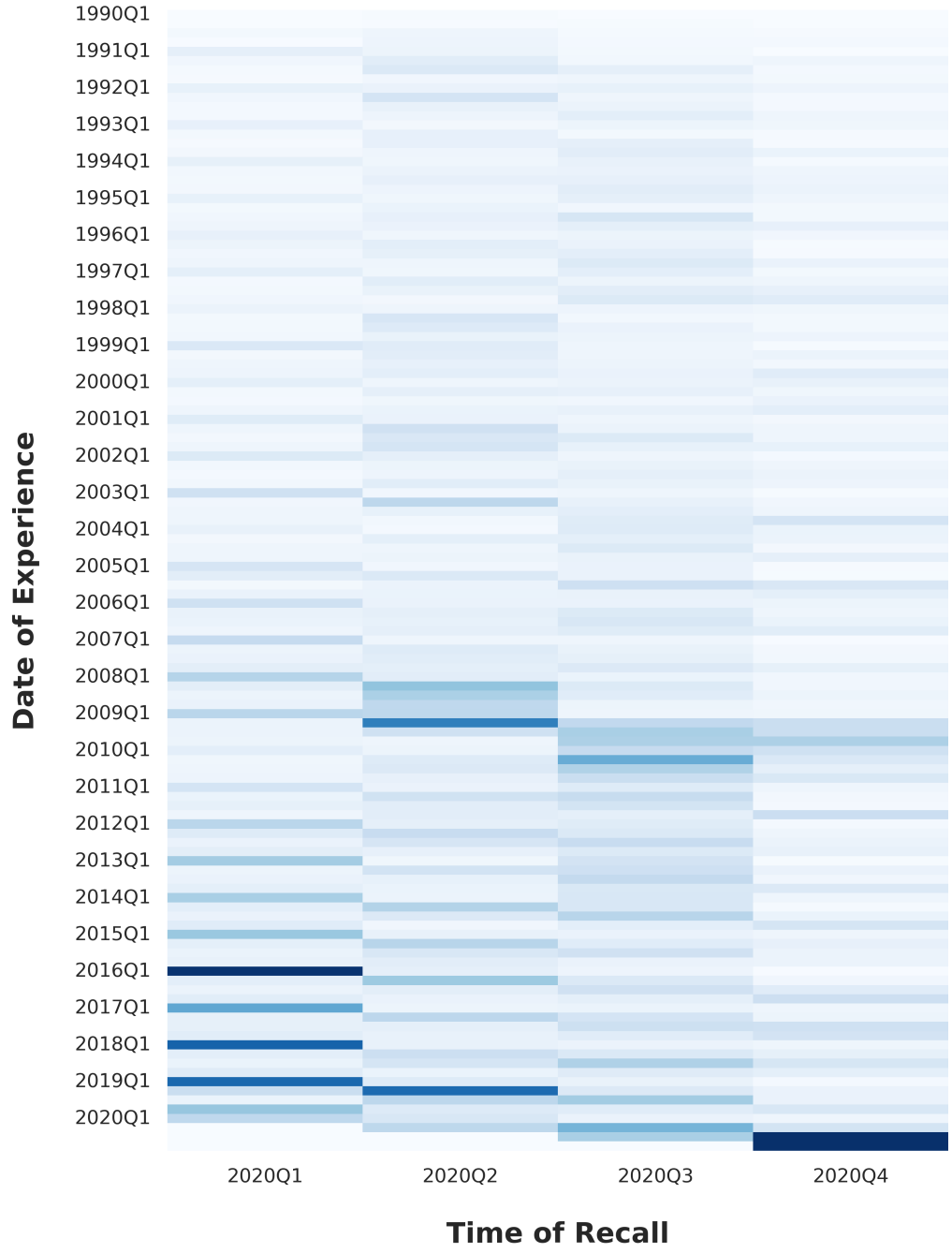



Figure 5: The representative analyst's recalled episodes during COVID

Furthermore, we compute for each period, the frequency of the recalled episodes happened in each of the past 15 years. Figure 6 shows the the average results for all periods during 2005-2020, the average results for periods during 2008 GFC and during the COVID. Being consistent with assumption made in the experience effect literature (Malmendier and Nagel, 2011), on average, the impact of past experiences is decreasing in the number of years before

today. However, such assumption does not hold during certain periods. For example, during the COVID pandemic, the impact of GFC on analysts is salient. Moreover, we can see from Figure 5, and the impact of GFC even dominate the recent episodes during the second and the third quarter of 2020. These results derived by our neuroscientific-founded approach suggest more robust disciplines of modeling the impact of past experiences from different time periods.

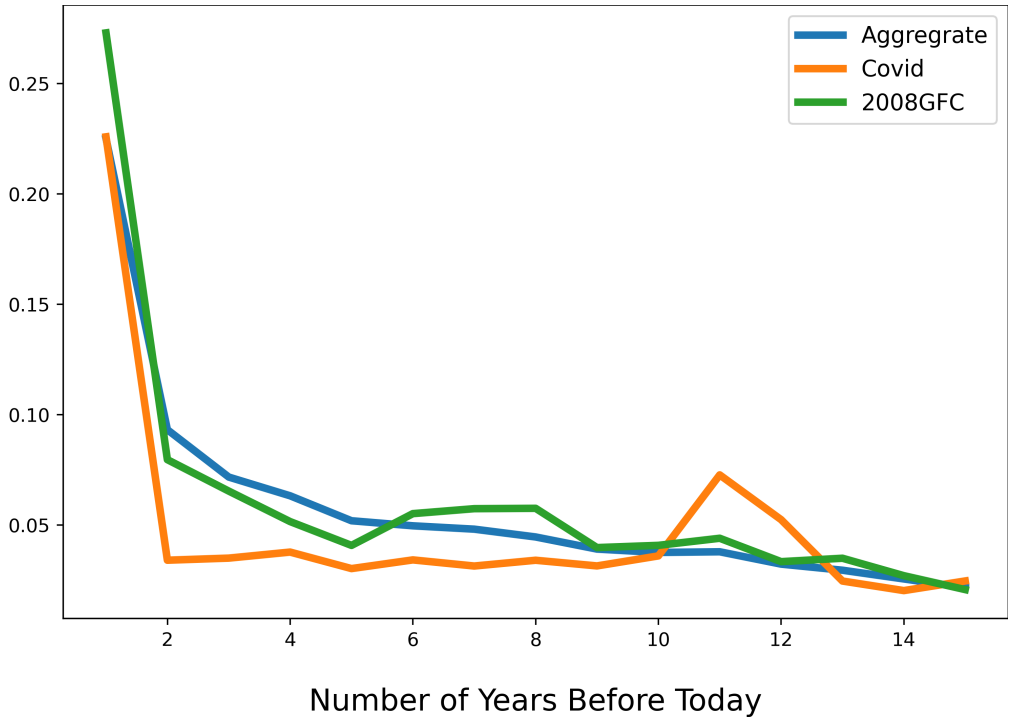



Figure 6: The frequency of recalled episodes happened in each of the past 15 years

## 4.2 Benchmark Recalls and Misleading Recalls

Given the analysts' recalls, we are interested in whether or not, the analysts recall the misleading episodes and thus make wrong forecasting decisions. Making decisions or forming belifs based on the recalled episodes (reference point) itself may be rational. Deviation from full rationality comes from the possibility that the contextual-cued recalls are misleading or distorted, i.e., reminding the agent of the wrong episodes. For example, Walters and Fernbach (2021) survey the investors and find that their memories are biased in the sense that the investors tend to recall their own portfolio returns as higher than achieved, then this leads to overconfidence in their investment decisions. Hence, in this section, we aim

to examine whether the analysts' recalls are biased and the implication of the misleading recalls.

We first need the benchmark recalls, i.e., the episodes that the analysts should have recalled in order to make accurate forecasts. Thus, we let the same memory model that is used to fit the analyst's forecasting decisions re-fit the realized earnings decisions $RE_{j,t}$, and we can get the econometrician's context vector $h^E_{j,t}$. Then, we find the econometrician's (benchmark) recalled episodes $(l^E, \tau^E)$ following the analogous process of memory retrieval for the representative analysts as described in Section 4.1.

Figure 7 presents the benchmark recalls. Compared with the analysts' recalls shown in Figure 4 and Figure 5, the benchmark recalls in each period are more concentrated. In general, the econometrician pays more attention to the recent episodes and most of the experiences happened in the distant past are irrelevant, while the analysts are affected more deeply by long-term memories. Moreover, sometimes the analyst's recalls are obviously wrong compared with the benchmark recalls. For instance, during 2010 to 2014, analysts recalled more episodes related to the end of the boom period, while the benchmark recalls did not fall into that period. Another example is during the COVID pandemic, the machine learning benchmark highly concentrated on the 2008 GFC whereas the analyst's recalls were more distracted.

Compared to the machine learning benchmark, analysts' recency effect is weaker during the regular time but stronger during the crisis time. It suggests that analysts do not fully react to environment changes. During the regular time, environment changes are represented by new input features and therefore the machine learning benchmark rapidly switch to recent similar scenarios. On the contrary, analysts focused on distant episodes which are actually less correlated with current situation. During the crisis time, macroeconomics conditions dramatically change and the machine learning benchmark rapidly changes its focus while analysts still immerse themselves in the old days. Both underreaction phenomenon can be explained by analysts' confirmation bias. During regular time, analysts overvalue their past experience and therefore they recall more distant episodes. On the other hand, they do not believe they need to change too much during the crisis time, resulting more recent recalls compared to the benchmark.

