Overreaction and horizon: long-term expectations overreact, but short-term expectations drive fluctuations *

Basil Halperin
Stanford

J. Zachary Mazlish
Oxford and GPI

January 2025

Abstract

Across a large cross-country panel of surveyed macroeconomic expectations, we establish four facts about average macroeconomic expectations: 1) Less than one-year ahead expectations under-revise. 2) Two or more year ahead expectations over-revise. 3) All horizon expectations tend to be too extreme. 4) Over-revision and over-extremity increase in the horizon of the forecast. These facts hold across advanced and emerging economies, and across a host of different macroeconomic variables. We then show that despite long-term expectations overreacting more, it is short-term expectations which are most strongly associated with ”booms and busts” in investment, GDP, and the stock market.

* *Email:* `basilh@stanford.edu, john.mazlish@economics.ox.ac.uk`
We are grateful for generous financial support to the George and Obie Shultz Fund at MIT and to the Institute for Humane Studies under grant number 017435.

1 Introduction

At this point in the history of economics, it is uncontroversial that ”rational expectations” is a Panglossian hypothesis. Yet, given the vast array of ways agents’ expectation formation *could* differ from rational expectations, it remains unclear which departures from rationality are most systematic.

In this paper, we show four facts about forecasters’ *average* macroeconomic expectations that are robust across different macroeconomic variables and a broad swathe of advanced and emerging market economies:

(i) When forecasting macroeconomic outcomes less than one-year in the future, average expectations tend to *under*-revise.

(ii) When forecasting outcomes two or more years in the future, average expectations tend to *over*-revise.

(iii) At all forecast horizons, expectations tend to be too extreme — unusually high forecasts are too high, and unusually low forecasts are too low.

(iv) The tendency to over-revise and the tendency to be too extreme both are stronger the further away in the future is the outcome that is being forecast. That is, overreaction increases in the horizon of the forecast.

The robustness of these facts — across different country contexts, different macroeconomic variables, and different time-periods — make them features that any model of average macroeconomic expectations that wants to be consistent with the data should match. Importantly, a number of popular recent models of agents’ expectation formation are inconsistent with fact (4), that over-reaction is increasing in forecasting horizon.

How do these properties of average macroeconomic expectations influence the business cycle? A recent literature has found that, in the US, long-run expectations are most strongly associated with subsequent business cycle horizon real and financial outcomes (Bordalo, Gennaioli, Porta, and Shleifer 2024, Bordalo, Gennaioli, Porta, OBrien, et al. 2024). In contrast to this literature, we find that, in our broad cross-country sample, *short-run* expectations are most predictive of business cycle frequency outcomes.

Our expectations data comes from Consensus Economics’ ”long-term forecasts”. The participants in the Consensus survey are economists who work at banks and consultancies. An important question is whether or not the surveyed expectations of these professional forecasters correspond to the expectations of actual economic agents.

Reassuringly, our findings about which type of expectations are consistent with the data (and which type of expectations are not) align with the recent experimental work of Afrouzi et al. (2023), which surveys Mechanical Turkers and college students, not professional economic forecasters. Furthermore, the findings in Afrouzi et al. (2023) consider expectations about abstract AR(1) processes. Relative to their paper, our contribution is to show that the same key properties agents’ expectations hold in a stylized experimental setting appear in actual macroeconomic expectations.

We now explain each of the paper’s main empirical findings and explicitly situate them relative to the literature.

Fact 1: Less than one-year ahead average expectations under-revise. In the language of Angeletos, Huo, and Sastry (2021), this fact is the "CG fact", since it first came to widespread attention in Coibion and Gorodnichenko (2015). This fact is consistent with noisy information models where the average forecast under-reacts not due to agents *irrationality*, but due to noise in individual agents' signals (Woodford 2001, Sims 2003). Since CG also use Consensus data, it is unsurprising that we find similar evidence of under-reaction. The difference between our data and theirs is ours covers 89 countries and includes annual forecasts out to ten-years in the future. In Coibion and Gorodnichenko (2015), they use *short-term* Consensus forecasts which cover 12 advanced economies and are forecasts of quarterly outcomes out to only six-quarters in the future. As facts (2) and (4) reveal, the horizon of the forecast importantly determines how the forecast departs from full-information rational expectations (FIRE).

Fact 2: Average forecasts of outcomes two or more years in the future *over*-revise. Leveraging the fact that in our data forecast horizons extend out to ten years in the future, we show that the CG under-revision fact "flips" to over-revision at two-year or longer horizons. This finding is consistent with the fact that individual forecasters tend to over-revise (Bordalo, Gennaioli, Ma, et al. 2020) and that individual over-revision combined with noisy signals causes the average forecast to flip from under-reacting to over-reacting (Angeletos, Huo, and Sastry 2021).[1] Bordalo, Gennaioli, Porta, and Shleifer 2024)(BGLS) also find that longer horizon *average* forecasts over-revise. Relative to BGLS, our contribution on fact (2) is twofold: 1) BGLS use just US data, while our survey data includes 89 countries. 2) BGLS focus on earnings-growth expectations of stock-analysts, while our results are for real (GDP, investment, consumption) and nominal (inflation) macroeconomic variables.

Fact 3: At all forecast horizons, macroeconomic expectations tend to be too extreme. The english language word "overreaction" can be mapped to multiple different formal properties of expectations. One sense is over-revision, detected by regressing forecast errors on forecast revisions. Another sense comes from regressing forecast errors on the lagged forecast. This regression shows whether when forecasts are "extreme" (high or low relative to the typical forecast) they tend to be rationally extreme, not extreme enough, or too extreme. We find that forecasts tend to be too extreme at all forecast horizons. As far as we are aware, this fact is novel, though other patterns of overreaction related to this one have been documented (Bordalo, Gennaioli, Porta, and Shleifer 2024, Kohlhas and Walther 2021).[2]

Fact 4: Longer horizon macroeconomic forecasts overreact more, in both senses —

[1]The co-existence of over and under-reaction depending on the setting has begun to be explained at a more cognitive level by work like Augenblick, Lazarus, and Thaler (2025) and Ba, Bohren, and Imas (2024). Connecting our findings with those papers would imply that longer horizon forecasts feature overreaction because they are more dependent on weak and noisy signals. We think this is a promising connection that should be explored more in future work.

[2]Specifically, Bordalo, Gennaioli, Porta, and Shleifer (2024) show that five-year ahead stock-analysts' earnings growth forecasts are too extreme, in this same sense, but do not show the result for different forecast horizons.

they over-revise more, and have more of a tendency to be too extreme. Versions of the fact that longer-horizon expectations overreact more are found in a number of recent papers (Bordalo, Gennaioli, Porta, and Shleifer 2024, Angeletos, Huo, and Sastry 2021, d'Arienzo 2020, and Afrouzi et al. 2023). The key property that emerges in our data is that both senses of overreaction continue to increase out to a ten-year forecast horizon. This matches the experimental findings of Afrouzi et al. (2023), and importantly, discriminates against the sort of models found in Bordalo, Gennaioli, Porta, and Shleifer (2024) and Angeletos, Huo, and Sastry (2021).

In both of the latter models, overreaction increases at medium-horizons as the influence of information noise (fact 1) fades out, but then decreases back towards zero at longer horizons. That is, there is no overreaction in beliefs infinitely far out into the future. Our results only speak to beliefs ten-years out in the future, but since overreaction is monotonically increasing out to ten years in our data, our findings fit more naturally with the framework of Afrouzi et al. (2023), where it is beliefs about the *long-run mean* that overreact.

