Surveying Generative AI's Economic Expectations *

J. Leland Bybee

Yale University

Abstract

I introduce a survey of economic expectations formed by querying a large language model (LLM)'s expectations of various financial and macroeconomic variables based on a sample of news articles from the Wall Street Journal between 1984 and 2021. I find the resulting expectations closely match existing surveys including the Survey of Professional Forecasters (SPF), the American Association of Individual Investors, and the Duke CFO Survey. Importantly, I document that LLM based expectations match many of the deviations from full-information rational expectations exhibited in these existing survey series. The LLM's macroeconomic expectations exhibit under-reaction commonly found in consensus SPF forecasts. Additionally, its return expectations are extrapolative, disconnected from objective measures of expected returns, and negatively correlated with future realized returns. Finally, using a sample of articles outside of the LLM's training period I find that the correlation with existing survey measures persists – indicating these results do not reflect memorization but generalization on the part of the LLM. My results provide evidence for the potential of LLMs to help us better understand human beliefs and navigate possible models of nonrational expectations.

Expectations matter. How much to consume or save, what price to set, and whether to hire or fire workers are just some of the fundamental decisions underlying macroeconomic dynamics that hinge upon agents' expectations of the future. Yet how those expectations are formed, and how best to model this process, remains an open question.

Coibion and Gorodnichenko (2015)

Beliefs are central to asset pricing. Asset prices are forward-looking, and essentially any asset-pricing model implies that investors price assets based on their beliefs about the joint distribution of some stochastic discount factor (SDF) M_{t+1} and payoffs X_{t+1}. An observer outside the field of asset pricing might therefore guess that a major part of the research efforts in asset pricing are devoted to understanding how investors form beliefs. This is, at least so far, not the case.

Brunnermeier et al. (2021)

Anyone who sets himself the problem of analyzing the causation of behavior will (in the absence of independent neurophysiological evidence) concern himself with the only data available, namely the record of inputs to the organism and the organism's present response.

Chomsky (1959)[1]

---

[1]This quote is taken from Chomsky's review of Skinner (1957) and presents Chomsky's summary of Skinner's approach. While the review itself is critical, the behaviorist approach, so summarized, is nevertheless relevant to the present study of LLMs.

1 Introduction

The rational expectations hypothesis of Muth (1961) remains the dominant model of beliefs in macroeconomics and finance. Its dominance is not hard to understand: rational expectations provides a tractable approach for modeling economic behavior (Lucas (1972)), while constraining the econometrician's degrees of freedom (Lucas (1976)). However, rational expectations has never been without evidence and alternative theories to question its dominance. Such models place us in the "wilderness" of alternative expectations of Sims (1980).[2] Successful attempts to navigate this wilderness have primarily relied on survey data for proxies of beliefs (Coibion and Gorodnichenko (2015), Bordalo et al. (2020), Angeletos et al. (2021), Nagel and Xu (2022a), Lochstoer and Muir (2022)).

In this paper I propose a new source for proxies of beliefs using large language models (LLMs). LLMs, are a class of statistical model designed to learn the structure of human language. These models accomplish this goal by estimating the probability of a token, s_i, given all previously observed tokens in a document,

$$p(s_i|s_1, \ldots, s_{i-1}), \tag{1}$$

using a particular neural network architecture, known as transformers (Vaswani et al. (2017)). While this objective may seem simple, by leveraging massive corpuses of training data and a very large parameter space, these models have exhibited an emergent ability to mimic "human-like" behavior (Brown et al. (2020), Wei et al. (2022), Bubeck et al. (2023)).[3]

This ability to mimic human-like interactions with text opens up the possibility for LLMs to simulate human behavior and beliefs. Rich causal beliefs can be embedded in text – this assumption underlies the entire academic publication process. As a statistical procedure designed to generate text similar to its training corpus, LLMs should reflect the beliefs embedded in their training data.[4] I leverage this insight to form expectations of an LLM by providing a historical sample of news articles from *The Wall Street Journal* (*WSJ*) to OpenAI's GPT-3.5 instance and asking it to predict various financial and macroeconomic quantities. I then aggregate these article level expectations into a time-series of monthly and quarterly expectations and compare the resulting belief proxies to existing surveys.

---

[2]Sims himself refers to the "wilderness of disequilibrium economics" but this wilderness has since been rephrased as Sargent (2001)'s "wilderness of bounded rationality" and Angeletos et al. (2021)'s "wilderness of alternative models for expectations formation and equilibrium".

[3]Emergent here refers to the scale of these models as opposed to smaller language models. For instance, Open AI's GPT-3 instance used roughly 45 terabytes of text for training data and has roughly 175 billion parameters.

[4]That LLMs reflect the biases and beliefs of their training corpus is central to much of the current work on AI alignment (Bender et al. (2021), Schramowski et al. (2022))

First, I validate how well GPT's expectations of the stock market match the expectations of the American Association of Individual Investors (AAII) survey and Duke CFO Survey. I find GPT's expectations significantly correlate with these existing surveys on a comparable level to alternative existing survey measures. I then compare the correlation between GPT's expectations of the S&P and various alternative measures of expected returns to that of the existing surveys. I find that GPT exhibits the same extrapolative expectations and significantly correlate with equity fund flows. Additionally, I find that GPT's expectations exhibit the same disconnect from objective measures of expected returns – matching the sign of the correlation for existing survey measures with the log dividend price ratio, Lettau and Ludvigson (2001)'s CAY measure, and several predictive proxies for expected returns. Finally, I find that GPT's expectations are negatively correlated with future realized returns, matching existing survey measures while differing from objective measures of expected returns.

Second, I compare GPT's expectations, for a series of macroeconomic variables studied in Coibion and Gorodnichenko (2015), to the Survey of Professional Forecaster (SPF). I find that GPT's expectations are significantly correlated with revisions in all but two of the SPF series. I then evaluate whether the variation in SPF revisions associated with GPT's expectations is a driver of the underreaction commonly found in consensus SPF forecasts. By running Coibion-Gorodnichenko (CG) regressions using both SPF revisions and GPT's expectations, I find that GPT is able to match the observed underreaction.

These results provides new evidence for navigating the wilderness of alternative expectations as GPT's expectations match many of the deviations from full-information rational expectations (FIRE) exhibited in existing survey series. There are two possible sources for deviations from FIRE in GPT's expectations, F_t^{gpt}, about the kth series, X_{t+h}^k, given as follows:

$$F_t^{gpt}(X_{t+h}^k) = g_k(\theta_t). \tag{2}$$

Here g_k corresponds to the estimated LLM weights prompted for the given series k and θ_t the provided news text at period t. First, note that g_k is static across the entire sample. Second, note that the news text provided to GPT is isolated and lacking in context – GPT's expectations at time t are based only on the set of news articles available at time t.[5] This puts constraints on the set of models which can explain GPT's deviations.

If the deviations in GPT's expectations emerge from the estimated LLM weights, g_k, then models to explain then models of beliefs must reflect a static reaction function to news. Models that rely on an overemphasis of representative or salient news (Bordalo et al. (2018), Bordalo et al. (2022)) may explain this bias if the representativeness is not dependent on

---

[5]In fact it is stronger than this, GPT's expectations are based on individual news articles and even lack the context of other news articles at time t.

recent context but some more general attribute of the given news articles. The potential of the LLMs to capture these biases indicates the importance of a nascent multidisciplinary literature experimentally studying the behavior of these models (Aher et al. (2022), Horton (2023), Argyle et al. (2023), Brand et al. (2023)).

Alternatively, if deviations in GPT's expectations emerge from attributes of the news articles themselves, θ_t, then models to explain these deviations need to address these attributes. In particular, for models of learning, extrapolation, and the representativeness of recent states to explain these deviations (Fuster et al. (2010), Afrouzi et al. (2020), Farmer et al. (2021)), it must be the case that these dynamics are embedded in the news text generating process itself. This potential source of deviations highlights the importance of work studying the origin and dynamics of narratives as a driver of beliefs (Shiller (2019), Bybee et al. (2021), Andre et al. (2021), Michalopoulos and Xue (2021), Bybee et al. (2022), Flynn and Sastry (2022)).

Further still, it is interesting to note that GPT is able to match distinct moments across a variety of existing survey series. Given GPT's expectations are based on the same algorithm and the same source of news, this suggest the distinct deviations observed across these surveys originate from the same source and models attempting to explain these moments should be able to jointly explain these results.

An alternative story that may contaminate my results is the possibility of data leakage from the training corpus. Importantly for my story, I assume that GPT has learned some generalization of the causal structure of beliefs contained in its training sample not that it is responding with memorized text. To address the relevance of this concern for my results I run two tests. First, using a sample of *WSJ* articles outside of GPT's training period – after September of 2021 – I test whether GPT's expectations continue to correlate with the existing survey measures, I find that this is indeed the case. Second, if GPT is relying on information about the future then realized values of the corresponding series should be more important for explaining GPT's expectations than the forecast revisions. I find that this is not the case.

An open question still remains: who's beliefs are GPT's? My results suggest GPT's beliefs closely match a number of different groups, including professional forecasters, individual investors, and CFOs. GPT's training corpus is based on a wide range of sources and topics, and as such it may provide something of a "representative agent". Nevertheless, LLMs designed to more directly reflect the beliefs and interest of distinct groups is likely the future of this field. This work may be done by fine tuning existing open source LLMs like that in Touvron et al. (2023) or building from the ground up from distinct corpuses as in Wu et al. (2023). My results provide a proof of concept that these methods represent a valuable route forward to better understand human behavior.