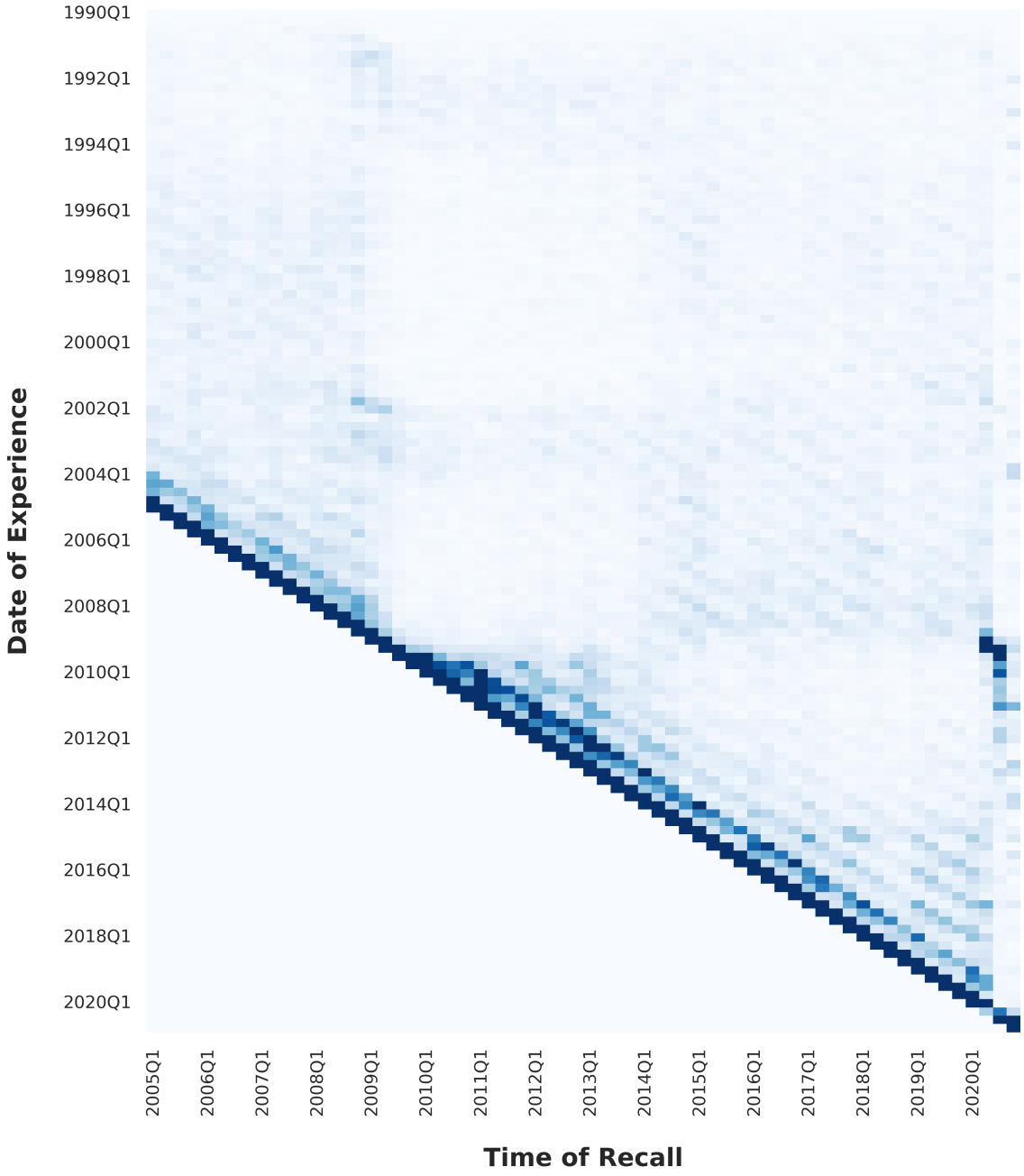



Figure 7: The benchmark recalled episodes

We develop a simple paired $t$-test to formally examine whether the analyst's recalls are different from the benchmark recalls. In other words, we wish to test if the analysts' recall $(l^A, \tau^A)$ is significantly different from the benchmark recall $(l^E, \tau^E)$. However, the distance between $(l^A, \tau^A)$ and $(l^E, \tau^E)$ is not directly measurable. Thus, we project the recalls into the same measurable space—we compare the realized earnings revisions in the recalled episodes. To be specific, for each firm $j$ at each time $t$, let $\left\{(l^A_{j,t,1}, \tau^A_{j,t,1}), (l^A_{j,t,2}, \tau^A_{j,t,2}), ..., (l^A_{j,t,P}, \tau^A_{j,t,P})\right\}$

and $\left\{(l^E_{j,t,1}, \tau^E_{j,t,1}), (l^E_{j,t,2}, \tau^E_{j,t,2}), ..., (l^E_{j,t,P}, \tau^E_{j,t,P})\right\}$ denote the top $P$[18] recalled episodes from the analysts' view and the econometrician's view, respectively (top $P$ solutions that maximize the similarity to $h^A_{j,t}$ and $h^E_{j,t}$, respectively). Further, we define the analysts' recall distortion as:

$$RD_{j,t} = \frac{1}{P}\sum_{p=1}^{P} RE_{l^A_{j,t,p}, \tau^A_{j,t,p}} - \frac{1}{P}\sum_{p=1}^{P} RE_{l^E_{j,t,p}, \tau^E_{j,t,p}}. \tag{4.3}$$

Then the null hypothesis that overall the analyst's recalls are indifferent from the benchmark recalls is equivalent to that the mean of the recall distortion is zero, $\overline{RD} = 0$. The simple paired $t$-test shows that the $t$-statistics is -3.64[19], which indicates that overall, the analysts' recalls are significantly distorted from the benchmark recalls. Put differently, analysts are making mistakes in choosing the "right analogies."

## 4.3 Recall Distortion and Forecast Errors

Given that analysts may choose the "wrong analogies," we next demonstrate the implication of the misleading recalls. Specifically, we show that the recall distortion could help explain and predict the analysts' forecast errors.

The forecast error of firm $j$ in month $t$ is defined as:

$$FE_{j,t} = AF_{j,t} - RE_{j,t}.$$

Since both $AF_{j,t}$ and $RE_{j,t}$ take the value of -1, 0, or 1, we have that $FE_{j,t} \in \{-2, -1, 0, 1, 2\}$. That is, the forecast error is zero, when the analysts make the right revision decision; and the forecast error will reach the maximum, when the analysts make the opposite revision decision, e.g. the analysts revise up the forecast while the realized earning is actually smaller than the analysts' consensus forecast from last period.

We then perform the test nested in the following predictive regression:

$$FE_{j,t} = \beta \times RD_{j,t} + \varepsilon_{j,t}, \tag{4.4}$$

in which we also include several fixed effects, e.g., month fixed effects, firm fixed effects, and industry and month cross fixed effects. Note that, although the time subscript for the regressor $RD$ is the same as the one for the dependent variable, we still perform the analysis in a predictive regression fashion. That is because, $RD_{j,t}$ is constructed using all the available information before time $t$ as the model inputs $X_{j,t}$ are the public signals that were announced

[18] In the analysis shown below, we set $P = 10$, but the results are robust to the choice of $P$.

[19] The standard errors are clustered on both industry and year dimension.

before time $t$ and the model is trained using all the data points before time $t$ as well, while $FE_{j,t}$ is available in month $t$ or even after $t$. In this first regression, we are interested in whether $\beta$ is significantly positive. The intuition of a positive $\beta$ is that the analysts are making over-optimistic (over-pessimistic) forecasts when recalling over-optimistic (over-pessimistic) episodes.

We also study the case where we take the absolute values for both the regressor $RD$ and dependent variable $FE$ :

$$|FE_{j,t}| = \beta \times |RD_{j,t}| + \varepsilon_{j,t}, \tag{4.5}$$

in this sense, we only focus on the magnitude and explore whether the severer recall distortion predicts larger forecast errors.

Another way to study the relationship between $FE_{j,t}$ and $RD_{j,t}$ is to use the ordered logit model in reason that the regressor $FE_{j,t}$ is a ordinal variable with 5 different values:

$$\begin{aligned} FE^*_{j,t} &= \beta \times RD_{j,t} + \varepsilon_{j,t} \\ FE_{j,t} &= \begin{cases} -2 & FE^*_{j,t} \neq \mu_{-2} \\ -1 & \mu_{-1} < FE^*_{j,t} \neq \mu_{-2} \\ 0 & \mu_0 < FE^*_{j,t} \neq \mu_{-1} \\ 1 & \mu_0 < FE^*_{j,t} \neq \mu_1 \\ 2 & \mu_1 < FE^*_{j,t} \end{cases} \end{aligned} \tag{4.6}$$

A key assumption of the ordered logit model is the odd ratio ($OR$) keeps constant for any value, i.e.

$$\frac{Prob(FE_{j,t} > k+1)}{Prob(FE_{j,t} > k)} = OR \quad \forall\ k < 2.$$

Table 3 reports the estimation results of regression (4.4), (4.5), and (4.6). In both regressions, we find a strongly significant positive relation between recall distortion and the forecast error. The size of the association barely changes when we control for several fixed effects. Therefore, we claim that, compared with the machine leaning benchmark, when analysts recall over-optimistic (over-pessimistic) episodes, they tend to make over-optimistic (over-pessimistic) forecasts as well. Besides, when the analysts' recalled episodes are distorted far away from the benchmark recalls, their forecast errors are larger. Column (3) and (6) perform the ordered logit regression and we can calculate the economic magnitude. 10% increase in the level of recall distortion lead to $7.88\%(= e^{0.1\times 0.759} - 1)$ probability increases of jumping to the relatively more optimistic level. Similarly, 10% increase in the absolute value of recall distortion lead to $1.08\%(= e^{0.1\times 0.108} - 1)$ probability increases of

jumping to the next level representing larger errors. In a nutshell, through this memory test, we provide a memory interpretation of the forecast errors - empirically what analysts have recalled could help explain their misbehavior and deviated expectations. The above results are based on the similarity measure (4.2), for robustness check, we also perform the analogous regression analysis based on the similarity measure (4.2), the corresponding results are reported in Appendix A.2. Since the machine learning model training process involves randomness, we also perform the robustness check for 100 different random seeds, and the results are shown in Appendix A.4.