The primary contribution of our paper is establishing that these four facts hold across four different macroeconomic variables — GDP growth, inflation, investment growth, and consumption growth. The basic pattern of the four facts emerges even if we split the sample into advanced and emerging market economies, split the sample in half, or remove the Great Financial Crisis (GFC). Furthermore, we show that using our forecasts helps with out-of-sample forecasting, and helps more the longer is the forecast horizon. This stands in contrast to the findings of Eva and Winkler (2023) about less than one-year expectations, and instills further confidence that the four facts we focus on are robust features of macroeconomic expectations across different regimes and information settings.

In addition to these features of expectations, we also investigate which expectations are associated with subsequent macroeconomic and financial outcomes. We first run local projections which control for a host of lagged macroeconomic variables to identify the effects of changing macroeconomic expectations on investment and GDP growth over the business cycle. We find evidence consistent with Bordalo, Gennaioli, Porta, OBrien, et al. (2024) that upward revisions in expectations are associated with short-term (< 2 year) "booms" and sharp "busts" right afterwards. However, unlike in Bordalo, Gennaioli, Porta, OBrien, et al. (2024), it is changes in *short-term* expectations that are most strongly associated with ensuing booms-and-busts. The pattern of the expansion and reversal of real activity matches the typical business cycle characteristics found in Angeletos, Collard, and Dellas (2020), and the fact that such a business-cycle arises in response to under-then over-reacting expecations fits with recent work such as Angeletos, Collard, and Dellas (2018), Bianchi, Ilut, and Saijo (2024), L'Huillier, Singh, and Yoo (2024), Cai (2024), and Bardóczy and Guerreiro (2023).

Next, we show that high GDP growth expectations predict weak stock-market returns. Similar to our results about the real economy, we find that short-term growth expectations are better predictors of up to five-year ahead stock-market return reversals than high long-term expectations, in contrast to the opposite findings of Bordalo, Gennaioli, Porta, and Shleifer (2024). Our results can be reconciled with theirs since in our *US data* we match their finding that long-term expectations are better predictors of weak returns. The finding that belief overreaction is important for explaining stock market movements fits within a long tradition, including important recent contributions such as Bianchi, Ludvigson, and Ma (2024) and McCarthy and Hillenbrand (2021). The specific boom-bust pattern we find in response to positive shocks matches the model implications in Mei and Wu (2024), and connects to other work on how irrational expectations jointly explain stock-market and business-cycle outcomes, such as Adam, Marcet, and Beutel (2017) and Winkler (2020). Finally, these findings also relate to recent work on the properties of long-horizon earnings expectations (Sias, Starks, and Turtle 2024, H Décaire and Guenzel 2023), and could explain the premium on near-future cash-flows found by Gormsen and Lazarus (2023).

These findings about the importance of short-term expectations pose some tension with our earlier findings that it is long-term expectations which overreact most. The discrepancy is not necessarily a puzzle, since different models predict substantially different effects of long-term expectations on current activity (Beaudry and Portier 2014, Dupor and Mehkari 2014). Further, our results are consistent with models which dampen the importance of long-term expectations relative to short-term expectations in agents' decision-making, such as Angeletos and Lian (2018) and Angeletos and Huo (2021). We leave a full modeling of all these dynamics to future work.

The rest of the paper proceeds as follows. Section 2 introduces the data. Section 3 establishes our four facts about average macroeconomic expectations. Section 4 shows that short-term expectations are more associated with "boom-bust" cycles in the macroeconomy than long-term expectations. Section 5 shows the same pattern for stock-return predictability. Section 6 briefly concludes.

2 Data

The survey data we use comes from Consensus Economics. Consensus surveys economists working at banks and consultancies to elicit a host of different macroeconomic forecasts. We make use of their "long-term" forecasts, focusing on forecasts of GDP growth, inflation, consumption growth, and investment growth.

All of our data is aggregate data. We only have to access to the average forecast and the standard deviation of forecasts, not individual forecasts.

Using GDP growth as an example, Consensus long-term forecasts work as follows: forecasters are asked to provide a point estimate for what GDP growth in a given country will be in the current year, as well as the next five years. They are also asked to provide a "long-term" forecast which is what they think average annual GDP growth will be 6-10 years from the current date. So, if you were taking the Consensus survey on the US in October 2024, you would be asked what you think GDP growth will be each year 2024-29 and what you think average GDP growth will be 2030-34.

Our data covers 89 different countries, with different sample lengths for different countries. The longest samples are for the G7 countries, which stretch back to 1989. Prior to 2014, consensus issued its survey two times per year — at the beginning of April (start of Q2) and the beginning of October (start of Q4). Since 2014, the survey has been quarterly.

In total, we have 4240 observations of 0-10 year GDP forecasts, 4205 observations of inflation expectations, 3185 observations of consumption forecasts, and 3185 observations of investment forecasts. Appendix Table (6) provides summary statistics on our expectations data coverage by country.

To construct forecast errors which compare surveyed outcomes to realized outcomes we use the World Bank's World Development Indicators (WDI).

3 Forecast Biases

3.1 Testing for Overreaction

Our primary test for overreaction follows the approach in Bordalo, Gennaioli, Porta, and Shleifer (2024). We define forecast errors as $e_{i,t+j} = x_{i,t+j} - \mathbb{E}_t\left(x_{i,t+j}\right)$, for any macroeconomic variable of interest x. As explained more below, for all of the primary results, forecasts x are z-scored with respect to macro variable, country, and forecast horizon.

We then run panel regressions of the following form:

$$e_{i,t+j} = \alpha + \beta_1 \Delta\, \mathbb{E}_t\left(x_{i,t+j}\right) + \beta_2\, \mathbb{E}_{t-1}\left(x_{i,t+j}\right) + f_{i,t,x} + \epsilon_{i,t} \tag{1}$$

The revision in expectations is given by $\Delta\, \mathbb{E}_t\left(x_{i,t+j}\right)$, and the revision can either represent the change in average expectations between *consecutive* surveys, or the change in expectations between surveys that are one year apart. Correspondingly, the lagged expectation $\mathbb{E}_{t-1}\left(x_{i,t+j}\right)$ can either represent the previous survey's average expectation or the average expectation from the survey one-year prior to date t.

Coefficients β_1 and β_2 are both indicators of overreaction. For both coefficients, $\beta < 0$ indicates that the average forecast overreacts, while $\beta > 0$ indicates that the average forecast underreacts.

The intuition for β_1 is that if $\beta_1 < 0$ that implies that a positive forecast revision is associated with a negative forecast error. Therefore, the forecast should not have been revised up as much as it was, since the forecast is (on average) above the realization. If the forecast systematically over-revises, that is a form of overreaction. Testing for over or underreaction by focusing on the coefficient on forecast revisions was introduced by Coibion and Gorodnichenko (2015).

The intuition for β_2 is that if $\beta_2 < 0$ that implies that a higher *level* of the forecast is associated with the forecast being too high relative to the outcome. This means that (relatively) high forecasts are too high while (relatively) low forecasts are too low — which is the sense of overreaction we call being "too extreme."

Since forecast revisions and forecast errors are known at time t, finding $\beta_1 \neq 0$ or

$\beta_2 \neq 0$ is a violation of full information rational expectations (FIRE).

Fixed effects are denoted by $f_{i,t,x}$, with x representing the variable that is being forecast. For our main specification, we pool the different macro variables together and include entity fixed effects where an entity is a country i and macro variable x pair. However, as shown in section X, our results are robust to including no fixed-effects, using just time fixed-effects, or using time and entity fixed-effects. Given that we have a panel with longitudinal and cross-sectional error correlation, we use Driscoll and Kraay (1998) standard errors, with groupings by both variable and country.

Why use two senses of overreaction, rather than running separate regressions? There are two answers. One is that both capture different relevant senses of overreaction. $\beta_1 < 0$ corresponds to over-revision while $\beta_2 < 0$ corresponds to over-extremity. The other answer is that otherwise the lagged forecast may be an omitted variable, since forecast revisions are predicted by the lagged forecast.