Figure 1: Prompt Format

```
Here is a piece of news:

"%s"

Do you think this news will increase or decrease %s?

Write your answer as:

{increase/decrease/uncertain}:
{confidence (0-1)}:
{magnitude of increase/decrease (0-1)}:
{explanation (less than 25 words)}
```

Note. Reports the prompt format for queries made to GPT. "%s" indicates where in the prompt the headline and target text are inserted.

2 Measuring Generative AI's Expectations

My primary dataset consists of all articles contained in *The Wall Street Journal* from 1984 to 2021 purchased from the Dow Jones Historical News Archive. A full summary of the data cleaning procedure is detailed in appendix A.1. Beyond this initial cleaning I sample 300 articles randomly from each month of data. This is done because each request made to OpenAI's API costs a marginal fee.

Queries are made to OpenAI's GPT-3.5 model instance. Figure 1 reports the prompt format used to query GPT. Each query to GPT receives a headline and a description of the desired series – for instance, for CPI I will ask about an increase or decrease in "the consumer price index in the United States". In response GPT generates a string of text corresponding to the requested format in the prompt.

Table 1 reports the surveyed series as well as summary statistics of the responses. Figure 7 in appendix B reports the coocurrence between the possible pairings of increase/decrease across the surveyed series.[6] Table 2 reports a series of example headlines and the corresponding responses for the S&P 500, CPI, and unemployment. Additionally, I request GPT to provide an explanation for its response which is included in table 2. These explanations provide a useful method to analyze the reasoning behind GPT's responses. For instance, when presented with a news article about the opening of access to the Japanese phone market for

---

[6]As opposed to classical supervised machine learning, GPT is not trained on a specific task, i.e. forecasting economic quantities, but rather on a general language modeling task. As such, this sort of classification is referred to as zero-shot learning in the machine learning literature.

Table 1: Summary Statistics of GPT Survey

Series	Prompt	Date Range	Count	Inc. %	Dec. %	Unc. %
SNP	the S&P 500 index	1984-2021	136345	15.13	26.84	58.02
CPI	the consumer price index in the United States	1984-2021	132736	7.86	6.45	85.69
HS	housing starts in the United States	1984-2021	132212	2.50	5.68	91.82
IP	industrial production in the United States	1984-2021	132892	10.11	11.72	78.17
DEFL	the GDP price deflator in the United States	1984-2021	132760	9.63	12.50	77.88
AAA	the AAA-rated bond's rate in the United States	1984-2021	133467	11.09	14.88	74.03
C	real consumption in the United States	1984-2021	131574	11.53	17.67	70.80
GF	federal government consumption in the United States	1984-2021	132839	9.86	10.88	79.26
GY	the real GDP of the United States	1984-2021	132148	20.54	20.91	58.56
NRI	real nonresidential investment in the United States	1984-2021	132961	17.46	22.94	59.59
RI	real residential investment in the United States	1984-2021	133157	8.66	16.49	74.85
GS	state and local government consumption in the United States	1984-2021	131428	13.44	17.30	69.26
3TB	the 3-month treasury bill rate	1984-2021	134609	15.45	11.21	73.34
UE	employment in the United States	1984-2021	120102	9.81	11.24	78.95

Note. Reports summaries of the surveyed labels, the text passed to GPT, the date range, the number of non-missing responses, the proportion of responses that were "increase", the proportion of responses that were "decrease", and the proportions that were "uncertain".

U.S. firms, GPT expects this will increase the S&P by providing growth opportunities for U.S. phone companies, will decrease unemployment by creating new job opportunities in the U.S. telecommunications industry, and lower the CPI by increasing competition. Similarly, when presented with an article about tax cuts and tightening government budgets during the Bush years, GPT expects this will increase the S&P by stimulating the U.S. economy, no impact unemployment due to uncertainty about which effect may dominate, and increase the CPI by increased demand as a result of the tax cuts. These explanations highlight GPT's ability to provide a clear chain of reasoning for its answers and open up new possible routes for understanding the causal structure of beliefs.

Table 2: Example Responses

Series	Direction	Headline/Response
		(1989-06-29) U.S. Reaches Accords Widening Access To the Mobile Phone Business in Japan
S&P	1	The news signals potential growth opportunities for US mobile phone companies, which could positively impact the S&P 500 index.
UE	-1	Increased access to the Japanese mobile phone market will likely create new job opportunities in the U.S. telecommunications industry.
CPI	-1	Increased competition and access to the Japanese mobile phone market may lead to lower prices for U.S. consumers.
		(2001-02-28) Bush Offers Tax Cuts and Tight Budgets to Aid 'Faltering' Economy
S&P	1	Tax cuts stimulate the economy, which could lead to increased corporate profits and higher stock prices.
UE	0	Tax cuts may stimulate investment, but tight budgets could reduce government spending and slow job growth.
CPI	1	Tax cuts usually stimulate spending, which can increase demand and raise prices, but budget tightening may counteract that effect.

Note. Reports a sample of example article headlines and the corresponding responses from GPT. The first column reports the series queried. The second column reports the direction of the query. The third column reports the headline and the corresponding response from GPT.

Finally, since GPT produces article level binary expectations, I aggregate these expectations to place them at a comparable frequency to the survey data. To do this aggregation I compute a "balance statistic": the proportion of articles where GPT responds with increases minus the proportion articles where GPT responds with decreases:

$$F_t^{gpt}(X_{t+h}^k) = \frac{\sum_{i \in A_t} \mathbb{I}(\text{Increase})_i^k - \mathbb{I}(\text{Decrease})_i^k}{\sum_{i \in A_t} \mathbb{I}(\text{Increase})_i^k + \sum_{i \in A_t} \mathbb{I}(\text{Decrease})_i^k}. \tag{3}$$

This approach is common for a number of survey measures and is used for other surveys such as the Gallup survey, the University of Michigan Survey of Consumers, and the American Association of Individual Investors survey.

3   Return Expectations

I first validate GPT's expectations of returns by comparing them to two publicly available benchmark return series used in Greenwood and Shleifer (2014). The first is the American Association of Individual Investors (AAII) Investor Sentiment Survey. The AAII survey is a weekly survey of members of the AAII running from 1987 up to the present day which measures the percentage of participants that are bullish, bearish, or neutral on the stock market for the next six months. Following Greenwood and Shleifer (2014), I aggregate the weekly responses to monthly for the majority of my analysis. The second survey is Duke or Graham and Harvey survey of chief financial officers (CFOs), started in 1998 by John Graham and Campbell Harvey. The survey requests CFO views on a variety of macro and firm specific quantities including their expectations of returns for the U.S. stock market over the next twelve months. The survey runs from 2000 to the present day.

Figure 8 reports the time series of GPT's expectations using a three-month window for aggregation, along with the corresponding survey series. Additionally, it reports the correlation between each survey series and various aggregates of GPT's expectations. I consider aggregates over one, two, and three month windows as well as an exponentially weighted average over daily aggregates. For the exponentially weighted average I report correlation for the optimal smoothing parameter λ that maximizes the correlation with the survey series.

I find that for all different aggregation methods GPT's expectations are significantly correlated with the existing survey measures. For the remainder of this section I focus on the three-month aggregation window. For this window, the average correlation between GPT's expectations and the existing survey measures is 0.47. For comparison Greenwood and Shleifer (2014) find an average correlation of 0.43 among the full set of return expectation surveys they evaluate.

These results suggest GPT's expectations closely match the variation in existing survey measures, as a result I next evaluate how well GPT's expectations match key properties of existing return expectations. To test this, I next compare the correlation between GPT's expectations and alternative measures studied previously in the literature vs. the existing survey measures.

An extensive literature in asset pricing has focused on the importance of extrapolative expectations (Cutler et al. (1990), Barsky and De Long (1993), Lakonishok et al. (1994), Barberis et al. (2015), Jin and Sui (2022)). As a result, I evaluate the correlation of different survey expectations against the past twelve-months returns (R_{t-12}) of the U.S. stock market, following Greenwood and Shleifer (2014) and Nagel and Xu (2022b). Additionally, a recent literature has leveraged the increased availability of realized trading behavior as a proxy for investor beliefs (Giglio et al. (2019), Gabaix and Koijen (2021), Alekseev et al. (2022)).

Figure 2: Time Series of Return Expectations

Note. Reports the time-series of GPT's monthly/quarterly standardized expectations overlaid with the AAII and CFO surveys respectively. Bottom table reports the correlation coefficients and corresponding *t*-stats in brackets. *t*-stats use Newey-West standard errors with a 12 month lag. one/two/three month avg. correspond to GPT expectations aggregated over one/two/three month windows respectively. opt. EWMA corresponds to an exponentially weighted average with the optimal tuning parameter.

	one month avg.	two month avg.	three month avg.	opt. EWMA
CFO (N=76)	0.49	0.56	0.59	0.59
	[4.84]	[5.88]	[6.21]	[6.24]
AAII (Q, N=138)	0.32	0.40	0.39	0.43
	[3.93]	[5.15]	[5.01]	[5.63]
AAII (M, N=414)	0.32	0.35	0.35	0.36
	[6.83]	[7.65]	[7.65]	[7.82]

Following this literature, I evaluate the correlation between different survey expectations and mutual fund flows into equities. To compute this measure I combine the Thomson Reuters Mutual Fund Holdings S12 database with the CRSP Survivor-Bias-Free US Mutual Funds and follow the data cleaning procedure detailed in Alekseev et al. (2022).