Table 3: Analysts' recall distortion and forecast errors

| | Forecast Error ($FE_{j,t}$) | | | Forecast Error ($\|FE_{j,t}\|$) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Linear | Linear | Ologit | Linear | Linear | Ologit |
| Recall Distortion $RD_{j,t}$ | 0.293*** | 0.283*** | 0.759*** | | | |
| | (0.013) | (0.014) | (0.021)) | | | |
| Recall Distortion $\|RD_{j,t}\|$ | | | | 0.043*** | 0.042*** | 0.108*** |
| | | | | (0.006) | (0.007) | (0.019) |
| Observations | 277,487 | 277,245 | 277,487 | 277,487 | 277,245 | 277,487 |
| Month fixed effects | Yes | No | No | Yes | No | No |
| Firm fixed effects | Yes | Yes | No | Yes | Yes | No |
| Industry*Month fixed effects | No | Yes | No | No | Yes | No |

This table presents results for regressions of the form

$$FE_{j,t} = \beta \times RD_{j,t} + \varepsilon_{j,t},$$

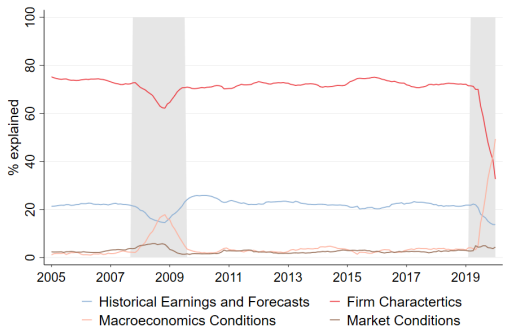
where the dependent variable is the analyst (consensus) forecast revision error for firm $j$ at time $t$ and $FE_{j,t} \in \{-2, -1, 0, 1, 2\}$, the independent variable is the analyst's recall distortion which are predicted by the model with all the information available before time $t$ as defined in Equation (4.3). Recalls are found based on the similarity measure shown in (4.1). Column (4)-(6) report the results that take the absolute value of both the dependent variable and the independent variable. Column (3) and (6) shows the results of the ordered logit model while the rest of the columns are associated with the linear regression model. Data are from the period 2005 to 2020. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

# 5 Context

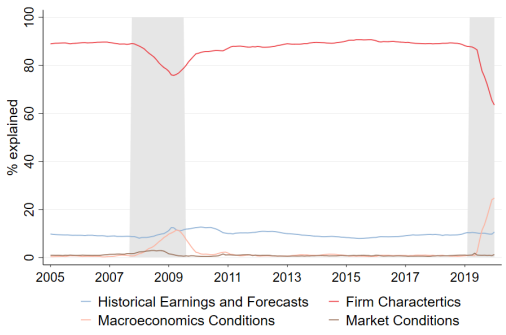
## 5.1 Decomposition of Contexts and Memories

Memory and context play the essential roles in memory models such as serving as the cue for recalls, getting better understanding of how memory and context actually look like from the field could gain us the insight of how the underlying memory mechanisms work out and guide the future memory modelling.

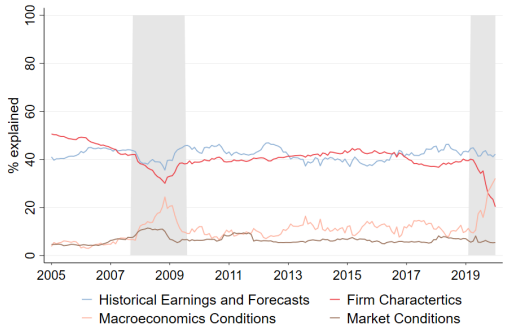
Both memory cells and the mental contexts are latent in the LSTM, therefore in this section, we provide the interpretation of the memory and context vectors by showing how the variable importance of the memory and context vary over time. The concept of variable importance is analogous to the notion as shown in Gu, Kelly, and Xiu (2020) and Kelly, Pruitt, and Su (2019). The linear variable importance of covariate $j$, $VI_j$ is approximated by the average reduction in panel $R^2$ of linearly regressing each of $K$ dimension of the extracted memory ($c$) and context ($h$) vectors on all input features ($X$) from setting all values of the covariate $j$ to zero, while holding the remaining model estimates fixed. Then, to better illustrate the time series of variable importance, we perform the linear regression based on the 1-year rolling window, and put all 79 covariates into four groups. Each of the four groups contains the firm characteristics (Part 1 and Part 2 of Table A1 except for the monthly stock return and P/E ratio related variables such as pe_exi, pe_inc, peg_trailing, capei), macroeconomics conditions (Part 3 of Table A1), market conditions (the firms' stock return over the previous month) and the historical earnings and forecasts (Part 4 of Table A1 and pe_exi, pe_inc, peg_trailing, capei), respectively.



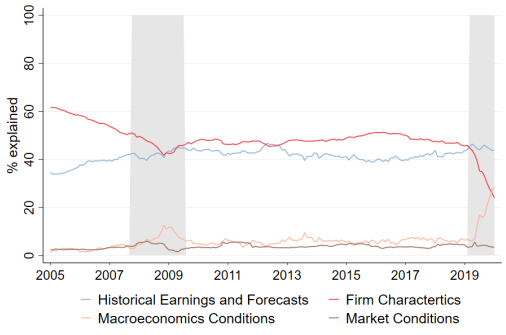

(a) Benchmark - Context




(b) Benchmark - Memory




(c) Analyst - Context




(d) Analyst - Memory

Figure 8: Decomposition of Context and Memory Vectors

Figures 8a - 8d demonstrate the time-series plot of the group variable importance of the representative analyst's and the machine learning benchmark context and memory vectors. Since different groups consist of different number of variables, we do not over-interpret the difference between group variable importance within each figure and at each point of time. However, we focus on the time-series variation in the variable importance and different patterns between figures. The findings are three-fold. First, the linear variable importance is time-varying. Regardless of analysts or benchmark, memory or context, when the economy is in recession, macroeconomic variables become more essential but firm fundamentals and historical earnings-related variables are less important. This pattern matches the intuition that the recession has the market-wide and systematic impact. The finding also provides evidence for the prediction by the theory of limited attention (Kacperczyk, Van Nieuwerburgh, and Veldkamp, 2016) which states that investors would allocate more attention to aggregate news during recession[20]. Second, we observe that context is more volatile than memory. This coincides with the interpretation that the context is considered as the short-term generalized information while the memory is taken as the long-term generalized information. Third, the analyst and the machine learning benchmark focus on different aspects, i.e., for the benchmark, firm fundamentals play the most important role in both context and memory, but for the analyst, firm fundamental and the historical earnings-related variables are almost equally weighted. The potential implications and explanations of different focuses of the analyst and the benchmark may include that analysts are subject to the encoding errors (early noise by Woodford, 2020), self-herding bias (Hirshleifer et al., 2019) as analysts overly focus on their past decisions and limited attention (Hirshleifer and Teoh, 2003). The difference between the decomposition of representative analyst's and the machine learning benchmark context and memory indicates that the analysts have distorted recalls due to the fact that they process the information experienced in their memory differently from the benchmark.

## 5.2 Experiences, Contexts, and Forecast Dispersion

To illustrate the importance and validity of mental context and memory in describing agents' information processing procedure, we systematically examine whether different mental contexts brought by analysts' different past experiences lead to dispersion in the forecasting decisions in a micro-founded way.

[20] Kwan, Liu, and Matthies (2022) show that institutional investors pay their attention to aggregate news during economic downturns with data on daily internet news reading.

A voluminous literature shows that financial market participants' different past experiences cause various financial decisions. For example, Malmendier and Nagel (2016) suggests that different experiences in inflation will significantly change investors' inflation expectation. In addition, being specific to analysts, Bradley, Gokkaya, and Liu (2017) and Hirshleifer et al. (2021) state that pre-analyst experiences in certain industry and analysts' first impressions indicate different forecast performances, respectively. Building on this literature but being different from previous work that examines the low-dimensional experiences in the reduced-form, in this section we present a systematic and comprehensive way of studying experience effect, in the sense that we show how to characterize the high-dimensional experiences within a structural and neuroscientific-founded model.

Specifically, for each firm $j$ at each time $t$, we are interested in whether variations in the model-induced context vectors for all the analysts who are issuing forecasts for this firm at this moment, can explain and predict the dispersion in their forecasting decisions.

Following Diether, Malloy, and Scherbina (2002), we define the forecast dispersion as the standard deviation of analysts' forecast revision decisions[21], that is

$$AF\_Dis_{j,t} = \sqrt{\frac{\sum_{i=1}^{N_{j,t}} \left(AF_{i,j,t} - \mu_{j,t}^{AF}\right)^2}{N_{j,t} - 1}},$$

where $N_{j,t}$ is the number of analysts who are making forecasts for firm $j$ in month $t$, and $\mu_{j,t}^{AF} = \frac{1}{N_{j,t}} \sum_{i=1}^{N_{j,t}} AF_{i,j,t}$ is the mean forecast revision decision[22].

The context dispersion is computed from $h_{i,j,t}$ which is the analyst $i$'s context vector for firm $j$ in month $t$. Thus, in order to define the context dispersion, we need to trace out the latent $h_{i,j,t}$ from the model. We first define that the individual analyst's covering experience starts from the date $s_{i,j}$ (analyst $i$'s first time covering the firm) up to date. Then we can compute $h_{i,j,t} = LSTM(X_{j,s_{i,j}}, X_{j,s_{i,j}+1}, ..., X_{j,t})$ from our well-trained model. Next, the context dispersion $Context\_Dis_{j,t}$ is measured as the total variation[23], which is the trace of the variance-covariance matrix of vectors $\{h_{i,j,t}\}_{i=1}^{N_{j,t}}$. The variance-covariance matrix is $K \times K$, where $K$ is the dimension of the vector $h_{i,j,t}$.