Figure (1) displays the coefficient (and standard errors) from regressing forecast revisions on the lagged forecast, by forecast horizon. All primary forecast variables — GDP, inflation, investment, and consumption — are pooled together.

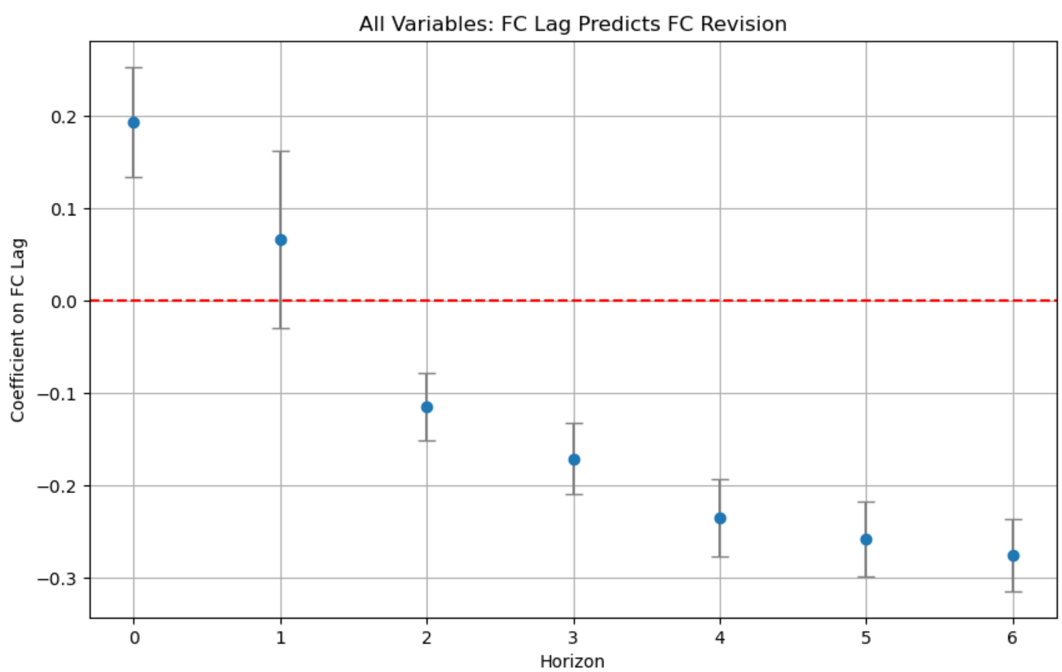


Figure 1: The regression coefficients in the figure are from the regression $\Delta\,\mathbb{E}_t(x_{i,t+j}) = \alpha + \beta\,\mathbb{E}_{t-1}(x_{i,t+j}) + \epsilon_{i,t}$, run separately for each horizon. GDP, inflation, consumption, and investment forecasts are pooled together. The forecast here is the forecast of average annualized growth (of that variable) between t and $t+j$, which is constructed by cumulating the individual-year growth forecasts. All forecasts are z-scored with respect to variable, country, and horizon, with lagged forecasts z-scored using an expanding window that drops the first 10 observations. Standard errors are Driscoll-Kraay with country-variable groupings.

The figure shows that there is significant momentum in current-year forecasts — a high-level of the forecast yesterday predicts upward revisions today — while there is significant reversal for two-year and beyond forecasts. That is, at longer horizons, a higher (lower) value of the forecast predicts you will revise down (up) your forecast, a sort of ”mean-reversion” in forecasts. Appendix Figure (10) shows that an almost identical relationship holds for each of the forecast variables GDP, inflation, consumption, and investment.

What does it mean if one coefficient is positive while the other is negative? That just means that overreaction in one sense happens, while underreaction in the other sense happens. There is no single formal definition of overreaction which captures all of the different intuitions the word corresponds to.

Why z-score forecasts? Z-scoring forecasts with respect to country and horizon prevents countries with greater volatility of forecast errors (and revisions) from dominating the results. It also prevents mean-biases from affecting the results.

For example, consider the case where the only two countries in the data are Venezuela and Switzerland, and we are running (1) on inflation expectations. Inflation expectations in Venezuela have average forecast errors and forecast revisions that are *far* more variable than in Switzerland, leading Venezuela to dominate the variance in the data and be correspondingly ”overweighted” in the coefficient estimates. Adding country fixed-effects would not alleviate this issue, since the fixed-effect would only adjust for the *average* forecast error in Venezuelan, but not the fact that the variability of Venezuelan forecast errors swamps that of Switzerland.

If we z-score expectations with respect to country and horizon, then the regression is telling us ”when expectations revise a large amount *relative to what is typical* for horizon H and country C, are forecast errors large or small *relative to what is typical* for horizon H and country C”? This corresponds better with an intuitive sense of what a ”systematic” tendency for over/underreaction would mean.

Furthermore, by z-scoring, the average (z-scored) forecast error in each country is automatically set to zero, thereby avoiding the possibility that finding under-reaction could stem from a mean-bias. In the case of US short-term interest rate forecasts, for most of the period 1981-2019 forecasters over-predicted the future level of short-term interest rates. Since interest rates mostly moved down over this period, this looks like a case of underreaction (forecasts revised down, but not as much as they should have). However, as shown in Farmer, Nakamura, and Steinsson (2024), this pattern of forecast misses can stem from having a mis-specified prior belief about the mean of the interest rate process, rather than any systematic tendency to underreact. The z-scored regression avoids this issue since the average forecast error and the average forecast revision would both be zero. Therefore, the coefficient on forecast revisions would only be positive (indicating underreaction) if unusually large downward revisions in interest rate forecasts corresponded with the actual interest rate being even more below the forecast than the typical error — which does fit the intuitive sense of underreaction.

The one issue naively z-scoring could give rise to is that ”knowledge of the future” would influence the z-scored $\mathbb{E}_{t-1}(x_{i,t+j})$ variable in an undesirable way. For example, if a given country's forecast of 10-year ahead GDP growth in 1998 ends up being the highest in the sample, it would end up with a large z-scored value. However, presumably forecasters *in 1998* did not think they were giving what would prove to be an unusually high forecast. To get around this issue, the lagged forecast is z-scored with respect to *only* the observations up to that point in time. We also drop any observations where there are not at least 5 previous forecasts to base the z-score on. Note that this issue does not affect the forecast error or forecast revision z-scores.

Finally, similar to the logic on preventing certain countries from dominating the estimates, z-scoring allows us to pool together GDP, inflation, investment, and consumption forecasts without worrying that the variance of revisions or errors from one variable would dominate the other variables.

Is noise in long-horizon expectations an issue? In a recent paper, De Silva and Thesmar (2024) show that if longer-horizon forecasts are noisier than shorter-horizon forecasts, the β_1 coefficient in a regression of forecast errors on forecast revisions is biased downwards. If Consensus long-horizon forecasts are noisier, that would present an issue for our estimates.

However, appendix figures (11) and (12) plot the mean and median of the standard deviation of forecasts for a given variable at each horizon, as well as the mean squared-error of forecasts at different horizons (by variable).[3] Contrary to the results of De Silva and Thesmar (2024) (and of Patton and Timmermann 2010), in our data, the standard-deviation of forecasts is if anything *decreasing* in horizon, and the mean-squared error is slightly elevated at medium-horizons but not noticeably larger at six-to-ten year horizons than zero-year horizons. Therefore, we do not need to make the sort of noise adjustments made in De Silva and Thesmar (2024).

Our results can be reconciled with the two just mentioned papers by noting that in De Silva and Thesmar (2024) they look at analyst stock-level earnings forecasts, which are plausibly different from macroeconomic variables, and that in Patton and Timmermann (2010), they only look at expectations out to two years in the future (for which in our data there is some evidence of increased noise). Note that Ahn and Farmer (2024) also find inflation forecast ″noise″ (forecaster disagreeement) decreasing in the horizon of the forecast.