An exhaustive literature in asset pricing has also studied empirical measures of objective expected returns. Importantly, there exists a well known disconnect between these objective measures and subjective survey measures (Greenwood and Shleifer (2014), Nagel and Xu

(2022b)). As a result, I next consider a number of objective expected return proxies. First, I consider the log dividend-price ratio and 12 month changes in the log dividend-price ratio following much of the empirical literature. Second, I consider the consumption wealth ratio (CAY) of Lettau and Ludvigson (2001). Finally, I consider two expected return indices formed by regressing future 12-month returns on various sets of predictors. In particular, I consider the same expected return index used in Greenwood and Shleifer (2014) (following Fama and French (1989)) which uses the log dividend price ratio, the Treasury-bill yield, the default spread (the yield on BAA minus yield on AAA-rated bonds) and the term spread (the yield on ten-year government bonds minus the yield on three-month Treasury bill). Additionally, I form an expected return index using the "kitchen-sink" predictors from Welch and Goyal (2008) which represent a common benchmark in the empirical asset pricing literature.

Figure 3 reports the correlations between each survey series and the corresponding measures discussed above. For each survey series, I include a GPT based expectation series formed at the same frequency over the same time sample. For all cases except changes in CAY and the GPT expectations proxy for AAII, GPT's correlations exhibit the same sign as the existing survey measures. These results are evidence that GPT does not only match the variation in existing survey measures but more importantly the deviations from rational expectations noted in the literature, indicating the potential of these measures as a tool for studying nonrational expectations.

Finally, it is the case that measures of expected returns should forecast future returns under models of rational expectations. However, if anything, existing survey measures are negatively correlated with future returns. As a result, I next compare the correlation of GPT's expectations with future returns to that of a number of alternative subjective and objective expected return measures. Figure 4 reports the correlation coefficients for a series of predictive regressions of future returns on the expectation measures discussed above. GPT exhibits the same negative correlation with future returns as existing survey measures, which is in contrast to the positive correlation observed for objective measures.

Figure 3: Survey Correlations with Existing Moments
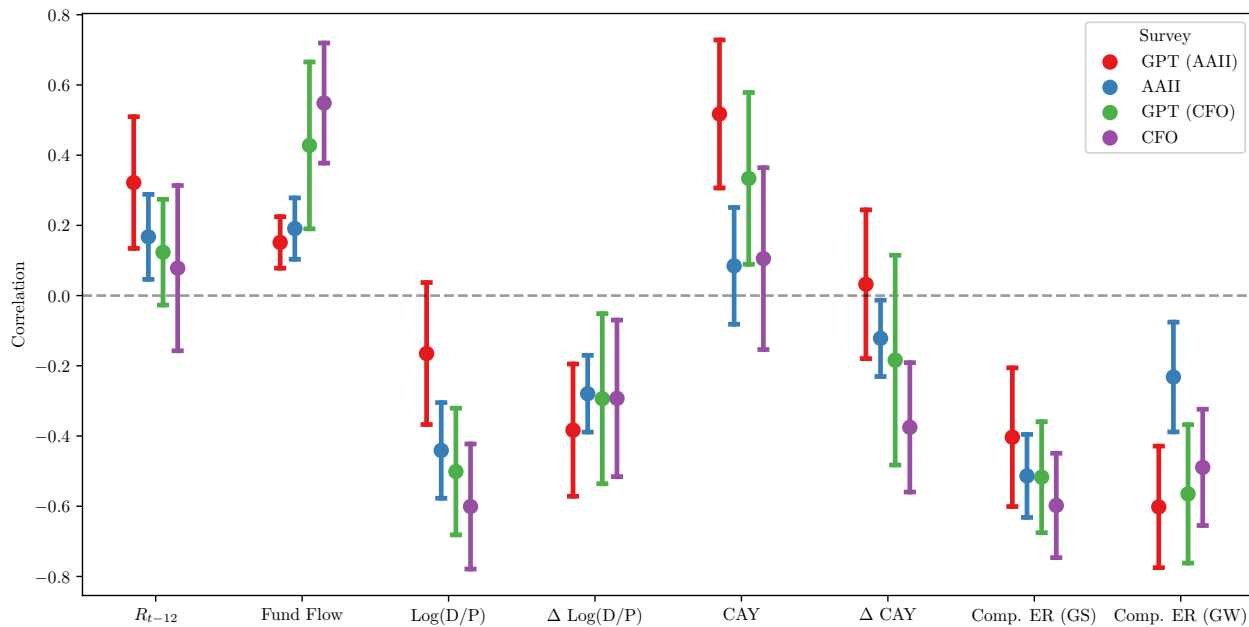
	GPT (AAII)		AAII		GPT (CFO)		CFO	
	Corr.	t-Stat	Corr.	t-Stat	Corr.	t-Stat	Corr.	t-Stat
R_{t-12}	0.32	2.82	0.17	2.27	0.12	1.35	0.08	0.54
Fund Flow	0.15	3.39	0.19	3.58	0.43	2.95	0.55	5.26
Log(D/P)	-0.17	-1.34	-0.44	-5.31	-0.50	-4.56	-0.60	-5.53
Δ Log(D/P)	-0.38	-3.34	-0.28	-4.19	-0.29	-1.99	-0.29	-2.15
CAY	0.52	4.03	0.08	0.83	0.33	2.24	0.11	0.67
Δ CAY	0.03	0.25	-0.12	-1.84	-0.18	-1.01	-0.38	-3.34
Comp. ER (GS)	-0.40	-3.35	-0.51	-7.13	-0.52	-5.36	-0.60	-6.59
Comp. ER (GW)	-0.60	-5.70	-0.23	-2.43	-0.56	-4.70	-0.49	-4.85

Note. Reports the correlation between each expectation series and the corresponding series and 90% confidence intervals, standard errors are Newey-West with a 12 month lag. Bottom table reports the corresponding correlation coefficients and t-stats. t-stats use Newey-West standard errors with a 12 month lag. GPT (AAII) corresponds to the correlation with GPT's expectations at the monthly frequency for dates where the AAII survey is available. GPT (CFO) corresponds to the correlation with GPT's expectations at the quarterly frequency for dates where the CFO survey is available. R_{t-12} corresponds to lagged 12 month returns. Fund Flow corresponds to the aggregate mutual fund flows into equities. Log(D/P) is the log dividend-price ratio and CAY the aggregate log consumption-wealth ratio of Lettau and Ludvigson (2001). Δ Log(D/P) and Δ CAY correspond to the respective 12-month changes. Comp. ER (GS) corresponds to the fitted values of a regression of one-year ahead returns on log dividend price ratio, the Treasury-bill yield, the default spread and the term spread —— the composite expected return measure used in Greenwood and Shleifer (2014). Comp. ER (GW) corresponds to the fitted values of a regression of one-year ahead returns on the "kitchen-sink" predictors from Welch and Goyal (2008).

Figure 4: Predictive Return Regressions
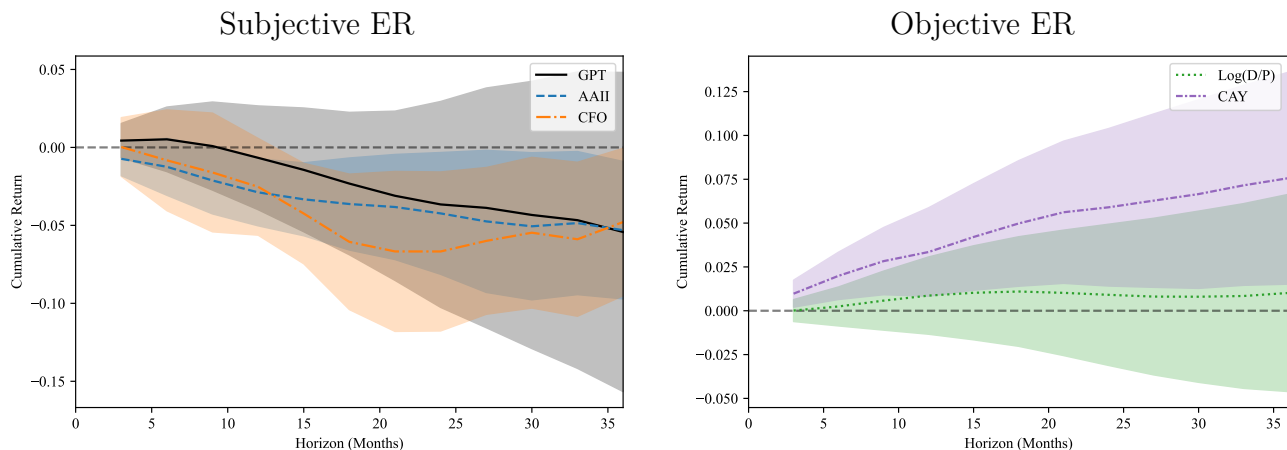
Note. Reports the coefficients for a series of predictive regressions of future cumulative returns over the given horizon on subjective and objective expected return proxies. Shaded bans report 90% confidence intervals using Newey-West standard errors with the corresponding horizon as the number of lags.