[21] In Diether, Malloy, and Scherbina (2002), the dispersion is defined as the standard deviation of earnings forecast divided by the absolute value of the mean earning forecast. Here, we are not scaling the standard deviation by the absolute value of the mean because the analysts' forecast decisions under this paper's setting have already been scaled as they are defined as the classification choices.

[22] We only keep the firms that are covered by two or more analysts. For the multiple forecasts made by the same analyst in the same month, we only keep the last one.

[23] The usual dispersion measure of multivariate variables is the generalized variance Wilks (1932), which is the determinant of the variance-covariance matrix. However, in this paper, the determinant is usually just zero since in many cases, $N_{j,t} < K$ and it leads to the singular variance-covariance matrix.

To further demonstrate that context is a better proxy for the experiences, we include the length of covering experiences as another proxy. The length of covering experiences is defined as

$$Length\_Cov\_Exp_{i,j,t} = t - s_{i,j} + 1.$$

Then we take the standard deviation of the lengths of covering experiences for firm $j$ at time $t$ as another dispersion measure $Length\_Dis_{j,t}$.

We perform the analysis based on the following predictive regression:

$$AF\_Dis_{j,t} = \beta_c \times Context\_Dis_{j,t} + \beta_l \times Length\_Dis_{j,t} + \varepsilon_{j,t}, \tag{5.1}$$

in which we also include the firm and month time fixed effect. Regression (5.1) is the predictive regression in terms of the interested variable $Context\_Dis_{j,t}$, since $Context\_Dis_{j,t}$ is computed from $h_{i,j,t}$ which is derived from the information available before time $t$ and the model trained with the data before time $t$, while $AF\_Dis_{j,t}$ is the dispersion of the realized analysts' forecast decisions that are available in month $t$.

Table 4: Dispersion of Covering Lengths, Contexts, and Analysts Forecast Revisions

| | *AF_Dis* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| *Context_Dis* | 0.225*** | 0.264*** | 0.202*** | | | | 0.226*** | 0.268*** | 0.209*** |
| | (0.029) | (0.029) | (0.029) | | | | (0.029) | (0.030) | (0.030) |
| *Length_Dis* | | | | 0.031 | -0.179 | -0.325*** | -0.008 | -0.234 | -0.382*** |
| | | | | (0.109) | (0.140) | (0.049) | (0.109) | (0.142) | (0.052) |
| Observations | 179,956 | 179,911 | 179,911 | 179,956 | 179,911 | 179,911 | 179,956 | 179,911 | 179,911 |
| Month fixed effects | Yes | No | Yes | Yes | No | Yes | Yes | No | Yes |
| Firm fixed effects | No | Yes | Yes | No | Yes | Yes | No | Yes | Yes |

This table presents results for regressions of the form

$$AF\_Dis_{j,t} = \beta_c \times Context\_Dis_{j,t} + \beta_l \times Length\_Dis_{j,t} + \varepsilon_{j,t},$$

where the dependent variable is the dispersion (standard deviation) of analysts' forecast revisions for firm $j$ at time $t$, the first independent variable is the dispersion (trace of the variance-covariance matrix) of analysts' mental contexts $h_{i,j,t}$ which are predicted by the model with all the information available before time $t$, and the second independent variable $Length\_Dis_{j,t}$ is the dispersion (standard deviation) of lengths of the analysts' covering experiences. Data are from the period 2005 to 2020. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

Table 4 reports the significantly positive relation between the dispersion in contexts and the dispersion in analysts' forecasting decisions. That is to say, the dispersion in analysts' mental contexts induced by different covering experiences could explain and predict the dispersion in their final forecasting decisions. However, without the memory model and the context, the dispersion of past experiences which is represented by the standard deviation of the lengths of the covering experience could not significantly explain the analysts' forecast dispersion, and the coefficients even become negative when including multiple fixed effects.

The results associated with the length of covering experiences are counter-intuitive which state that when analysts have more dispersed experiences they would make more consistent forecasts, which are in contrast with the experience effect (Malmendier and Nagel, 2011). The results indicate that the memory-model-induced mental contexts are efficient in characterizing generalized information that are applied to decision-making or expectation formation from different personal experiences.

For robustness check, we also generate another memory dispersion measure based on cosine-distance. The definition and the corresponding estimation results are shown in Appendix A.3. We also perform the robustness check for 100 different random seeds to rule out the possibility that the results are led by the randomness in the training process and the results are shown in Appendix A.4.

# 6 Role of Specific Memory Channels

LSTM is a comprehensive memory model that contains multiple memory channels. In this section, we examine the contribution of the specific memory feature and channel - temporal contiguity and selective forgetting, to analysts' belief formation processes.

## 6.1 Temporal Contiguity

Compared to other models, LSTM has a unique feature that it stores the long-term memory. Then combined with the autoregressive structure of mental context, LSTM has the channel of displaying temporal contiguity - one of the fundamental principles of human memory system. Temporal contiguity refers to the evidence that when people recall an event, they also tend to recall other temporally successive events. In the psychology experiments, Kahana (1996) first details the tendency that after an item is recalled from a specific serial position, the item recalled next mostly comes from a neighboring serial position. Two properties of the temporal contiguity effect are also documented: the forward asymmetry that it is more likely to make next recalls in the forward direction than in the backward direction, and invariant across time scales in the sense that contiguity effect is significant even for recalled events in the distant past (that's also one of the reasons why we need to incorporate long-term memory). Beyond the discussion in psychology literature, Wachter and Kahana (2022) illustrate the temporal contiguity effect in a theoretical financial setting, but without presenting empirical

results of the importance and existence of temporal contiguity[24]. In this section, we step further and provide empirical evidence to show the significance of temporal contiguity effect in analyst's belief formation processes.

To clearly show empirical evidence of temporal contiguity, we need to distinguish the temporal contiguity effect from the nature of similarity between two adjacent vectors of economic and financial variables in time. Hence, we design a simulation study to avoid the correlation among input features. But the simulation is based on the empirically estimated model from fitting the representative analyst's forecasting processes (the same recursively trained model shown in Section 3.2), in order to emphasize that the findings from the simulation study directly reflect the analyst's belief formation processes.

The design of the simulation study is as following. In each simulation, first, we generate a set of input vectors:

$$X_t^{\text{sim}} = \mathbb{E}[X] + 10u_t \times \sigma(X), \quad t = 1, 2, ..., T,$$

where $\mathbb{E}[X]$ and $\sigma(X)$ represent the time-series mean, and the standard deviation (the square root of diagonal elements of the covariance matrix) of the original input features $X$, respectively. And $\times$ is element-wise product operator. $u_t$ are randomly drawn and mutually orthogonal vectors whose $L^2$ norms are 1. The inner product of any two simulated input vectors $X_t^{\text{sim}}$ is a fixed number (the square of the $L^2$ norm of the feature expected value $\|\mathbb{E}[X]\|^2$), in this way we maximally exclude the correlation between input features and keep the simulated inputs close to the true distribution (the multiplier 10 on $u_t$ also serves for this purpose)[25]. We simulate $T = 70$ periods for $X_t^{\text{sim}}$[26].

Second, in period $T+1$, we duplicate the simulated input vector in period $\tau$, i.e., $X_{T+1}^{\text{sim}} = X_\tau^{\text{sim}}$. $\tau$ is randomly selected from the interval $(10, T-10)$. The first 10 periods and the last

[24] Wachter and Kahana (2022) provide a memory explanation for the narratives that depression would come right after seeing the financial crisis. Specifically, they show that in agents memory, the Great Depression in 1930 came right after the stock market crash of 1929. Then the re-appearance of a financial crisis today retrieves their memory on crisis in 1929, as well as the memory on the depression since the state of financial crises and depressions are associated in time in their memory, even the features of crisis and depression are assumed orthogonal. The temporal contiguity effect elicits all events happened around 1929.

[25] Our simulation results are robust with the selection of the inner product. If simulated input vectors are completely orthogonal to each other (the inner products are equal to 0), i.e.,

$$X_t^{\text{sim}} = 10u_t,$$

the temporal contiguity effect still prevails.

[26] Since the dimension of the original input features $X$ is 79, we can only maximally generate 79 mutually orthogonal vectors.

10 periods are excluded to avoid the potential impacts of the primacy effect and the recency effect.

Third, we use the empirically estimated model to simulate all mental contexts $\{h_t^{\text{sim}}\}_{t=1}^{T+1}$ according to the set of simulated input features $\{X_t^{\text{sim}}\}_{t=1}^{T+1}$, then use $h_{T+1}$ as the cue to search for the recalls. Next, we study the similarity around $\tau$ to examine the temporal contiguity[27]. We use the similarity of mental contexts $Similarity(h_{\tau+l}, h_{T+1})$ defined in (4.1) as a measure of the likelihood that the episode in period $\tau + l$ is recalled according to the cue of $h_{T+1}$.