3.2 Results

Figure 2 and Table 1 show the results where each forecast variable is pooled together in regression (1). The forecasted annualized growth rate is the forecast of x_{t+h} for $h \leq 5$, while the ″horizon 6-10″ forecast is the forecast of the average annual value from $t+6$ to $t+10$.

[3] Consensus provides us directly with the standard deviation of forecasts at each horizon, rather than the underlying individual forecasts which could be used to calculate them.

Table 1: All Variables

Horizon	0	1	2	3	4	5	6-10
Revision	0.16***	0.04	-0.06**	-0.06***	-0.08***	-0.10***	-0.21***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)
Lag	-0.06***	-0.10***	-0.13***	-0.14***	-0.17***	-0.17***	-0.44***
	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Observations	8513	7776	7074	6400	5735	5067	3221

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

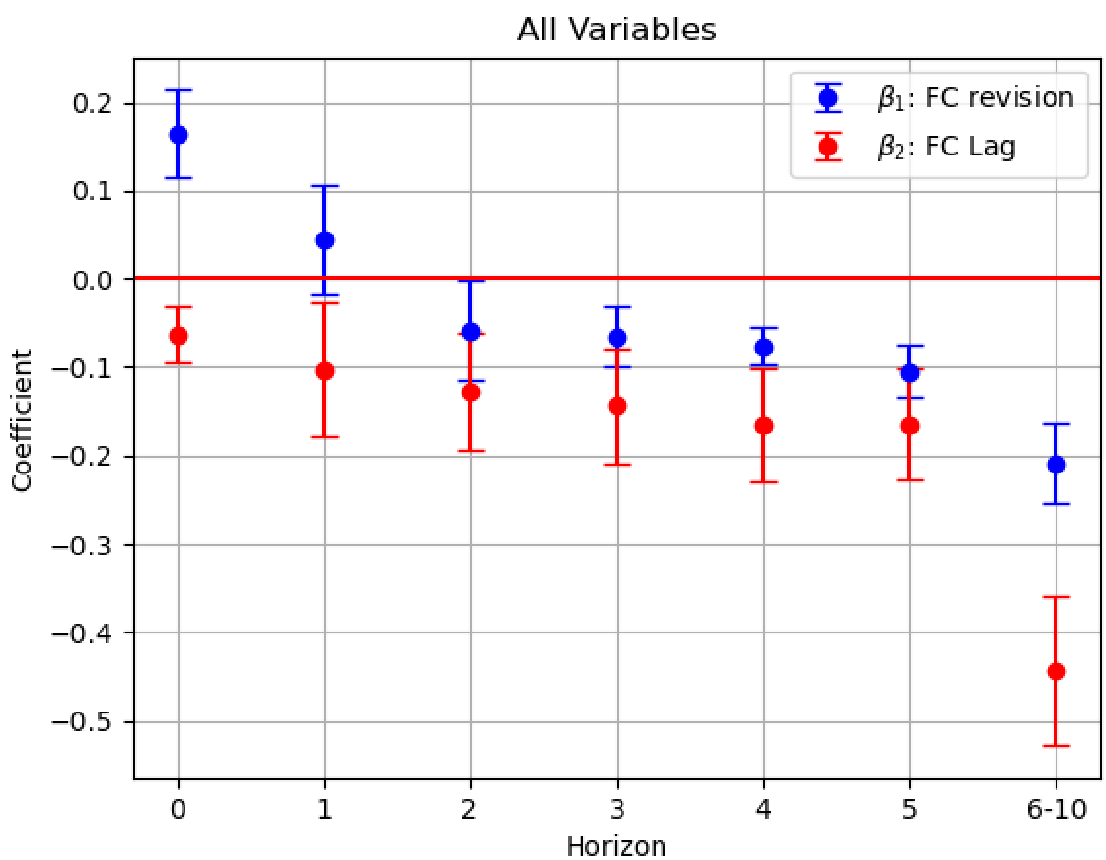


Figure 2: The figure plots the blue β_1 and red β_2 coefficients from the regression $e_{i,t+j} = \alpha + \beta_1 \Delta \mathbb{E}_t(x_{i,t+j}) + \beta_2 \mathbb{E}_{t-1}(x_{i,t+j}) + f_{i,t,x} + \epsilon_{i,t}$, where $e_{i,t+j}$ are forecast errors, $\Delta \mathbb{E}_t(x_{i,t+j})$ are consecutive survey forecast revisions, and $\mathbb{E}_{t-1}(x_{i,t+j})$ is the previous survey's forecast. All forecasts are pooled and are z-scored with respect to variable, country, and horizon, with expanding window z-scores for the lagged forecast. All observations where the outcome was 2020 or later are removed. Country-variable fixed effects are included and standard errors are Driscoll-Kraay with country-variable groupings.

All four facts discussed in the introduction appear in this result.

Fact 1 is that less than one-year forecasts under-revise, which is shown by the significantly positive β_1 coefficient. This is the same fact established by Coibion and Gorodnichenko (2015).

Fact 2 is that two-year and longer horizon forecasts *over*-revise, as shown by the significantly negative β_1 coefficients at horizons two and beyond.

Fact 3 is that forecasts are too extreme at every horizon, as shown by the fact that β_2 is everywhere significantly less than zero.

Fact 4 is that both senses of overreaction are increasing in forecast horizon, as shown by the monotonically decreasing β_1 and β_2 coefficients. The degree of overreaction of six-to-ten year forecasts is more than double that of five-year forecasts, for both senses of overreaction.

As discussed earlier, the discrepancy between the β_1 and β_2 coefficients at short-horizons implies that forecasts *under*-revise but *over*-extend. That is, when the average forecast updates, it tends to not update enough. But when the forecast is high relative to its typical value (up to that point in time), it tends to be too high (and when the forecast is low relative to its typical value, it tends to be too low).

The fact that β_1 coefficients flip from under to over-reaction between the one-to-two year horizon is consistent with the findings of Del Negro (2024) that professional forecasters flip from being under-confident to over-confident at greater than one-year horizons.

Using z-scores also allows us to interpret the magnitudes of the coefficients. The β_1 coefficient of -0.21 at a ten-year forecast horizon implies that a one standard deviation revision in average annualized six-to-ten-year expectations tends to produce a 0.21 standard deviation forecast error, where the forecast is above the outcome.

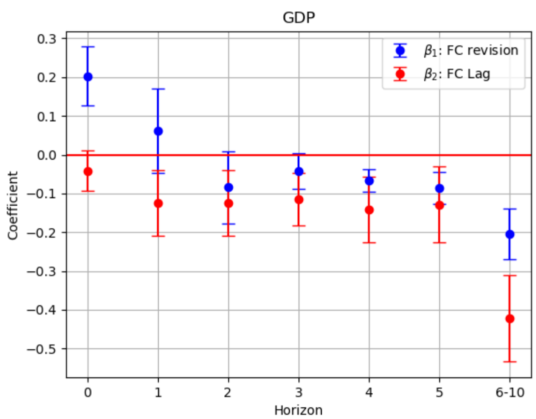
For our main specification we pool all the different macro variables together, since it gives us the tightest standard errors and the most precise estimates. As the next section shows, the pattern of results is broadly similar across variables, which justifies pooling the data and supports our interpretation of our findings. We want to argue we are discovering general patterns of forecaster over (and under) reaction — not patterns which are variable specific. In addition to the robustness checks we are about to discuss, appendix Figure (13) shows that if we restrict our sample to only survey observations from Q2 or Q4 (so the survey frequency doesn't change post-2014) we continue to get the same pattern of results.