4 Macroeconomic Expectations

I next validate GPT's expectations by comparing them to a standard set of macroeconomic expectations: the Survey of Professional Forecasters (SPF). The SPF is a quarterly survey of macroeconomic forecasts conducted by the Philadelphia Federal Reserve and remains the gold standard for work studying macroeconomic expectations (Coibion and Gorodnichenko (2012), Coibion and Gorodnichenko (2015), Bordalo et al. (2020), Angeletos et al. (2021), Farmer et al. (2021)). As professional forecasters, the respondents are some of the most informed agents in the economy and as such errors in their forecasts are notable.

The SPF covers a wide range of macroeconomic quantities during my sample, including CPI, real GDP, the unemployment rate and the federal funds rate. I compare GPT's expectations to that of the SPF for the 13 series examined in Coibion and Gorodnichenko (2015). For GPT's expectations I use a quarterly aggregate of the article level expectations over the quarter prior to the surveyed period. The SPF releases forecasts at a variety of horizons, however, since the horizon of GPT's expectations is not known, I compare GPT's expectations to an average over the 1-4 quarter horizon forecasts. My results are robust to this choice and results for each of the 1-4 quarter horizon forecasts are shown in appendix D.

I first evaluate how well GPT's expectations correlate with SPF forecasts. I report results for both levels $F_t(X_{t+h}^k)$ and revisions $F_t(X_{t+h}^k) - F_{t-1}(X_{t+h}^k)$. Revisions capture the new information embedded in the forecast and given GPT's expectations are formed from granular pieces of recent news it is likely that they will be more closely related to revisions than levels. Figure 5 reports these correlations – a full set of time-series plots of GPT's expectations vs.

the corresponding SPF revisions is available in appendix C. For all but two of the sampled series, federal government consumption and state and local government consumption, GPT's expectations significantly correlated with the corresponding revisions at the 90% confidence level.

Much of the literature studying deviations from FIRE focuses on the predictability of forecast errors. In particular, considerable focus is given to the rigidity of beliefs to new information measured using Coibion and Gorodnichenko (2015) (CG) regressions. CG regressions regress forecast errors on the corresponding revisions in expectations – positive coefficients capture underreaction, while negative coefficients capture overreaction.

Given GPT's significant correlation with SPF revisions, I next evaluate whether the well documented underreaction in macroeconomic expectations can be explained by the component of revisions associated with GPT. To do this, for the series which exhibit a significant correlation between GPT and SPF revisions (that is, excluding government consumption), I regress SPF forecast errors on the corresponding version of GPT's expectations. Figure 6 reports the correlations for these regressions and compares them to the correlations using the original SPF revisions. For all but two of the series, the correlation between GPT's expectations and SPF forecast errors is positive and the overall pooled results are positive and significant at the 90% confidence level.[7]

---

[7]Unemployment exhibits significant overreaction in the original SPF revisions, going against the literature. My sample includes the 2020 Covid-19 pandemic which may impact this result. To test this, I re-estimate the same procedure limiting myself to the period prior to 2019. These results are reported in appendix E. The significant positive correlation between GPT's expectations and SPF forecast errors remains and the original CG results align with the literature.

Figure 5: GPT/SPF Correlations
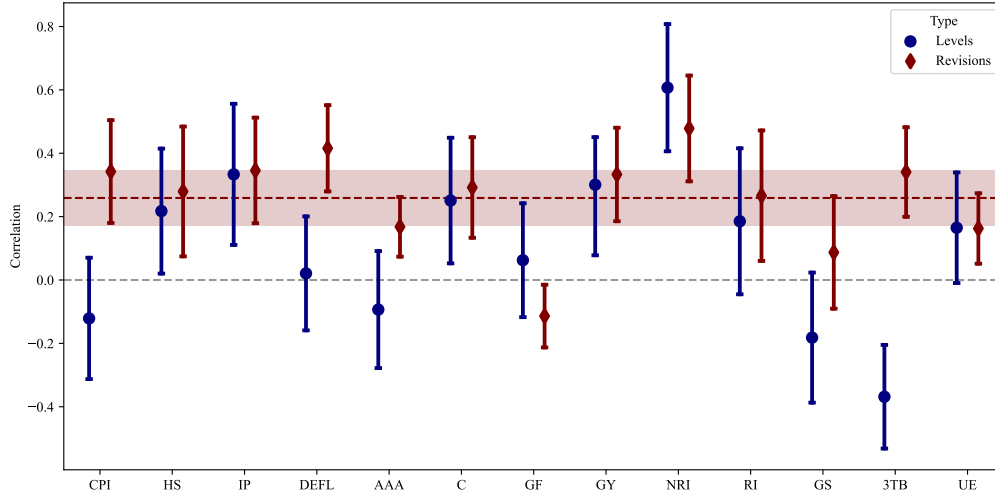
	Levels		Revisions	
	Corr.	T-Stat.	Corr.	T-Stat.
Panel	0.07	1.49	0.26	4.93
CPI	-0.12	-1.04	0.34	3.46
Housing Starts	0.22	1.81	0.28	2.24
Industrial Production	0.33	2.46	0.35	3.40
GDP Deflator	0.02	0.19	0.42	5.01
AAA Corporate Bond Yield	-0.09	-0.83	0.17	2.92
Consumption Growth	0.25	2.07	0.29	3.02
Federal Government Spending	0.06	0.57	-0.11	-1.89
GDP Growth	0.30	2.22	0.33	3.70
Nonresidential Investment	0.61	4.96	0.48	4.70
Residential Investment	0.19	1.32	0.27	2.12
State Government Spending	-0.18	-1.45	0.09	0.80
3-Month Treasury Bill	-0.37	-3.70	0.34	3.96
Unemployment Rate	0.16	1.55	0.16	2.40

Note. Reports the correlation between levels and revisions of SPF expectations and GPT expectations. Additionally reports 90% confidences intervals for the correlation coefficients. Dashed maroon line reports the panel correlation coefficient and shaded maroon band the corresponding 90% confidence interval. Standard errors for the single variable regressions are Newey-West, panel standard errors are Driscoll-Kraay. Bottom table reports the corresponding correlation coefficients and *t*-stats.

Figure 6: Coibion-Gorodnichenko Regressions
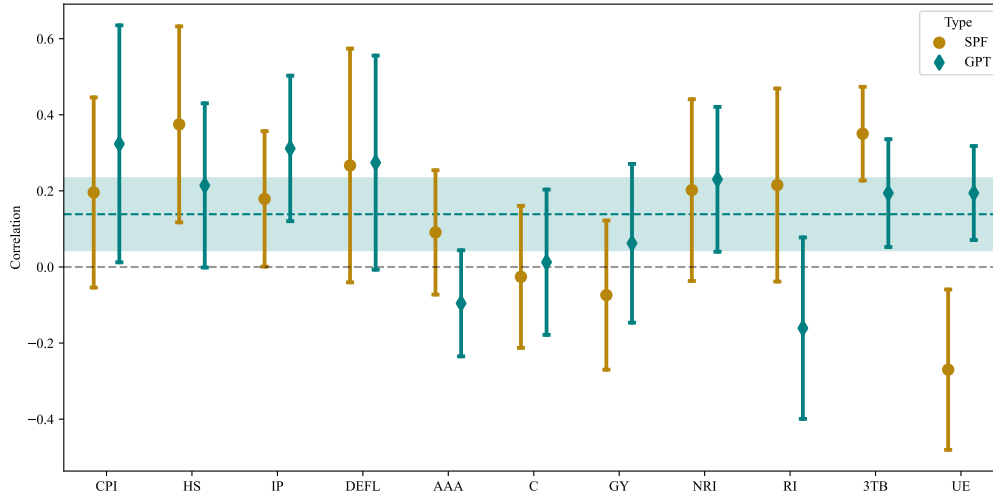
| | SPF | | GPT | |
	Corr.	T-Stat.	Corr.	T-Stat.
Panel	0.12	1.31	0.14	2.39
CPI	0.20	1.28	0.32	1.70
Housing Starts	0.37	2.39	0.21	1.63
Industrial Production	0.18	1.65	0.31	2.67
GDP Deflator	0.27	1.42	0.27	1.60
AAA Corporate Bond Yield	0.09	0.91	-0.10	-1.12
Consumption Growth	-0.03	-0.23	0.01	0.11
GDP Growth	-0.07	-0.62	0.06	0.49
Nonresidential Investment	0.20	1.39	0.23	1.99
Residential Investment	0.22	1.39	-0.16	-1.11
3-Month Treasury Bill	0.35	4.67	0.19	2.25
Unemployment Rate	-0.27	-2.10	0.19	2.58

Note. Reports the CG coefficients for SPF revisions and GPT expectations. Additionally reports 90% confidences intervals for the coefficients. Dashed teal line reports the panel CG coefficient and shaded teal band the corresponding 90% confidence interval. Standard errors for the single variable regressions are Newey-West, panel standard errors are Driscoll-Kraay. Bottom table reports the corresponding correlation coefficients and *t*-stats.