If the temporal contiguity effect exists, it is expected to find similarity is decreasing in the distance to $K$ with a forward asymmetry. In other words, temporal contiguity implies the following conditions should hold:

$$\begin{aligned}
&\forall\ 0 < i < j, Similarity(h_{\tau+i}, h_{T+1}) > Similarity(h_{\tau+j}, h_{T+1}),\\
&\forall\ 0 < i < j, Similarity(h_{\tau-i}, h_{T+1}) > Similarity(h_{\tau-j}, h_{T+1}),\\
&\forall\ i > 0, Similarity(h_{\tau+i}, h_{T+1}) > Similarity(h_{\tau+i}, h_{T+1}).
\end{aligned} \tag{6.1}$$

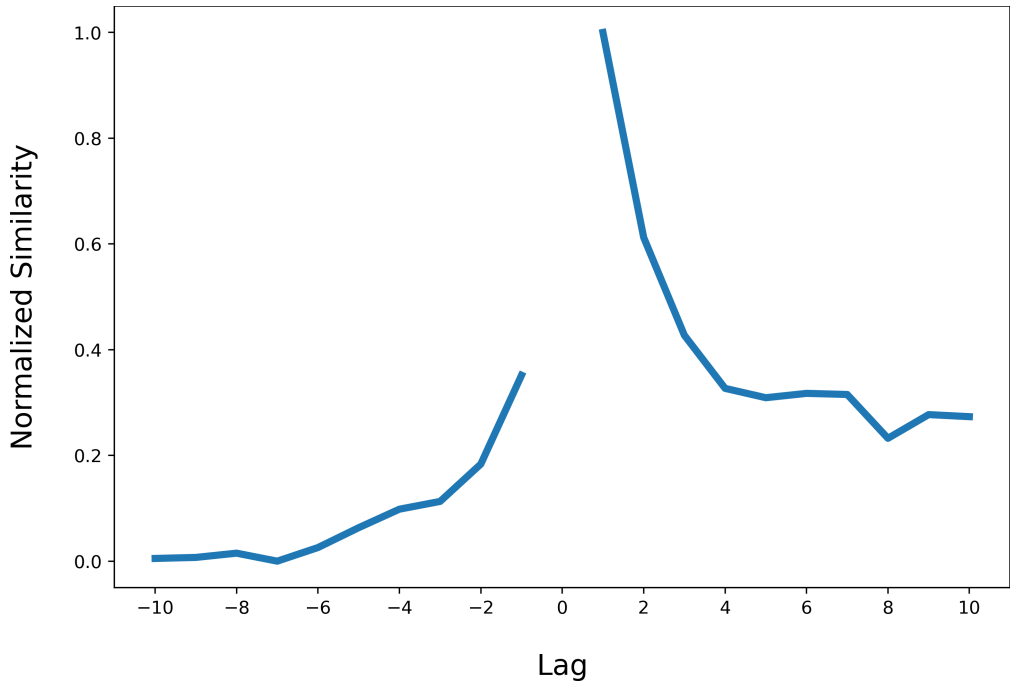



Figure 9: Temporal Contiguity of LSTM

Figure 9 displays the estimated average temporal contiguity effect of our trained LSTM model using simulated data. The simulation is implemented 10000 times. $Similarity(h_\tau, h_{T+1})$

[27]The process is similar to the standard memory experiment that the participants are shown $T$ different words successively, and then asked to recall one of the word (the $\tau$-th) they have studied. Next, the participants are asked to make free recalls. See for example, Healey, Long, and Kahana (2019)

is abstracted from the figure as it is the highest (this confirms the model's ability of finding the right recall) and irrelevant from the examination of temporal contiguity. The $X$-axis stands for $l$ which is the number of positive or negative lag on $\tau$, while the $Y$-axis shows $Similarity(h_{\tau+l}, h_{T+1})$ which are min-max normalized and the figure indicates that the highest similarity appears right after recall ($l = 1$) and the similarity decreases with the absolute value of the lag, while the overall similarity is smaller with negative lags than positive lags. In general, the similarity pattern shown in Figure 9 is consistent with the conditions (6.1) implied by temporal contiguity. This confirms that our LSTM model can not only theoretically but also empirically display the fundamental memory principal - temporal contiguity, and the temporal contiguity effect is indeed essential in modeling analysts' belief formation processes[28]. On the other hand, the findings indicate the need of a memory model like LSTM which can implement temporal contiguity to describe the analysts' belief formation processes.

## 6.2 Selective Forgetting

In this part, we break down the forget gate in the full LSTM to understand the role of forgetting in the analysts' and the econometrician's expectation formation process, in the spirit of performing counterfactual analysis. Literature has shown that forgetting affects agents' decision making. For example, Walters and Fernbach (2021) argue that selective forgetting[29] as a memory bias is presented among investors which is associated with misleading recalls and then leads to investors' overconfidence. In what next, we demonstrate what episodes in the representative analyst's and the econometrician's memory are fading away, and those are consistent with the discrepancy between their recalls.

In the full LSTM, the forget gate chooses the content and extent to erase from the existing memory cell. Now we remove the forget gate and the structure of the new memory model

[28] We argued that LSTM contains the channels of long-term memory and autoregressive context structure, and the two channels should jointly work to generate temporal contiguity. But the two channels are not necessarily salient, i.e., the two channels can diminish according to different model parameters and the empirical data the model fits in. For example, if the model parameters make forget gate always empty the memory cell in the sense that the channel of long-term memory is blocked, then LSTM collapses to RNN, which we show in Appendix A.5 does not produce temporal contiguity. Thus, employing the empirically estimated model in the simulation study indicates that actually two channels are both significant and generate temporal contiguity in analysts' forecasting processes.

[29] In their study, selective forgetting refers to that participants are more likely to readily forget consequential losing trades than the consequential winning trades, then less likely to recall the losing trades than the winning trades.

is showing in Figure 10. In this new memory model, the memory cell is only updated by inputting new information and without selective forgetting:

$$c_t = input_t \times \tilde{c}_t.$$

Next, we fist re-estimate the new model, then extract the latent context vectors and get the analyst's and the benchmark recalls according to this new memory model. The process is analogous to what is described in Section 4. It means that agents are no longer able to selectively forget any information. But it does not mean that agents can have complete knowledge of past experience. They are still subject to the limited space of memory or mental context: new information flushes in and old information shrinks and squeezes. selective forgetting is not compatible with this setting.

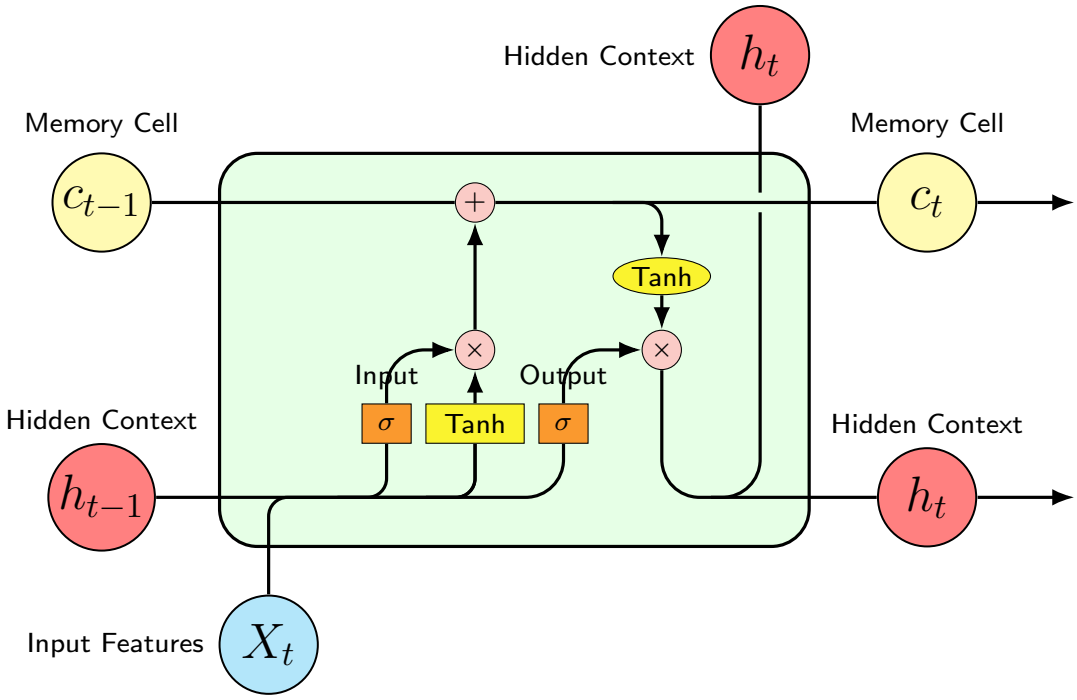



Figure 10: The Structure of LSTM without Forget Gate

To better illustrate the impact of the forget gate and the difference it brings in, we first define the magnitude of the recency effect. For each month, we compute the percentage of recalls which are the episodes that happened within the past three years[30]. Then we average across months (from January 2005 to December 2020), and get the overall magnitude of the

[30]We can alternatively define "recency" as past one year or five years, but the following patterns and conclusions remain unchanged.

recency effect:

$$Magnitude\_Recency = \frac{1}{T}\sum_{t}\frac{\#\ of\ month\ t'\ recalled\ episodes\ within\ the\ last\ three\ years}{\#\ of\ month\ t'\ total\ recalls}.$$

Table 5 reports the magnitude of the recency effect for the analysts and the benchmark under the full LSTM and LSTM without forget gate.

Table 5: The overall magnitude of the recency effect with and without forgetting mechanism

| | Analyst | Benchmark |
|---|---|---|
| LSTM - Full | 38.99% | 75.97% |
| LSTM - No Forget | 50.70% | 65.23% |

This table presents results for the overall magnitude of the recency effect for the representative analyst and the machine learning benchmark under the full LSTM or the LSTM without forget gate. The recency effect is defined as the average percentage of recalls which are the episodes that happened within the past three years.

The recency effect for analysts increases when forget mechanism is blocked while the recency effect for the machine learning benchmark decreases when the same intervention is implemented. An explanation is the machine learning benchmark is able to selectively forget the past experience in an optimal way and blocking the forgetting mechanism makes it wrongly recall the distant episodes which should be forgotten. On the other hand, a different story happens when analysts' forgetting mechanism is blocked. It is possible that analysts selectively forget all other scenarios but misleading images in their memory. When they are able to keep a full view of the past and realize that the distant experience actually does not fit the current situation, they recall recent experiences more often and the recency effect of analysts gets increased. The difference between contents that the analysts and the machine learning benchmark selectively forget provides an interpretation of analysts' recall distortion relative to the machine learning benchmark.

The results are also reconciled with the findings in De La O and Myers (2022) that the analysts consider the world more stationary than it truly is[31]. Then it is reasonable that analysts choose not to forget and rely too much on past experiences to extrapolate and form their expectation today.