3.2.1 Robustness

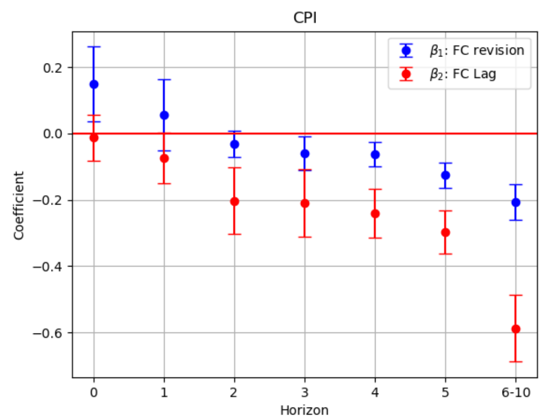
GDP, Inflation, Consumption, and Investment: Figure (3) shows the results from the same regression where the forecast variable is only one of GDP, inflation, consumption growth, or investment growth.

The four facts from the pooled variable regression are supported by each individual variable: 1) For all variables, β_1 is positive at short horizons, indicating under-revision, and is significant for each variables other than consumption. 2) For horizon 2 and longer, β_1 is always negative and is significant for all but three observations, for which it is very close to significant (horizons 2 and 3 for GDP, horizon 2 for inflation). 3) All point estimates of $\beta_2 < 0$ and all but three of those estimates are significant (inflation at horizons one and two, consumption at horizon two). 4) Both β_1 and β_2 have a downward trend for each variable. The coefficients are no longer monotonically decreasing, but overreaction clearly tends to increase in horizon, and horizon six-to-ten forecasts exhibit the greatest degree of overreaction for both β_1 and β_2 for each variable.

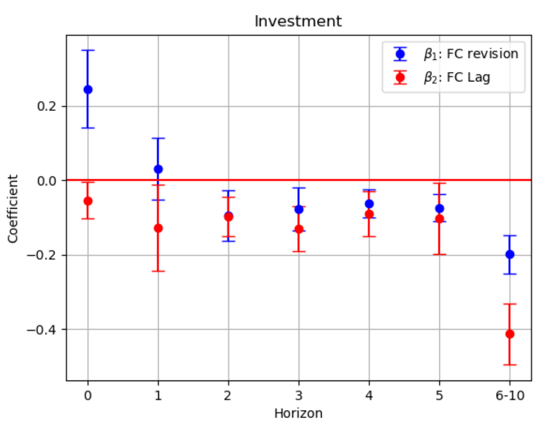
As the figures evince, since the patterns of how under and overreaction change with forecast horizon are so consistent across variables, it makes sense to pool the variables together and get tighter estimates.



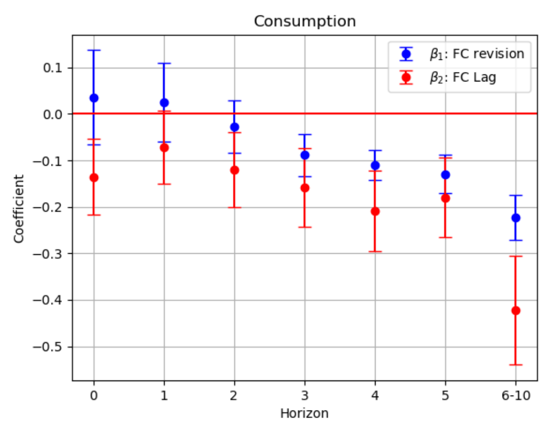
(a) GDP



(b) Inflation



(c) Investment



(d) Consumption

Figure 3: Regression coefficients from $e_{i,t+j} = \alpha + \beta_1 \Delta \mathbb{E}_t[x_{i,t+j}] + \beta_2 \mathbb{E}_{t-1}[x_{i,t+j}] + f_{i,t,x} + \varepsilon_{i,t}$ for different forecast variables x. Country-variable fixed effects are included and standard errors are Driscoll-Kraay with country-variable groupings.

Time Splits: Figure (4) shows that the facts are robust to different ways of splitting the sample along the time-dimension. The first two sub-plots split the sample into "halves", where each half is made so that there are an equal number of observations *within* a forecast-horizon grouping.

As plots 4a and 4c show, all four facts hold in each sub-sample. Where coefficients which correspond to facts are not significant, they are very close to being so, suggesting that power is the issue in the smaller sample. The first-half of the sample shows a slightly less clear pattern of monotonically decreasing coefficients, but still exhibits a downward trend in each coefficient and the most overreaction at six-to-ten year horizons.

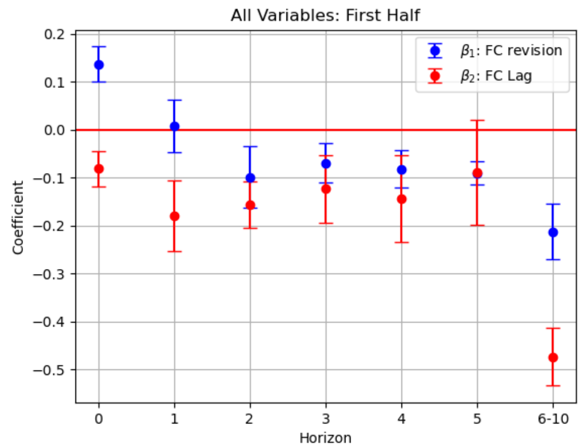
The third subplot 4c removes the Great Financial Crisis from the sample by taking out *all* observations where the forecast date *starts* before 2008 but includes 2007 or later in the forecast horizon. Given that we also remove Covid from the sample, for longer horizon forecasts, this reduces the sample quite a bit. The six-ten year out forecast sample goes from 3223 observations to 793. Despite the stringency of this test, all four facts continue to hold exactly.

Country Splits: Figure (5) splits the sample into 23 advanced economies and 63 emerging markets (all non-advanced economies in the sample) to show that the pattern of results is not specific to forecasting certain types of economies. In both samples, the four facts hold exactly. Once again, the downward trend in β_2 coefficients is only evident once the six-to-ten year horizon is included, but the six-to-ten year forecasts clearly exhibit the most over-extremity.

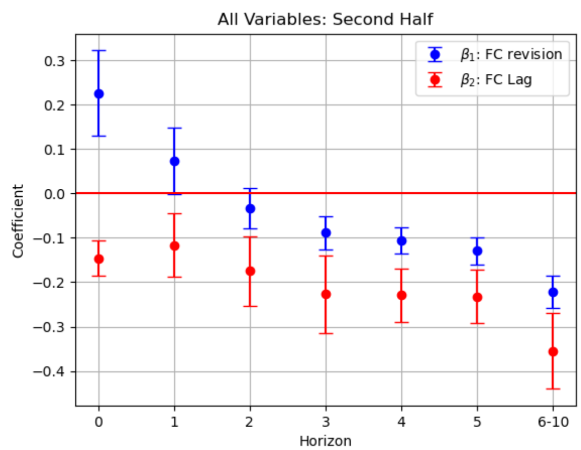
Different Fixed-Effects: Finally, Figure (6) shows that each of the four facts holds across four different ways of including fixed-effects in regression (1). The top-left subfigure 6a is a repeat of our main specification, which uses group fixed-effects. The groups are country i and macro variable x pairs.

The top-right figure 6b drops the group fixed-effects and uses time fixed-effects where the time periods are quarters. The bottom-left figure 6c includes both group and time fixed-effects. The bottom-right figure 6d uses no fixed-effects.