5 Memorization or Generalization?

Given GPT's training period overlaps with my sample period I next evaluate whether GPT's expectations result from data leakage or memorization of the training sample. The concern about memorization is different from standard concerns for overfitting. LLMs are trained to predict the next token in a sequence and direct overfitting here would indicate prompts in the format I specify here exist in the training sample which seems unlikely. However, GPT's training corpus could potentially contain information and discussions about the future values of the variables I'm interested in which it recites when prompted. To address this concern I run two tests. First, I extend my sample of *WSJ* articles to include articles from after GPT's training period – after September of 2021. Second, I test whether GPT's expectations are driven by revisions in the SPF forecasts or by the realized values of the variables.

To evaluate GPT's out-of-sample expectations, I scrape all available *WSJ* articles from the *WSJ* archive from September 2021 to March 2023. I then apply the cleaning procedure detailed in appendix A.2 and query GPT for its expectation for each article as with the main sample. AAII return expectations are released weekly, and as a result this gives me 79 observations outside of GPT's training period. For the SPF only five quarters of data are available – as a result I run a panel regression of revisions in SPF expectations on GPT's expectations for the full set of series studied above. Table 3 reports the results for these tests and indicates that the significant correlation I observe for my main sample is not the result of memorization.

Table 3: Correlation with GPT OOS

	AAII (Weekly)	SPF Avg.	SPF 1	SPF 2	SPF 3	SPF 4
Correlation	0.46	0.61	0.49	0.64	0.57	0.53
T-Stat	[4.54]	[6.11]	[4.50]	[6.61]	[5.45]	[4.93]
N	79	65	65	65	65	65

Note. Reports the correlation between GPT's out-of-sample expectations and various benchmarks. Each SPF column reports a different horizon and standard errors are Driscoll-Kraay.

Next, I evaluate whether GPT's expectations are driven by revisions in the SPF forecasts or by the realized values of the variables. To do this I run a panel regression of revisions in SPF expectations on GPT's expectations for the full set of series studied above. Table 4 reports the results for these tests and indicates that the significant correlation I observe for my main sample is not the result of GPT having knowledge about future realizations beyond the SPF expectations.

Table 4: Correlation between GPT and SPF revisions and realized values

Horizon	Avg.	1	2	3	4
SPF Revision	0.25	0.20	0.17	0.16	0.14
	[11.65]	[8.92]	[7.92]	[7.02]	[6.19]
Realized	0.02	0.01	0.02	0.02	0.01
	[1.80]	[0.47]	[1.44]	[1.36]	[0.97]
R^2	6.71	4.07	3.27	2.53	1.95

Note. Reports the correlation between GPT's expectations and SPF revisions and realized values. Standard errors are Driscoll-Kraay.

6   Conclusion

In this paper I introduce a new methodology to form expectation proxies and study beliefs using large language models. I find that the resulting expectation proxies exhibit many of the same deviations from full-information rational expectations as their current survey counterparts. These results provide new evidence to help guide the development of models of expectations formation.

These facts impose new constraints on models of beliefs depending on the origins of these deviations. If these deviations emerge from the estimated LLM itself then further work is needed to study the static behavior of these methods. On the other hand, if the deviations emerge from the *WSJ* articles themselves then further study of the narrative dynamics of these articles is needed. Overall, there is still much to learn about LLMs as human simulacra but the results here indicate the potential of these methods to open a new agenda of study.

References

Afrouzi, Hassan, Spencer Yongwook Kwon, Augustin Landier, Yueran Ma, and David Thesmar, 2020, Overreaction in expectations: Evidence and theory, *Available at SSRN 3709548* .

Aher, Gati, Rosa I Arriaga, and Adam Tauman Kalai, 2022, Using large language models to simulate multiple humans, *arXiv preprint arXiv:2208.10264* .

Alekseev, Georgij, Stefano Giglio, Quinn Maingi, Julia Selgrad, and Johannes Stroebel, 2022, A Quantity-Based Approach to Constructing Climate Risk Hedge Portfolios, *NBER* .

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart, 2021, Narratives about the macroeconomy .

Angeletos, George-Marios, Zhen Huo, and Karthik A. Sastry, 2021, Imperfect Macroeconomic Expectations: Evidence and Theory, *NBER Macroeconomics Annual* .

Argyle, Lisa P., Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate, 2023, Out of One, Many: Using Language Models to Simulate Human Samples, *Political Analysis* 1–15.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2015, X-CAPM: An extrapolative capital asset pricing model, *Journal of Financial Economics* 115, 1–24.

Barsky, Robert B, and J Bradford De Long, 1993, Why does the stock market fluctuate?, *The Quarterly Journal of Economics* 108, 291–311.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, 2021, On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? xn–cc6c, in *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623 (Association for Computing Machinery, New York, NY, USA).

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer, 2020, Overreaction in macroeconomic expectations, *American Economic Review* 110, 2748–82.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2018, Diagnostic expectations and credit cycles, *The Journal of Finance* 73, 199–227.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2022, Salience, *Annu. Rev. Econ.* 14, 521–544.

Brand, James, Ayelet Israeli, and Donald Ngwe, 2023, Using gpt for market research, *Available at SSRN 4395751* .

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei, 2020, Language Models are Few-Shot Learners, *arXiv* .

Brunnermeier, Markus, Emmanuel Farhi, Ralph S. J. Koijen, Arvind Krishnamurthy, Sydney C. Ludvigson, Hanno Lustig, Stefan Nagel, and Monika Piazzesi, 2021, Review Article: Perspectives on the Future of Asset Pricing, *Rev. Financ. Stud.* 34, 2126–2160.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang, 2023, Sparks of Artificial General Intelligence: Early experiments with GPT-4, *arXiv* .

Bybee, Leland, Bryan T Kelly, Asaf Manela, and Dacheng Xiu, 2021, Business news and business cycles .

Bybee, Leland, Bryan T Kelly, and Yinan Su, 2022, Narrative asset pricing: Interpretable systematic risk factors from news text, *Johns Hopkins Carey Business School Research Paper* .

Chomsky, Noam, 1959, A review of bf skinner's verbal behavior, *Language* 35, 26–58.

Coibion, Olivier, and Yuriy Gorodnichenko, 2012, What Can Survey Forecasts Tell Us about Information Rigidities?, *Journal of Political Economy* .

Coibion, Olivier, and Yuriy Gorodnichenko, 2015, Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts, *Am. Econ. Rev.* 105, 2644–78.

Cutler, David M, James M Poterba, Lawrence H Summers, et al., 1990, Speculative dynamics and the role of feedback traders, *American Economic Review* 80, 63–68.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Farmer, Leland, Emi Nakamura, and Jón Steinsson, 2021, Learning about the long run, Technical report, National Bureau of Economic Research.

Flynn, Joel P, and Karthik Sastry, 2022, The macroeconomics of narratives, *Available at SSRN 4140751* .

Fuster, Andreas, David Laibson, and Brock Mendel, 2010, Natural Expectations and Macroeconomic Fluctuations, *J. Econ. Perspect.* 24, 67–84.

Gabaix, Xavier, and Ralph S. J. Koijen, 2021, In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis, *NBER* .

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2019, Five Facts about Beliefs and Portfolios, *NBER* .

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of Returns and Expected Returns, *Review of Financial Studies* 27.

Horton, John J., 2023, Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?, *NBER* .

Jin, Lawrence J, and Pengfei Sui, 2022, Asset pricing with return extrapolation, *Journal of Financial Economics* 145, 273–295.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny, 1994, Contrarian investment, extrapolation, and risk, *The journal of finance* 49, 1541–1578.

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, Aggregate Wealth, and Expected Stock Returns, *Journal of Finance* 56, 815–849.

Lochstoer, Lars A., and Tyler Muir, 2022, Volatility Expectations and Returns, *Journal of Finance* 77, 1055–1096.

Lucas, Robert E., 1972, Expectations and the neutrality of money, *J. Econom. Theory* 4, 103–124.

Lucas, Robert E, 1976, Econometric policy evaluation: A critique, in *Carnegie-Rochester conference series on public policy*, volume 1, 19–46, North-Holland.

Michalopoulos, Stelios, and Melanie Meng Xue, 2021, Folklore, *Q. J. Econ.* 136, 1993–2046.

Muth, John F, 1961, Rational expectations and the theory of price movements, *Econometrica: Journal of the Econometric Society* 315–335.

Nagel, Stefan, and Zhengyang Xu, 2022a, Asset Pricing with Fading Memory, *Rev. Financ. Stud.* 35, 2190–2245.

Nagel, Stefan, and Zhengyang Xu, 2022b, Dynamics of Subjective Risk Premia, *NBER* .

Sargent, Thomas J, 2001, *The conquest of American inflation* (Princeton University Press).

Schramowski, Patrick, Cigdem Turan, Nico Andersen, Constantin A. Rothkopf, and Kristian Kersting, 2022, Large pre-trained language models contain human-like biases of what is right and wrong to do, *Nat. Mach. Intell.* 4, 258–268.

Shiller, Robert J., 2019, *Narrative Economics* (Princeton University Press, Princeton, NJ, USA).

Sims, Christopher A, 1980, Macroeconomics and reality, *Econometrica: journal of the Econometric Society* 1–48.