[31] De La O and Myers (2022) show that a structural model of fundamental extrapolation where agents form their expectation from a slowly adjusted weighted sum of current and past realizations could replicate the empirical evidence accurately.

# 7 Conclusion

In this paper, we present a novel approach to shed light on the impact of memory on financial market participants' belief formation processes. We extract the analysts' recalls and mental contexts by adapting the machine learning memory model - LSTM. We also provide a machine learning benchmark for the recalls and mental contexts to examine whether the analysts' memory deviates them away from the full rationality. Our analysis shows that analysts' recalls are significantly distorted from the benchmark, which contributes to their forecast errors. Such recall distortion can be explained by the evidence that analysts' mental context are mainly influenced by past earnings and forecasting decisions, rather than current firm fundamentals, and that analysts tend to selectively ignore recent episodes.

Our detailed investigation of recalls and mental contexts offers new insights for theoretical modeling and empirical research on memory in financial markets. Additionally, our approach is well-suited for complex real-world scenarios where agents confront high-dimensional and non-stationary conditions with non-linear interactions between variables, and it can be readily customized to explore the influence of memory in other settings. Our application also highlights the potential of machine learning techniques in analyzing economic agents' behavior in the age of big data.

# Appendix

## A.1 Input Features

The following table reports all 79 input features applied in this paper. For detailed data processing (e.g. imputation), one can refer to van Binsbergen, Han, and Lopez-Lira (2020).

Table A1: Input Features

Part 1. Firm Fundamentals−WRDS Financial Ratios

| Variable | Definition | Variable | Definition |
|---|---|---|---|
| Accrual | Accruals/Average Assets | adv_sale | Advertising Expenses/Sales |
| aftret_eq | After-tax Return on Average Common Equity | aftret_equity | After-tax Return on Total Stockholders Equity |
| aftret_invcapx | After-tax Return on Invested Capital | at_turn | Asset turnover |
| bm | Book/Market | capei | Shillers Cyclically Adjusted P/E ratio |
| capital_ratio | Capitalization Ratio | cash_debt | Cash Flow/Total Debt |
| cash_lt | Cash Balance/Total Liabilities | cash_ratio | Cash Ratio |
| cfm | Cash Flow Margin | curr_debt | Current Liabilities/Total Liabilities |
| curr_ratio | Current Ratio | debt_asset | Total Debt/Total Assets |
| debt_at | Total Debt/Total Assets | debt_capital | Total Debt/Capital |
| debt_ebitda | Total Debt/EBITDA | debt_invcap | Long-term Debt/Invested Capital |
| divyield | Dividend Yield | dltt_be | Long-term Debt/Book Equity |
| dpr | Dividend Payout Ratio | efftax | Effective Tax Rate |
| equity_invcap | Common Equity/Invested Capital | evm | Enterprise Value Multiple |
| fcf_ocf | Free Cash Flow/Operating Cash Flow | gpm | Gross Profit Margin |
| GProf | Gross Profit/Total Assets | int_debt | Interest/Average Long-term Debt |
| int_totdebt | Interest/Average Total Debt | intcov | After-tax Interest Coverage |
| intcov_ratio | Interest Coverage Ratio | inv_turn | Inventory Turnover |
| invt_act | Inventory/Current Assets | lt_ppent | Total Liabilities/Total Tangible Assets |
| npm | Net Profit Margin | ocf_lct | Operating CF/Current Liabilities |
| opmad | Operating Profit Margin After Depreciation | opmbd | Operating Profit Margin Before Depreciation |
| pay_turn | Payables Turnover | pcf | Price/Cash flow |
| pe_exi | P/E (Diluted, Excl. EI) | pe_inc | P/E (Diluted, Incl. EI) |
| PEG_trailing | Trailing P/E to Growth ratio | pretret_earnat | Pre-tax Return on Total Earning Assets |
| pretret_noa | Pre-tax return on Net Operating Assets | profit_lct | Profit Before Depreciation/Current Liabilities |
| ps | Price/Sales | ptb | Price/Book |
| ptpm | Pre-tax Profit Margin | quick_ratio | Quick Ratio (Acid Test) |
| RD_SALE | Research and Development/Sales | rect_act | Receivables/Current Assets |
| rect_turn | Receivables Turnover | roa | Return on Assets |
| roce | Return on Capital Employed | roe | Return on Equity |

| | | | |
|---|---|---|---|
| sale_equity | Sales/Stockholders Equity | sale_invcap | Sales/Invested Capital |
| sale_nwc | Sales/Working Capital | short_debt | Short-Term Debt/Total Debt |
| totdebt_invcap | Total Debt/Invested Capital | | |

Part 2. Other Firm Fundamentals

| Variable | Definition | Variable | Definition |
|---|---|---|---|
| asset_g | Growth Rate in Total Assets | invest_g | Growth Rate in Capital Expenditure |
| sales_g | Growth Rate in Sales | return | Monthly Stock Return |

Part 3. Macroeconomic Variables

| Variable | Definition | Variable | Definition |
|---|---|---|---|
| con_g | Log Difference of Consumption in Goods and Services | IPT_g | Log Difference of Industrial Production Index |
| GDP_g | Log Difference of Real GDP | unemployment | Unemployment Rate |

Part 4. Earnings-Related Variables

| Variable | Definition | Variable | Definition |
|---|---|---|---|
| Realized_EP_ANN | Realized Annual Earnings from Last Period/Stock Price from Last Month | Realized_EP_QTR | Realized Quarter Earnings from Last Period/Stock Price from Last Month |
| AF_EP_lag | Mean Analyst Forecast from Last Period /Stock Price from Last Month | NUMEST_lag | Number of Forecasts from Last Period |
| Realized_ANN_g | Growth Rate in Realized Annual Earnings | Realized_QTR_g | Growth Rate in Realized Quarter Earnings |
| AF_g_lag | Lag 1 Growth Rate in Mean Analyst Forecast | Maturity | Months to Fiscal End Date/12 |

## A.2 Robustness Check - Recall Distortion and Forecast Errors

Table A2 reports the results of the analogous regression analysis to that shown in Table 3, but the recalls are found based on the similarity measure defined in (4.2). The estimation results are robust to the choice of the similarity measure.

Table A2: Analysts' recall distortion and forecast errors with cosine similarity

| | $FE_{j,t}$ | | | $\lvert FE_{j,t}\rvert$ | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Linear | Linear | Ologit | Linear | Linear | Ologit |
| $RD_{j,t}$ | 0.290*** | 0.279*** | 0.747*** | | | |
| | (0.014) | (0.015) | (0.022)) | | | |
| $\lvert RD_{j,t}\rvert$ | | | | 0.047*** | 0.047*** | 0.130*** |
| | | | | (0.006) | (0.006) | (0.023) |
| Observations | 277,487 | 277,245 | 277,487 | 277,487 | 277,245 | 277,487 |
| Month fixed effects | Yes | No | No | Yes | No | No |
| Firm fixed effects | Yes | Yes | No | Yes | Yes | No |
| Industry*Month fixed effects | No | Yes | No | No | Yes | No |

This table presents results for regressions of the form

$$FE_{j,t} = \beta \times RD_{j,t} + \varepsilon_{j,t},$$

where the dependent variable is the analyst (consensus) forecast revision error for firm $j$ at time $t$ and $FE_{j,t} \in \{-2, -1, 0, 1, 2\}$, the independent variable is the analyst's recall distortion which are predicted by the model with all the information available before time $t$ as defined in Equation (4.3). Recalls are found based on the similarity measure shown in (4.1). Column (4)-(6) report the results that take the absolute value of both the dependent variable and the independent variable. Column (3) and (6) shows the results of the ordered logit model while the rest of the columns are associated with the linear regression model. Data are from the period 2005 to 2020. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

## A.3 Robustness Check - Dispersion of Context and Forecasts

For robustness check, we present an alternative context dispersion measure. Let $\overline{h_{j,t}} = \frac{1}{N_{j,t}} \sum_{i=1}^{N_{j,t}} h_{i,j,t}$ denote the mean vector of $\{h_{i,j,t}\}_{i=1}^{N_{j,t}}$, then the alternative memory dispersion is defined as the mean of the cosine distance between each $h_{i,j,t}$ and $\overline{h_{j,t}}$:

$$Context\_Dis_{j,t}^{alt} = \frac{1}{N_{j,t}} \sum_{i=1}^{N_{j,t}} \left(1 - \frac{h_{i,j,t} \cdot \overline{h_{j,t}}}{\|h_{i,j,t}\| \|\overline{h_{j,t}}\|}\right).$$

Table A3 then reports the estimation results, the interested relation remains significantly positive.

Table A3: Dispersion of Covering Lengths, Contexts, and Analysts Forecast Revisions

| | *AF_Dis* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| $Context\_Dis^{alt}$ | 0.750*** | 0.697*** | 0.599*** | | | | 0.750*** | 0.710*** | 0.625*** |
| | (0.084) | (0.085) | (0.083) | | | | (0.084) | (0.089) | (0.087) |
| *Length_Dis* | | | | 0.031 | -0.179 | -0.325*** | -0.003 | -0.223 | -0.378*** |
| | | | | (0.109) | (0.140) | (0.049) | (0.108) | (0.143) | (0.052) |
| Observations | 179,956 | 179,911 | 179,911 | 179,956 | 179,911 | 179,911 | 179,956 | 179,911 | 179,911 |
| Month fixed effects | Yes | No | Yes | Yes | No | Yes | Yes | No | Yes |
| Firm fixed effects | No | Yes | Yes | No | Yes | Yes | No | Yes | Yes |

This table presents results for regressions of the form

$$AF\_Dis_{j,t} = \beta_c \times Context\_Dis_{j,t}^{alt} + \beta_l \times Length\_Dis_{j,t} + \varepsilon_{j,t},$$

where the dependent variable is the dispersion (standard deviation) of analysts' forecast revisions for firm $j$ at time $t$, the first independent variable is the dispersion (cosine distance) of analysts' mental contexts $h_{i,j,t}$ which are predicted by the model with all the information available before time $t$, and the second independent variable $Length\_Dis_{j,t}$ is the dispersion (standard deviation) of lengths of the analysts' covering experiences. Data are from the period 2005 to 2020. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

## A.4 Robustness Check - Random Seeds

For robustness check, we present estimation results in table 3 and table 4 with 100 different random seeds. Figure A1 shows the point estimates of Forecast error $FE_{j,t}$ on recall distortion $RD_{j,t}$ with and without taking absolute values of the dependent variable and the independent variable. Figure A2 shows the point estimates of analyst forecast revision dispersion ($AF\_Dis$) on mental context distance $Context\_Dis$ with and without controlling for covering experience distance $Length\_Dis$. Figure A3 shows the point estimates of analyst forecast revision dispersion ($AF\_Dis$) on covering experience distance $Length\_Dis$ with and without controlling for mental context distance $Context\_Dis$. In sum, all of these three figures show that our estimation results are robust to different random seeds.