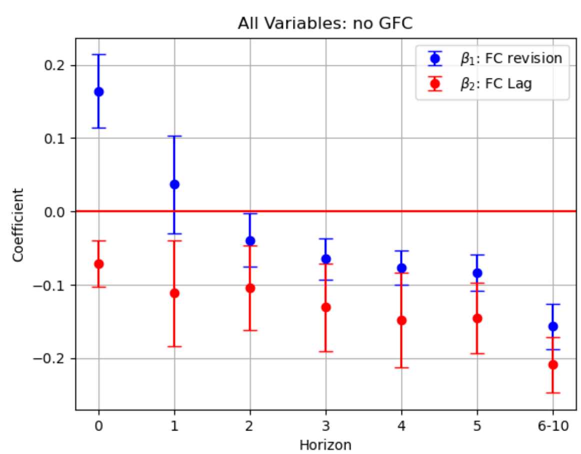
The results are broadly unchanged across all four different specifications. The biggest difference across results is that the two specifications which use time fixed-effects find slightly higher β_2 coefficients at short horizons (though the point estimate remains everywhere negative). Conceptually, we prefer the specifications without time fixed-effects because if forecasters are making a similar forecast error across countries/variables in a given time period, that is something a measure of overreaction *should* capture.



(a) First Half

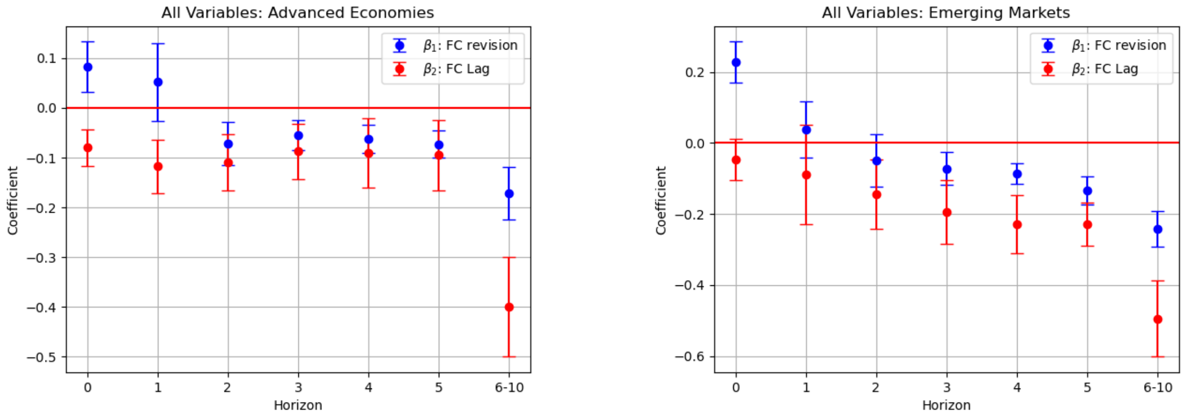


(b) Second Half



(c) No GFC

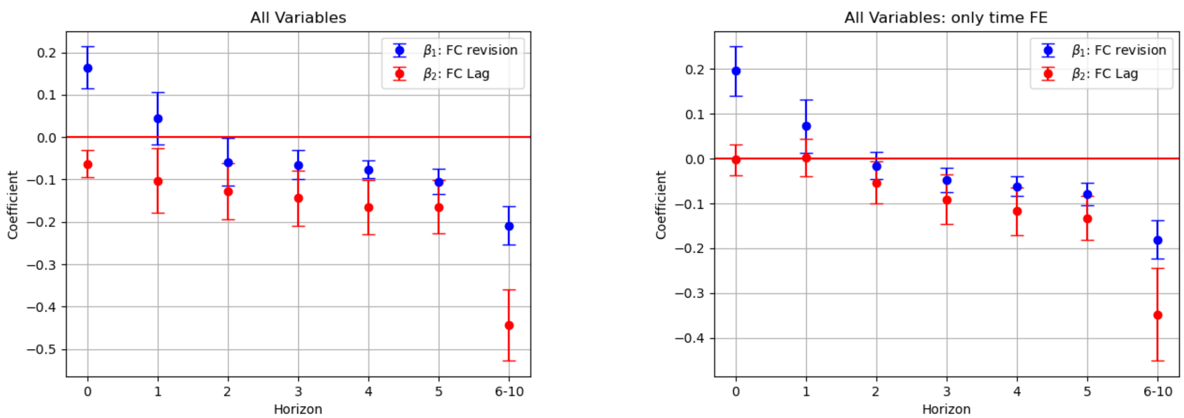
Figure 4: Regression coefficients from $e_{i,t+j} = \alpha + \beta_1 \Delta \mathbb{E}_t[x_{i,t+j}] + \beta_2 \mathbb{E}_{t-1}[x_{i,t+j}] + f_{i,t,x} + f_{i,t,x} + \varepsilon_{i,t}$ for different splits of the sample across the time dimension, with x pooled across all forecast variables. Country-variable fixed effects are included and standard errors are Driscoll-Kraay with country-variable groupings.



(a) Advanced Economies

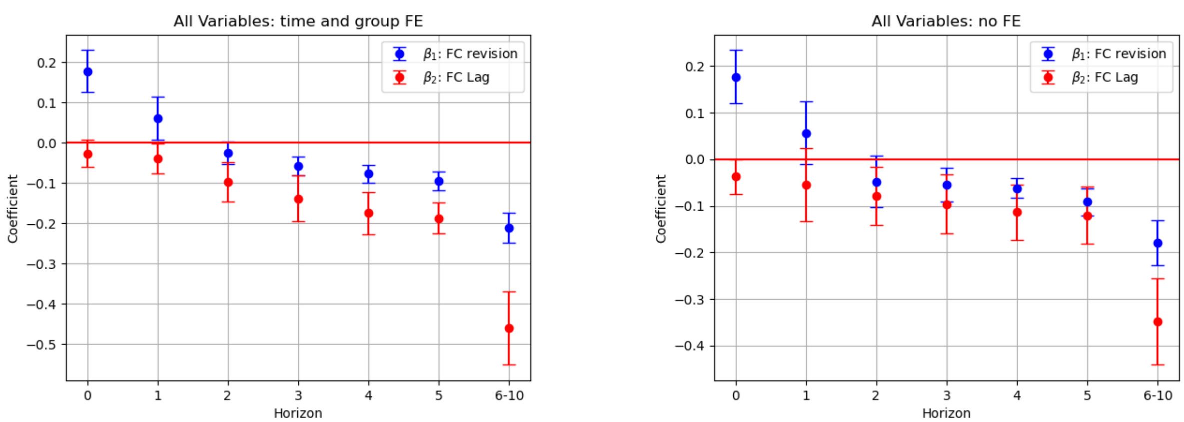
(b) Emerging Markets

Figure 5: Regression coefficients from $e_{i,t+j} = \alpha + \beta_1 \Delta\mathbb{E}_t[x_{i,t+j}] + \beta_2 \mathbb{E}_{t-1}[x_{i,t+j}] + f_{i,t,x} + \varepsilon_{i,t}$, splitting the sample into advanced and emerging economies, with x pooled across all forecast variables. Country-variable fixed effects are included and standard errors are Driscoll-Kraay with country-variable groupings.



(a) Group FE

(b) Time FE



(c) Group and Time FE

(d) No FE

Figure 6: Regression coefficients from $e_{i,t+j} = \alpha + \beta_1 \Delta\mathbb{E}_t[x_{i,t+j}] + \beta_2 \mathbb{E}_{t-1}[x_{i,t+j}] + f_{i,t,x} + \varepsilon_{i,t}$ for different fixed-effects $f_{i,t,x}$. Group fixed-effects are country i and macro variable x groupings. Standard errors are Driscoll-Kraay with country-variable groupings.

3.3 Out-of-sample Forecasting

In a recent paper, Eva and Winkler (2023) argue that in order to reject rational expectations, forecast errors should be more predictable out of sample when adjusting for forecast biases. Yet, for the forecast biases they test — such as the Coibion and Gorodnichenko (2015) aggregate under-revision bias — adjusting for the bias does not consistently improve out-of-sample forecasts. This is an important challenge to the literature arguing for systematic departures from FIRE, since it suggests the biases found by researchers are not very stable over time.