Skinner, BF, 1957, *Verbal behavior* (Copley Publishing Group).

Touvron, Hugo, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample, 2023, LLaMA: Open and Efficient Foundation Language Models, *arXiv* .

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, 2017, Attention Is All You Need, *arXiv* .

Wei, Jason, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus, 2022, Emergent Abilities of Large Language Models, *arXiv* .

Welch, Ivo, and Amit Goyal, 2008, A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, *Rev. Financ. Stud.* 21, 1455–1508.

Wu, Shijie, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann, 2023, BloombergGPT: A Large Language Model for Finance, *arXiv* .

A   Constructing the *WSJ* Samples

A.1   Constructing the *WSJ* 1984-2021 Sample

I conduct data processing steps in the following order:

1. Remove all articles prior to January 1984 and after December 2021.

2. Exclude articles with page-citation tags corresponding to any sections other than A, B, C, or missing.

3. Exclude articles corresponding to weekends.

4. Exclude articles with subject tags associated with obviously non-economic content such as sports. List of exclusions available from authors on request.

5. Exclude articles with the certain headline patterns (such as those associated with data tables or those corresponding to regular sports, leisure, or books columns). List of exclusions available from authors on request.

6. Exclude articles with less than 100 words.

7. Exclude articles with headlines less than 10 words.

8. Sample 300 articles for each month of data.

A.2   Constructing the *WSJ* Oct 2021-Mar 2023 Sample

I conduct the following data processing steps in the following order:

1. Exclude articles with headlines less than 8 words.

2. Exclude articles not in the following sections: "U.S", "BUSINESS", "WORLD", "POL-ITICS", "WSJ NEWS EXCLUSIVE", "HEARD ON THE STREET", "FINANCE", "EARNINGS", "MARKETS", "U.S. MARKETS", "U.S. ECONOMY", "ECONOMY".

B   Coocurrence of GPT Responses

This section reports the coocurrence between the article level increase/decrease responses for the GPT expectations series studied.

Increase (rows) × Increase (columns)

	SNP	CPI	HS	IP	DEFL	AAA	C	GF	GY	NRI	RI	GS	3TB	UE
SNP	--	0.04	0.02	0.07	0.06	0.07	0.07	0.05	0.11	0.11	0.06	0.07	0.07	0.00
CPI	0.04	--	0.01	0.03	0.05	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.05	0.01
HS	0.02	0.01	--	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.00
IP	0.07	0.03	0.02	--	0.05	0.04	0.07	0.04	0.09	0.09	0.05	0.06	0.04	0.00
DEFL	0.06	0.05	0.02	0.05	--	0.05	0.06	0.05	0.07	0.07	0.05	0.06	0.06	0.01
AAA	0.07	0.04	0.02	0.04	0.05	--	0.05	0.04	0.07	0.07	0.04	0.05	0.07	0.01
C	0.07	0.04	0.02	0.07	0.06	0.05	--	0.05	0.10	0.09	0.06	0.07	0.04	0.00
GF	0.05	0.04	0.02	0.04	0.05	0.04	0.05	--	0.06	0.06	0.05	0.07	0.05	0.01
GY	0.11	0.04	0.02	0.09	0.07	0.07	0.10	0.06	--	0.14	0.08	0.09	0.06	0.00
NRI	0.11	0.04	0.02	0.09	0.07	0.07	0.09	0.06	0.14	--	0.08	0.09	0.06	0.00
RI	0.06	0.03	0.02	0.05	0.05	0.04	0.06	0.05	0.08	0.08	--	0.06	0.04	0.00
GS	0.07	0.04	0.02	0.06	0.06	0.05	0.07	0.07	0.09	0.09	0.06	--	0.06	0.00
3TB	0.07	0.05	0.01	0.04	0.06	0.07	0.04	0.05	0.06	0.06	0.04	0.06	--	0.02
UE	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.02	--

Increase (rows) × Decrease (columns)

	SNP	CPI	HS	IP	DEFL	AAA	C	GF	GY	NRI	RI	GS	3TB	UE
SNP	--	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05
CPI	0.02	--	0.01	0.02	0.01	0.01	0.02	0.01	0.02	0.02	0.02	0.02	0.00	0.03
HS	0.00	0.00	--	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
IP	0.00	0.00	0.00	--	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.06
DEFL	0.01	0.00	0.01	0.01	--	0.01	0.01	0.00	0.01	0.02	0.01	0.01	0.00	0.04
AAA	0.02	0.00	0.01	0.01	0.01	--	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.03
C	0.00	0.01	0.00	0.00	0.01	0.01	--	0.00	0.00	0.00	0.00	0.00	0.01	0.06
GF	0.02	0.00	0.01	0.01	0.01	0.01	0.01	--	0.02	0.02	0.01	0.01	0.01	0.04
GY	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.00	--	0.00	0.00	0.00	0.01	0.09
NRI	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	--	0.00	0.00	0.01	0.08
RI	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	--	0.00	0.01	0.05
GS	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	--	0.01	0.06
3TB	0.06	0.01	0.02	0.03	0.02	0.02	0.04	0.02	0.04	0.05	0.04	0.03	--	0.03
UE	0.07	0.02	0.03	0.06	0.04	0.04	0.06	0.04	0.07	0.07	0.06	0.06	0.03	--

Decrease (rows) × Increase (columns)

	SNP	CPI	HS	IP	DEFL	AAA	C	GF	GY	NRI	RI	GS	3TB	UE
SNP	--	0.02	0.00	0.00	0.01	0.02	0.00	0.02	0.00	0.00	0.00	0.02	0.06	0.07
CPI	0.00	--	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.01	0.02
HS	0.00	0.01	--	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.02	0.03
IP	0.00	0.02	0.00	--	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.03	0.06
DEFL	0.01	0.01	0.00	0.01	--	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.04
AAA	0.01	0.01	0.00	0.00	0.01	--	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.04
C	0.00	0.02	0.00	0.00	0.01	0.02	--	0.01	0.00	0.00	0.00	0.01	0.04	0.06
GF	0.00	0.01	0.00	0.00	0.00	0.01	0.00	--	0.00	0.00	0.00	0.00	0.02	0.04
GY	0.00	0.02	0.00	0.00	0.01	0.02	0.00	0.02	--	0.00	0.00	0.01	0.04	0.07
NRI	0.00	0.02	0.00	0.00	0.02	0.02	0.00	0.02	0.00	--	0.00	0.01	0.05	0.07
RI	0.00	0.02	0.00	0.00	0.01	0.02	0.00	0.01	0.00	0.00	--	0.01	0.04	0.06
GS	0.00	0.02	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	--	0.03	0.06
3TB	0.01	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	--	0.03
UE	0.05	0.03	0.02	0.06	0.04	0.03	0.06	0.04	0.09	0.08	0.05	0.06	0.03	--

Decrease (rows) × Decrease (columns)