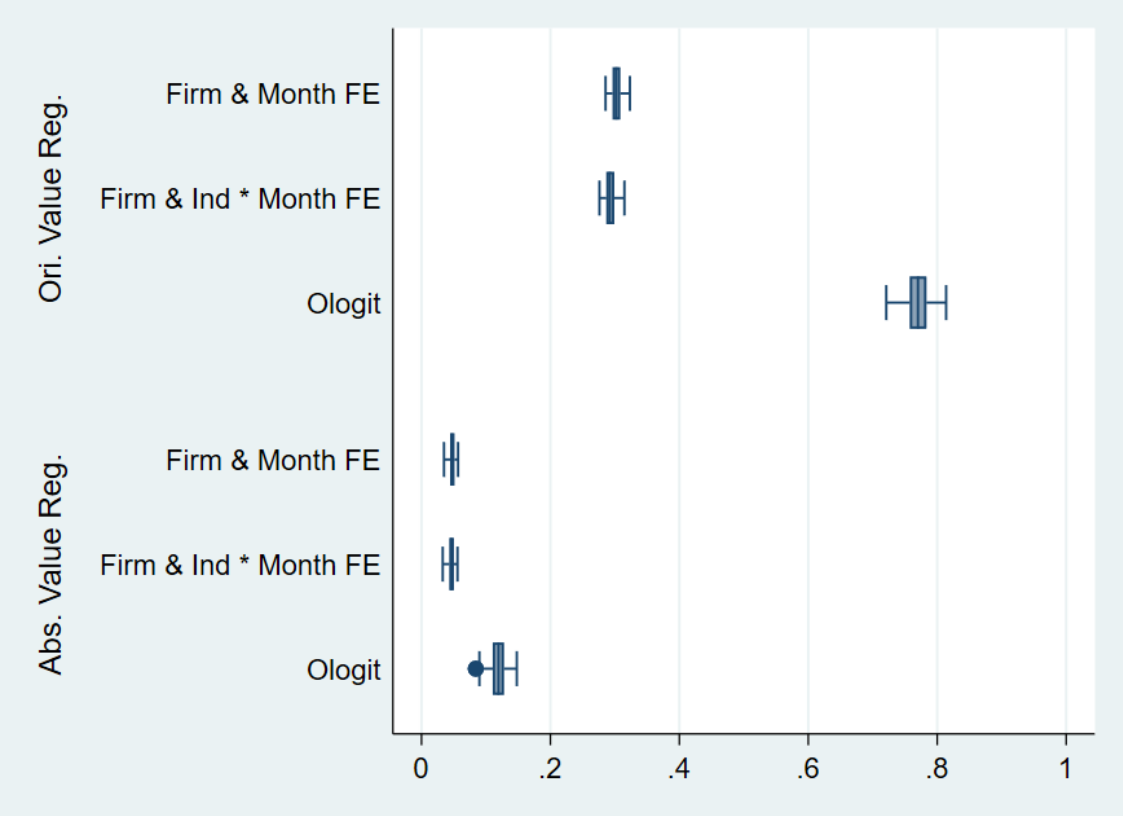



Figure A1: Point estimates on Recall Distortion $RD_{j,t}$ with 100 random seeds

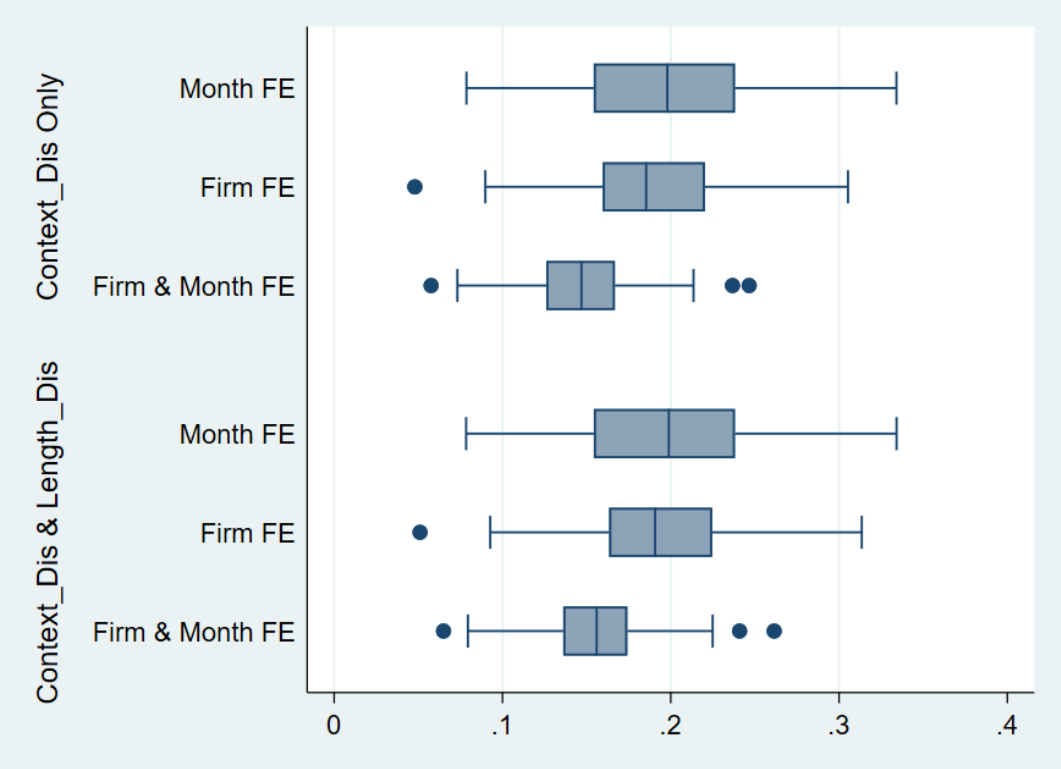



Figure A2: Point estimates on *Context_Dis* with 100 random seeds

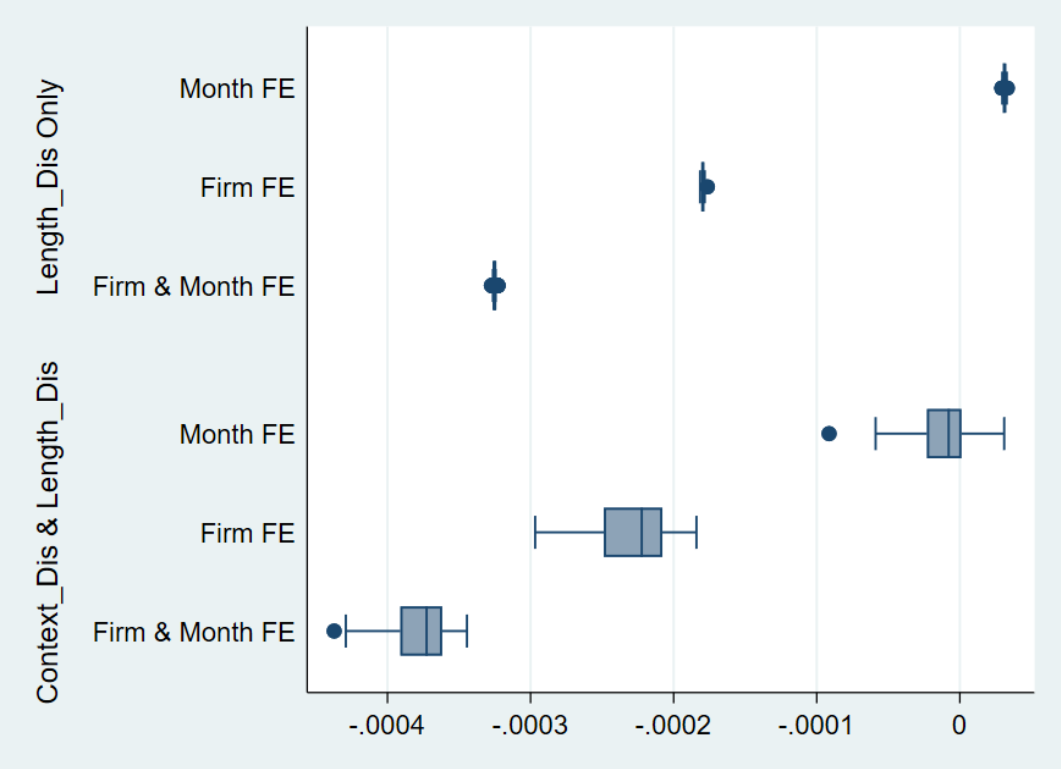



Figure A3: Point estimates on *Length_Dis* with 100 random seeds

## A.5 Temporal Contiguity - RNN

In this section, we present the simulation results for RNN which examine the temporal contiguity effect. The simulation design is analogous to what described in Section 6.1, the only difference is now we employ the trained RNN model instead of LSTM. The RNN model is also recursively trained to fit the represent analyst's forecasting decisions as stated in Section 3.2.