However, the forecast biases tested in Eva and Winkler (2023) all use US only data where the forecast horizon is 1-year or less. Therefore, they are not able to test whether the cross-country and long-horizon biases we have documented work out of sample. Here we conduct such an exercise and show that our forecast biases *do* help predict forecast errors out of sample, providing additional support to the realism and importance of long-horizon overreaction.

Our out-of-sample forecasting test follows the procedure in Eva and Winkler (2023) closely. The basic idea is as follows: for each year-quarter date unit in our dataset, we figure out which country-horizon-variable groupings have at least ten observations up to that point in time. We then z-score all observations up to that point in time using country-horizon-variable groupings. As before, the z-scores for the lagged forecast are rolling. Then, on this "training" data we run regression (1).

Then, we take the next year-quarter date unit in our dataset, and add the country-horizon-variable data which has corresponding country-horizon-variable data in the test set. We then re-z-score all observations. Then, we compute a "bias-adjusted" forecast as follows:

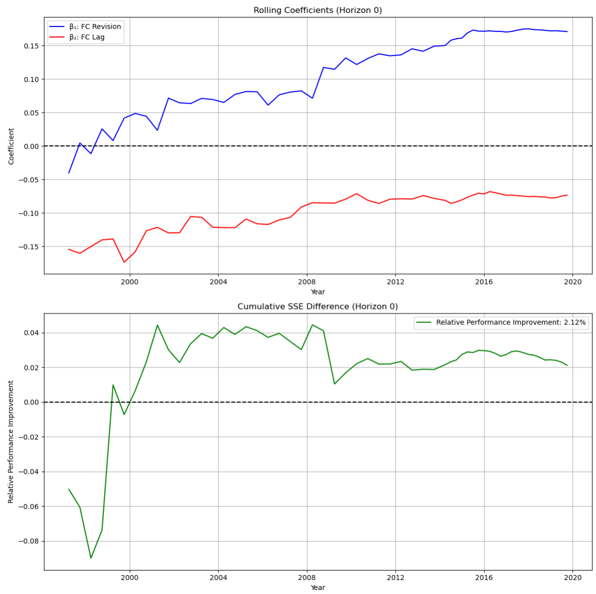
$$\mathbb{E}_t[x^*_{i,t+j}] = \mathbb{E}_t[x_{i,t+j}] + \beta_{1,t-1}\Delta\mathbb{E}_t[x_{i,t+j}] + \beta_{2,t-1}\mathbb{E}_{t-1}[x_{i,t+j}] \tag{2}$$

The $*$ indicates that $\mathbb{E}_t[x^*_{i,t+j}]$ is a bias-adjusted forecast. The time subscripts on the β coefficients reflect the fact that they are estimated on the data up to but not including the forecast date t. Therefore, the adjusted forecast uses only information that a forecaster at the time who had access to the history of forecasts would be able to use.

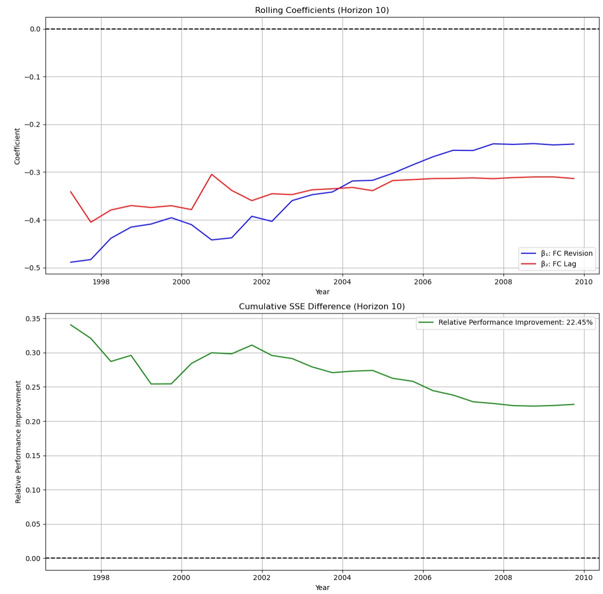
We then compare the performance of the adjusted forecast with the actual forecast and compute a rolling sum of squared errors for each: $\text{SSE}^a = x_{i,t+j} - \mathbb{E}_t[x_{i,t+j}]$ and $\text{SSE}^* = x_{i,t+j} - \mathbb{E}_t[x^*_{i,t+j}]$, where SSE^a is the sum of squared errors from the actual forecasts and SSE^* is the sum of squared errors from the bias adjusted forecasts. Finally, we compare the performance of the two by calculating the following relative performance metric:

$$\text{RP} = \frac{\text{SSE}^a - \text{SSE}^*}{\text{SSE}^a} \tag{3}$$

The metric RP gives the % increase in cumulative squared errors that comes from using the actual forecasts as opposed to the bias-adjusted forecasts. A positive value of RP therefore indicates that actual forecasts are worse than bias-adjusted forecasts, since they produce RP% higher cumulative squared errors.



(a) Horizon 0 OOS Forecasting



(b) Horizon 6-10 OOS Forecasting

Figure 7: The top sub-plots on the left and right show the rolling estimates of β_1 and β_2 as the sample increases in regression (1). The bottom sub-plots shows the % difference in sum of squared forecasting errors from using the original forecasts versus the bias-adjusted forecasts, as defined in (3).

Figure 7 compares the OOS results for horizon zero and horizon six-to-ten. The left-hand panel 7a is for horizon zero. The top of the left-hand panel shows the β_1 and β_2 coefficients at each point in time, with the β_1 coefficients in blue and β_2 coefficient in red. The bottom panel shows RP over time, with a positive value indicating the bias-adjusted forecast outperforms the actual forecast.

The idea behind plotting the coefficients over time is it allows you to see how stable the over or under-reaction is. Two things to note are that 1) the dates on the x-axis represent the date of the *forecast*, not the date of the outcome. This is why the horizon six x-axis stops in 2014: any forecasts made after that point would include Covid. 2) we do not use country-variable fixed effects this time around since they might have undue influence in the smaller samples at the beginning of the data. Still, the end of sample coefficients here will be slightly different than the no group fixed effects coefficients plotted in 6d because they do not include the last year-quarter unit and because of the five data point burn-in period.

Both the horizon zero and horizon six-to-ten plots show a fair amount of coefficient stability. The β_1 coefficient starts out negative at the very beginning of the horizon zero sample, but otherwise the signs of the coefficients are consistent, and by the last five years or so of the sample the coefficients are very stable.

The horizon zero cumulative performance figure on the bottom-left shows the bias-adjusted forecast slightly outperforming the un-adjusted forecast after starting the sample under-performing. By the end of the sample the cumulative relative performanc eimprovement is 2.12%. The improvement from bias-adjusting at the ≤ 1 year horizon is better than the negative results in Eva and Winkler (2023), but the small magnitude of the advantage is consistent with their overall message.

The horizon six-to-ten cumulative performance figure in the bottom-right tells a very different story. The bias-adjusted forecast always outperforms the unadjusted forecast by at least 20%, and a 20-30% performance gap remains steady for most of the sample. Outperformance is 22.45% by the very end. Note that a flat line here is consistent with the bias-adjusted forecast continually outperforming, since the y-axis is in percentage terms — if the forecasts started performing equally well, the line would decline towards zero performance difference.

Table 2 shows that as early as the one-year ahead horizon the bias-adjusted forecast outperforms the unadjusted forecast by over 10%. At horizon two and beyond — where there is overreaction as measured by both β_1 and β_2 — adjusting for overreaction leads to 16-23% better forecasting performance. Appendix figures 14-18 show plots like figure 7 for horizons one through five.