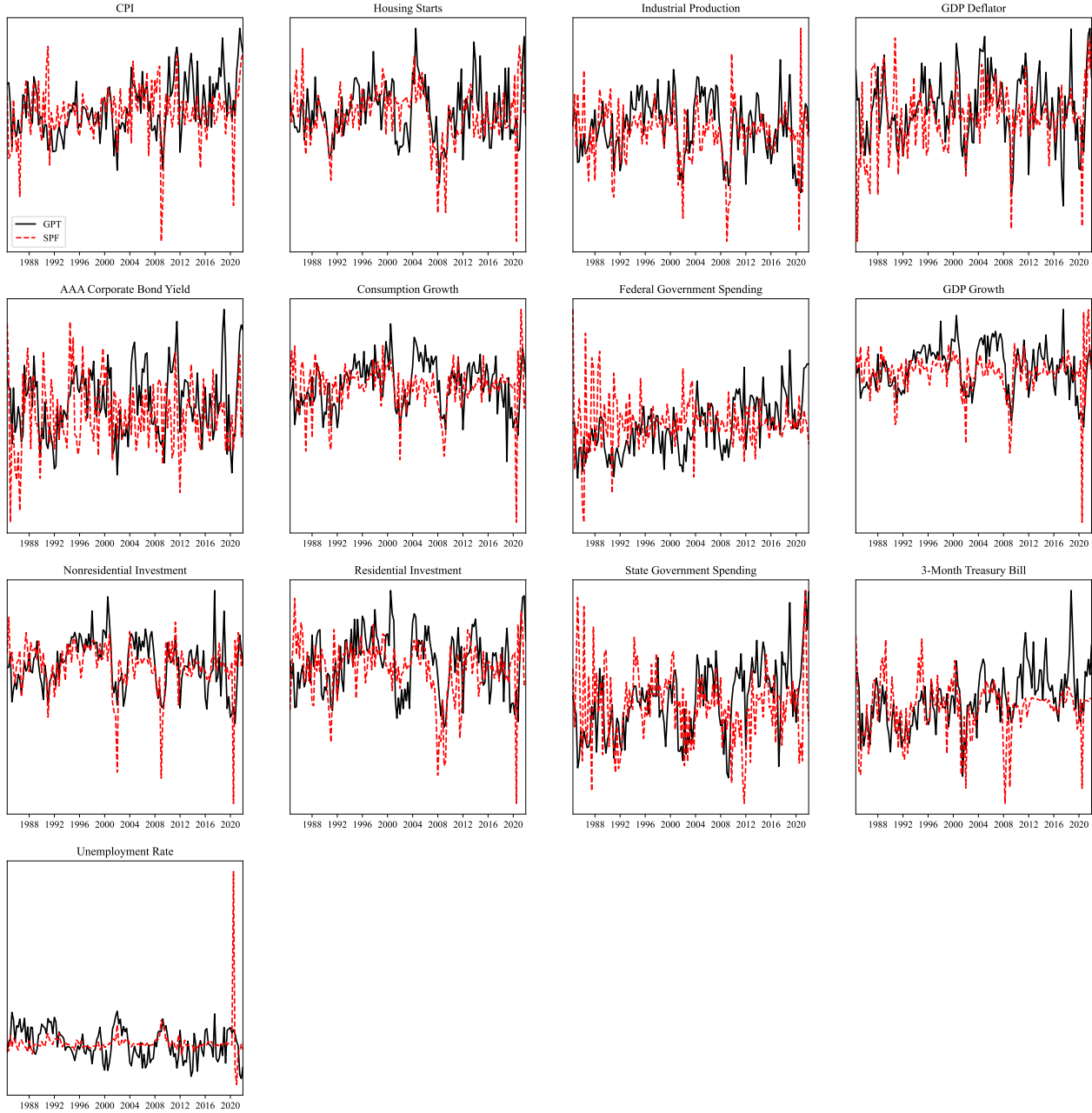|	SNP	CPI	HS	IP	DEFL	AAA	C	GF	GY	NRI	RI	GS	3TB	UE
SNP	--	0.05	0.05	0.10	0.10	0.11	0.14	0.08	0.16	0.17	0.13	0.12	0.08	0.00
CPI	0.05	--	0.02	0.04	0.05	0.05	0.05	0.04	0.05	0.05	0.04	0.05	0.04	0.00
HS	0.05	0.02	--	0.04	0.03	0.03	0.05	0.03	0.05	0.05	0.05	0.04	0.02	0.00
IP	0.10	0.04	0.04	--	0.07	0.06	0.09	0.06	0.10	0.10	0.09	0.08	0.04	0.00
DEFL	0.10	0.05	0.03	0.07	--	0.08	0.08	0.06	0.10	0.09	0.08	0.08	0.06	0.00
AAA	0.11	0.05	0.03	0.06	0.08	--	0.08	0.06	0.10	0.10	0.08	0.08	0.07	0.00
C	0.14	0.05	0.05	0.09	0.08	0.08	--	0.08	0.14	0.14	0.11	0.11	0.06	0.00
GF	0.08	0.04	0.03	0.06	0.06	0.06	0.08	--	0.08	0.08	0.07	0.08	0.05	0.00
GY	0.16	0.05	0.05	0.10	0.10	0.10	0.14	0.08	--	0.16	0.13	0.12	0.07	0.00
NRI	0.17	0.05	0.05	0.10	0.09	0.10	0.14	0.08	0.16	--	0.14	0.12	0.07	0.00
RI	0.13	0.04	0.05	0.09	0.08	0.08	0.11	0.07	0.13	0.14	--	0.10	0.06	0.00
GS	0.12	0.05	0.04	0.08	0.08	0.08	0.11	0.08	0.12	0.12	0.10	--	0.06	0.00
3TB	0.08	0.04	0.02	0.04	0.06	0.07	0.06	0.05	0.07	0.07	0.06	0.06	--	0.00
UE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	--

Note. Reports the cooccurence proportion of each indicator for different surveyed series.

C   Time-Series of GPT/SPF Expectations

This section reports time series plots the GPT expectations and average revisions for the SPF series studied in section 4.

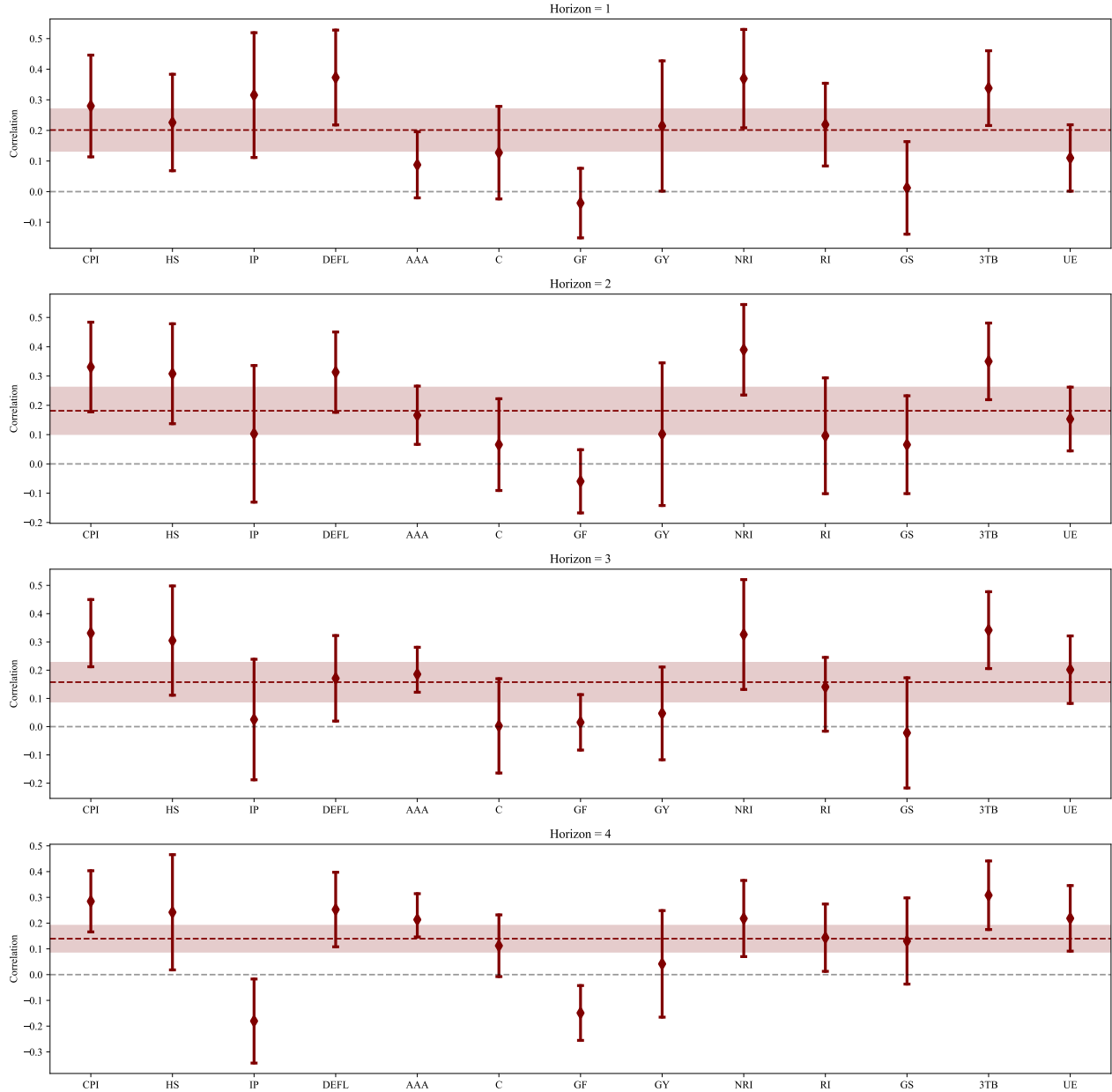Figure 8: Time Series of GPT Expectations and SPF Revisions

Note. Reports the time series of GPT expectations (black) and corresponding SPF revisions (red).

D   GPT/SPF Correlations Over Different Horizons

This section reports results comparable to section 4 but partitioned over different horizons.

Figure 9: SPF Revision Correlation Across Horizons

Note. Reports the correlation between revisions of SPF expectations and GPT expectations for different horizons. Additionally reports 90% confidences intervals for the correlation coefficients. Dashed maroon line reports the panel correlation coefficient and shaded maroon band the corresponding 90% confidence interval. Standard errors for the single variable regressions are Newey-West, panel standard errors are Driscoll-Kraay.