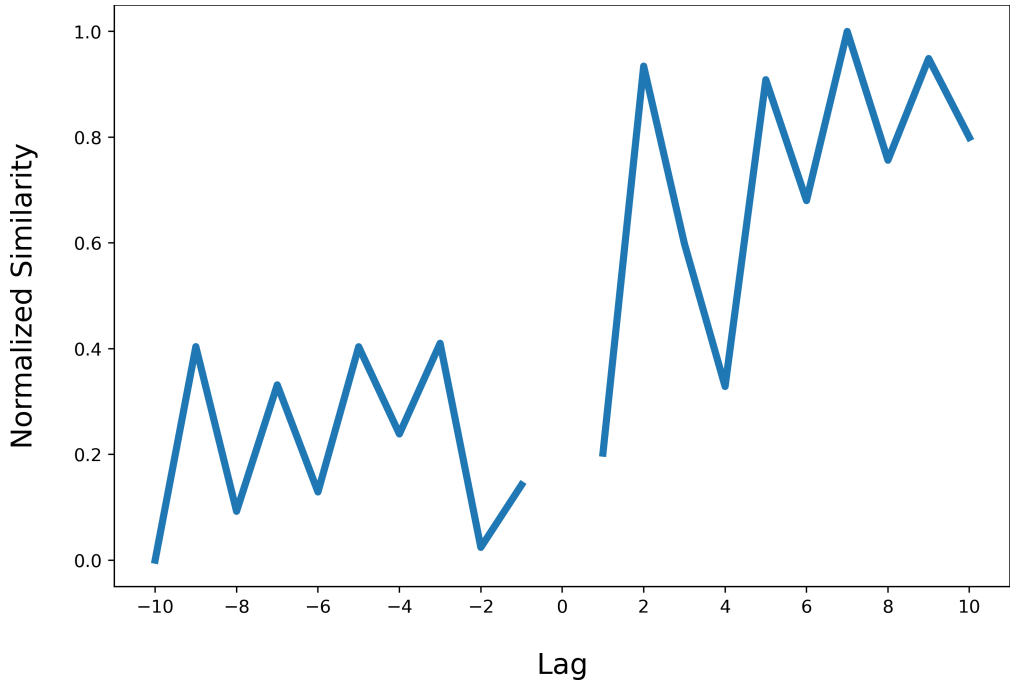



Figure A4: Temporal Contiguity of RNN

Figure A4 demonstrates that RNN does not produce temporal contiguity. The similarity presented in Figure A4 is not consistent with the conditions (6.1) implied by temporal contiguity. However, as indicated in Figure 9, temporal contiguity should be significant in analysts' forecasting processes. As implied by Howard and Kahana (2002), the channel of long-term memory is essential in generating temporal contiguity. Lack of the channel of long-term memory disables RNN to produce temporal contiguity despite of containing the autoregressive context structure. This also makes RNN inferior to model analysts' belief formation processes, compared to LSTM.

# References


Barberis, Nicholas and Lawrence Jin. 2021. "Model-free and Model-based Learning as Joint Drivers of Investor Behavior." Working paper .

Bengio, Yoshua, Patrice Simard, and Paolo Frasconi. 1994. "Learning long-term dependencies with gradient descent is difficult." IEEE transactions on neural networks 5 (2):157–166.

Bernanke, Ben S. 2015. The Courage to Act. New York: W.W.Norton & Company.

Bianchi, Francesco, Sydney C Ludvigson, and Sai Ma. 2022. "Belief distortions and macroeconomic fluctuations." American Economic Review 112 (7):2269–2315.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2019. "Diagnostic expectations and stock returns." The Journal of Finance 74 (6):2839–2874.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. "Memory, Attention, and Choice." The Quarterly Journal of Economics 135 (3):1399–1442.

Bradley, Daniel, Sinan Gokkaya, and Xi Liu. 2017. "Before an analyst becomes an analyst: Does industry experience matter?" The Journal of Finance 72 (2):751–792.

Brunnermeier, Markus, Emmanuel Farhi, Ralph SJ Koijen, Arvind Krishnamurthy, Sydney C Ludvigson, Hanno Lustig, Stefan Nagel, and Monika Piazzesi. 2021. "Perspectives on the Future of Asset Pricing." The review of financial studies 34 (4):2126–2160.

Charles, Constantin. 2022. "Memory Moves Markets." Working Paper .

Chen, Luyang, Markus Pelger, and Jason Zhu. 2023. "Deep learning in asset pricing." Management Science .

De La O, Ricardo and Sean Myers. 2021. "Subjective cash flow and discount rate expectations." The Journal of Finance 76 (3):1339–1387.

———. 2022. "Which Subjective Expectations Explain Asset Prices?" Working paper .

Diether, Karl B, Christopher J Malloy, and Anna Scherbina. 2002. "Differences of opinion and the cross section of stock returns." The Journal of Finance 57 (5):2113–2141.

Dubey, Shiv Ram, Satish Kumar Singh, and Bidyut Baran Chaudhuri. 2022. "Activation functions in deep learning: A comprehensive survey and benchmark." Neurocomputing 503:92–108.

Elman, Jeffrey L. 1990. "Finding structure in time." Cognitive science 14 (2):179–211.

Gillund, Gary and Richard M Shiffrin. 1984. "A retrieval model for both recognition and recall." Psychological review 91 (1):1.

Glenberg, Arthur M and Naomi G Swanson. 1986. "A temporal distinctiveness theory of recency and modality effects." Journal of Experimental Psychology: Learning, Memory, and Cognition 12 (1):3.

Goetzmann, William N, Akiko Watanabe, and Masahiro Watanabe. 2022. "Evidence on Retrieved Context: How History Matters." Working paper .

Graves, Alex, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou et al. 2016. "Hybrid computing using a neural network with dynamic external memory." Nature 538 (7626):471–476.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. "Empirical asset pricing via machine learning." The Review of Financial Studies 33 (5):2223–2273.

Healey, M Karl, Nicole M Long, and Michael J Kahana. 2019. "Contiguity in episodic memory." Psychonomic bulletin & review 26 (3):699–720.

Hirshleifer, David, Yaron Levi, Ben Lourie, and Siew Hong Teoh. 2019. "Decision fatigue and heuristic analyst forecasts." Journal of Financial Economics 133 (1):83–98.

Hirshleifer, David, Ben Lourie, Thomas G Ruchti, and Phong Truong. 2021. "First Impression Bias: Evidence from Analyst Forecasts." Review of Finance 25 (2):325–364.

Hirshleifer, David and Siew Hong Teoh. 2003. "Limited attention, information disclosure, and financial reporting." Journal of accounting and economics 36 (1-3):337–386.

Hochreiter, Sepp and Jürgen Schmidhuber. 1997. "Long short-term memory." Neural computation 9 (8):1735–1780.

Howard, Marc W and Michael J Kahana. 2002. "A distributed representation of temporal context." Journal of mathematical psychology 46 (3):269–299.

Jagtap, Ameya D, Kenji Kawaguchi, and George Em Karniadakis. 2020. "Adaptive activation functions accelerate convergence in deep and physics-informed neural networks." Journal of Computational Physics 404:109136.

Jiang, Zhengyang, Hongqi Liu, Cameron Peng, and Hongjun Yan. 2022. "Investor Memory and Biased Beliefs: Evidence from the Field." Working paper .

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp. 2016. "A rational theory of mutual funds' attention allocation." Econometrica 84 (2):571–626.

Kahana, Michael J. 1996. "Associative retrieval processes in free recall." Memory & cognition 24 (1):103–109.

———. 2020. "Computational models of memory search." Annual review of psychology 71:107–138.

Kelly, Bryan T, Seth Pruitt, and Yinan Su. 2019. "Characteristics are covariances: A unified model of risk and return." Journal of Financial Economics 134 (3):501–524.

Kingma, Diederik P and Jimmy Ba. 2014. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 .

Kwan, Alan, Yukun Liu, and Ben Matthies. 2022. "Institutional Investor Attention." Working Paper .

Malmendier, Ulrike and Stefan Nagel. 2011. "Depression babies: Do macroeconomic experiences affect risk taking?" The Quarterly Journal of Economics 126 (1):373–416.

———. 2016. "Learning from inflation experiences." The Quarterly Journal of Economics 131 (1):53–87.

Nagel, Stefan. 2021. Machine learning in asset pricing. Princeton University Press.

Nagel, Stefan and Zhengyang Xu. 2022. "Asset pricing with fading memory." The Review of Financial Studies 35 (5):2190–2245.

Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio. 2013. "On the difficulty of training recurrent neural networks." In International conference on machine learning. PMLR, 1310–1318.

Polyn, Sean M, Kenneth A Norman, and Michael J Kahana. 2009. "A context maintenance and retrieval model of organizational processes in free recall." Psychological review 116 (1):129.

So, Eric C. 2013. "A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts?" Journal of Financial Economics 108 (3):615–640.

Teräsvirta, Timo. 2006. "Forecasting economic variables with nonlinear models." Handbook of economic forecasting 1:413–457.

van Binsbergen, Jules H, Xiao Han, and Alejandro Lopez-Lira. 2020. "Man vs. Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases." Working paper .

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention is all you need." Advances in neural information processing systems 30.

Wachter, Jessica A and Michael Jacob Kahana. 2022. "A Retrieved-Context Theory Of Financial Decisions." Working paper .

Walters, Daniel J and Philip M Fernbach. 2021. "Investor memory of past performance is positively biased and predicts overconfidence." Proceedings of the National Academy of Sciences 118 (36):e2026680118.

Weston, Jason, Sumit Chopra, and Antoine Bordes. 2014. "Memory networks." arXiv preprint arXiv:1410.3916 .

Wilks, Samuel S. 1932. "Certain generalizations in the analysis of variance." Biometrika :471–494.

Woodford, Michael. 2020. "Modeling imprecision in perception, valuation, and choice." Annual Review of Economics 12:579–601.