Figure 10: Coibion-Gorodnichenko Coefficients Across Horizons
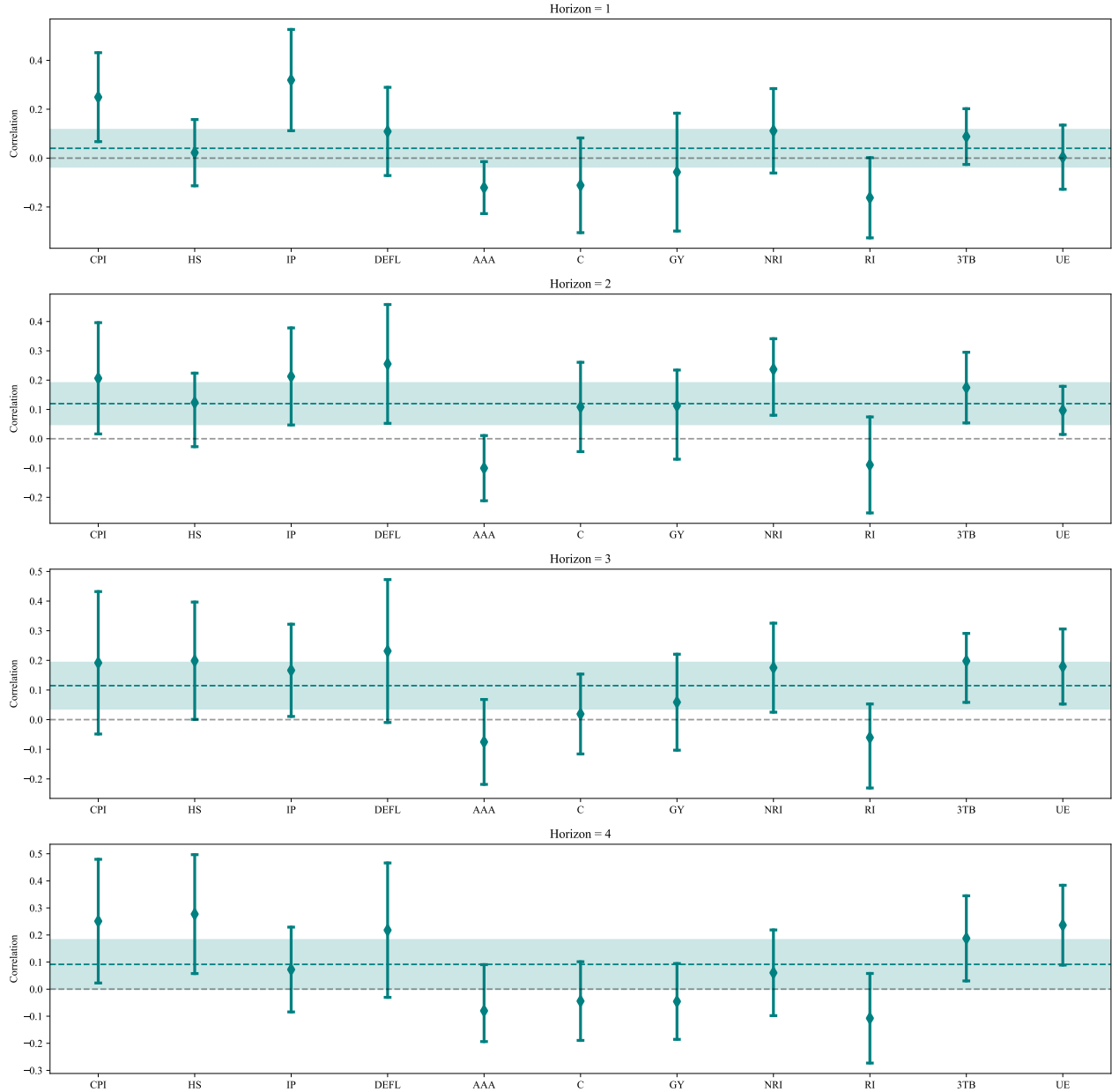
Note. Reports the CG coefficients for SPF revisions and GPT expectations across different horizons. Additionally reports 90% confidences intervals for the coefficients. Dashed teal line reports the panel CG coefficient and shaded teal band the corresponding 90% confidence interval. Standard errors for the single variable regressions are Newey-West, panel standard errors are Driscoll-Kraay.

E Covid

This section reports results comparable to 4 but excludes all data after 2019 to avoid the effects of the Covid-19 pandemic.

Figure 11: GPT/SPF Correlations (Pre 2019)
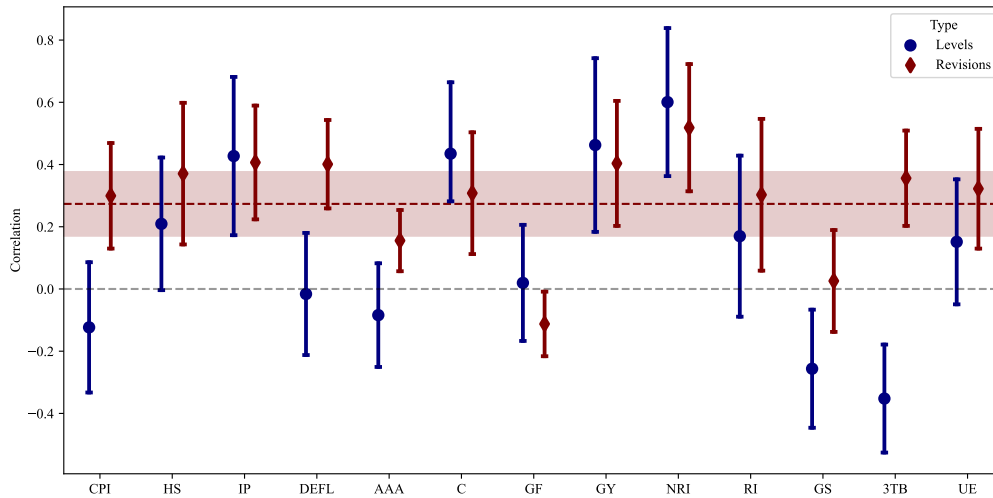
	Levels		Revisions	
	Corr.	T-Stat.	Corr.	T-Stat.
Panel	0.07	1.23	0.27	4.35
CPI	-0.12	-0.97	0.30	2.90
Housing Starts	0.21	1.61	0.37	2.67
Industrial Production	0.43	2.75	0.41	3.65
GDP Deflator	-0.02	-0.13	0.40	4.63
AAA Corporate Bond Yield	-0.08	-0.83	0.16	2.59
Consumption Growth	0.43	3.11	0.31	2.58
Federal Government Spending	0.02	0.17	-0.11	-1.78
GDP Growth	0.46	2.72	0.40	3.30
Nonresidential Investment	0.60	4.14	0.52	4.16
Residential Investment	0.17	1.08	0.30	2.04
State Government Spending	-0.26	-2.22	0.03	0.26
3-Month Treasury Bill	-0.35	-3.33	0.36	3.81
Unemployment Rate	0.15	1.24	0.32	2.75

Note. Reports the correlation between levels and revisions of SPF expectations and GPT expectations. Additionally reports 90% confidences intervals for the correlation coefficients. Dashed maroon line reports the panel correlation coefficient and shaded maroon band the corresponding 90% confidence interval. Standard errors for the single variable regressions are Newey-West, panel standard errors are Driscoll-Kraay. Bottom table reports the corresponding correlation coefficients and *t*-stats.

Figure 12: Coibion-Gorodnichenko Regressions (Pre 2019)
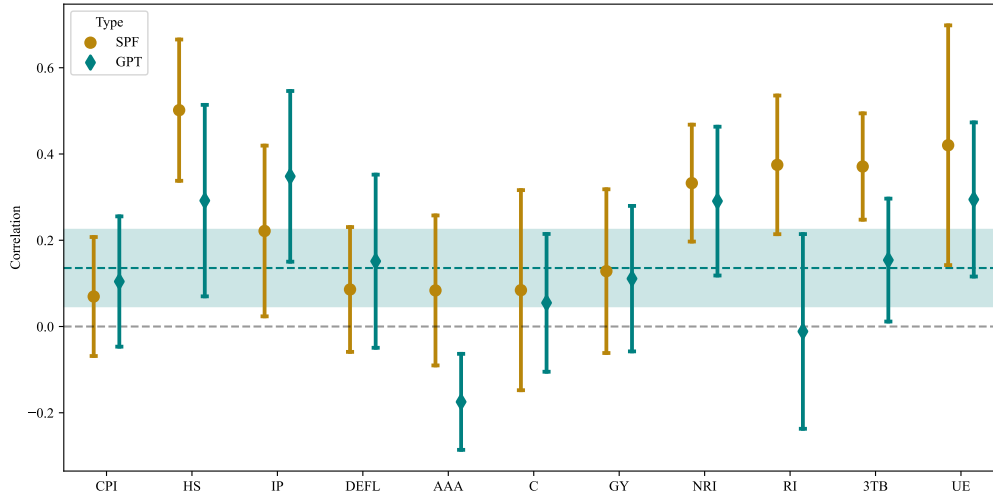
| | SPF | | GPT | |
	Corr.	T-Stat.	Corr.	T-Stat.
Panel	0.21	4.19	0.14	2.50
CPI	0.07	0.83	0.10	1.13
Housing Starts	0.50	5.02	0.29	2.16
Industrial Production	0.22	1.84	0.35	2.89
GDP Deflator	0.09	0.97	0.15	1.24
AAA Corporate Bond Yield	0.08	0.79	-0.17	-2.57
Consumption Growth	0.08	0.60	0.05	0.56
GDP Growth	0.13	1.11	0.11	1.08
Nonresidential Investment	0.33	4.02	0.29	2.76
Residential Investment	0.37	3.82	-0.01	-0.08
3-Month Treasury Bill	0.37	4.94	0.15	1.77
Unemployment Rate	0.42	2.48	0.29	2.70

Note. Reports the CG coefficients for SPF revisions and GPT expectations. Additionally reports 90% confidences intervals for the coefficients. Dashed teal line reports the panel CG coefficient and shaded teal band the corresponding 90% confidence interval. Standard errors for the single variable regressions are Newey-West, panel standard errors are Driscoll-Kraay. Bottom table reports the corresponding correlation coefficients and *t*-stats.