Table 2: Final Cumulative SSE Difference by Horizon (in percentages)

0	1	2	3	4	5	10
2.1%	11.0%	17.8%	16.1%	16.0%	16.1%	22.4%

3.4 Discussion

Across all of GDP, inflation, consumption, and investment; across emerging markets and advanced economies; from pre-to-post GFC, the same four facts hold: forecasts under-revise at short-horizons, over-revise at two-year and longer horizons, over-extend at all forecasts horizons, and over-revise and over-extend more the farther out in time is the outcome they are forecasting. At six-to-ten year horizons, forecast overreaction is strongest in both dimensions.

Other papers have shown that longer-horizon expectations overreact (Bordalo, Gennaioli, Porta, and Shleifer 2024, Angeletos, Huo, and Sastry 2021, Afrouzi et al. 2023, d'Arienzo 2020), but as far as we are aware, we are the first to show that overreaction in specifically *macroeconomic* expectations increases with horizon, and does so across a broad and representative cross-country sample. Furthermore, we are the first to show that adjusting for overreaction allows for improved out-of-sample forecasting.

As discussed more extensively in Afrouzi et al. (2023), our results are consistent with a model where agents' over-update their belief about the long-run mean in response to current data realizations, but are inconsistent with other popular models of overreaction. In particular, standard over-extrapolation and precision over-estimation — the types of overreaction used in Angeletos, Huo, and Sastry (2021) — will not produce overreaction that increases in horizon. Similarly, diagnostic expectations will not produce overreaction that increases in horizon unless the diagnostic expectations are specifically with respect to an object like the long-run mean.

4 Expectations and the economy

In Bordalo, Gennaioli, Porta, OBrien, et al. (2024) they show that changes in an index of ”long-term” expected US stock market earnings growth predicts a short-term boom in real variables like growth and investment followed by a subsequent bust.[4]

In this section, we show that, in our broad cross-country sample, the relationship between *GDP* growth forecasts and subsequent investment and GDP growth follows a slightly different pattern. There are systematic ”booms” and ”busts” in GDP and investment growth, but it is changes in *short-term* GDP growth forecasts which are most predictive of those booms and busts. This stands in contrast to the long-term forecasts emphasized by Bordalo, Gennaioli, Porta, OBrien, et al. (2024). In light of our previous section’s findings — that it is long-horizon expectations which overreact most — these results suggest movements in *forecasters* long-horizon expectations are relatively inert: their movement is not associated with changes in economic agents activity. That could be because forecasters expectations are not representative of economic decision makers’ expectations, or because even economic decision makers short-term actions do not respond significantly to their long-term beliefs.

In order to examine whether changes in growth expectations influence business cycle fluctuations, we follow the local projections approach used in Bordalo, Gennaioli, Porta, OBrien, et al. (2024). Specifically, we run panel local projections of the following form:

$$y_{i,t+h} = \alpha + \beta_1 \Delta_1 \mathbb{E}_t\left(g_{i,t+j}\right) + \beta_2 \mathbb{E}_t(y_t) + \beta_3 \mathbf{X}^*_{i,t} + f_{i,t} + \epsilon_{i,t} \tag{4}$$

Dropping country i subscripts for convenience, the dependent variable here is any macroeconomic variable of interest y_{t+h}, where h indicates that the projection is h years after t. The coefficient of interest is β_1, which multiplies the one-year change in j-year ahead GDP growth expectations, $\Delta_1 \mathbb{E}_t\left(g_{t+j}\right)$.

The variable $\mathbb{E}_t(y_t)$ indicates the ”main” control variable, which is the current-year forecast of the dependent variable $\mathbb{E}_t(y_t)$. Controlling for the current-year forecast is used to control for the fact that if the forecast revision is measured in the middle of the year, there is information about current-year economic conditions that is not controlled for by our other lagged controls.

The other control $\mathbf{X}^*_t$ is a vector of lagged macroeconomic variables which allow us to control for standard business cycle dynamics. The full set of controls is the one-year lag of the country’s GDP growth, investment growth, inflation, and stock market return; the

[4]”Long-term” is in quotations marks because the long-term expectations they use are expectations of companies average annual earnings growth over the next three-to-five years, but the horizon is not specified more than that.

change in GDP growth and investment growth from $t-2$ to $t-1$ and $t-3$ to $t-2$; the two-year lag of inflation and the stock market return; and the $t-2$ to $t-1$ and $t-3$ to $t-2$ change in the country's real interest rate. Details on the data sources for stock market returns and real interest rates are provided in appendix X. In robustness checks, we use a smaller subset of these control variables — just the one-year lag of GDP, investment, and stock market returns as well as the $t-2$ to $t-1$ change in investment and GDP.

These controls help us to isolate the role that changes in expectations alone have on business cycle dynamics. Any causal influence of past macroeconomic aggregates on the change in expectations that also effects future macroeconomic aggregates independently of their effect on expectations will be stripped out.

Once again, $f_{i,t}$ indexes the (possible) presence of country, time, or country and time fixed-effects. We also continue to z-score all variables with respect to country.

Since all variables are z-scored, the coefficients represent the impact of the independent variable on the dependent variable in standard deviation units.

Again using Driscoll-Kraay standard errors, the dark gray bands represent 68% confidence intervals, while the light gray bands represent 95% confidence intervals.

4.1 Results

Figure (8) is a two-by-two panel. In the left-hand column, the dependent variable is investment growth — the traditional propagator of "animal spirits" driven booms and busts. In the right-hand panel, the dependent variable is GDP growth.

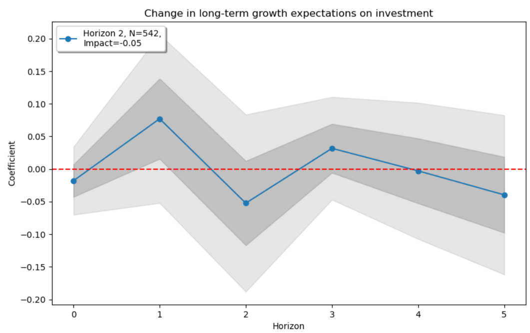
In the first row, the independent variable is the one-year change in long-term growth expectations, where long-term growth expectations are expectations for average annual growth six-to-ten years ahead. In the second row, the independent variable is the one-year change in "short-term growth expectations", where we define short-term growth expectations as the average annual growth expected for the current year and the subsequent two years.

The first panel row shows, at best, a very mild boom-bust pattern. Investment growth is significantly positive relative to the 68% CI at the one-year horizon and GDP growth is significant negative relative to the 68% CI at the two-year horizon, but no horizon responses are significant relative to the 95% CI. Furthermore, the magnitude of the "boom" and "bust" predicted by changing long-term expectations is less than one-tenth of a standard deviation.

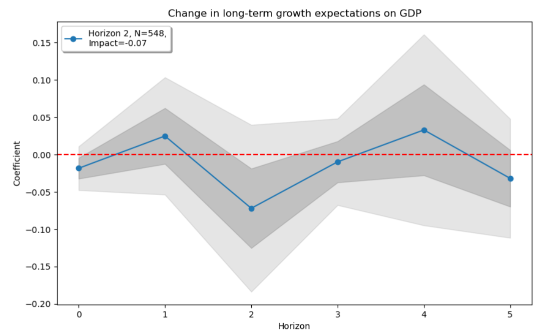
By contrast, the second row shows that changes in short-term growth expectations do seem to be strongly associated with booms and busts. Both investment and GDP boom in a highly significant fashion in the initial year, and both of them bust dramatically two-years later.[5] The magnitudes of the booms and busts are a bit over a third of a standard deviation — about four-to-five times larger than the association with changes in long-term expectations. The pattern of results in this panel matches well the "main business

[5] The p-value on the two-year horizon coefficient where GDP is the dependent variable is .0516.

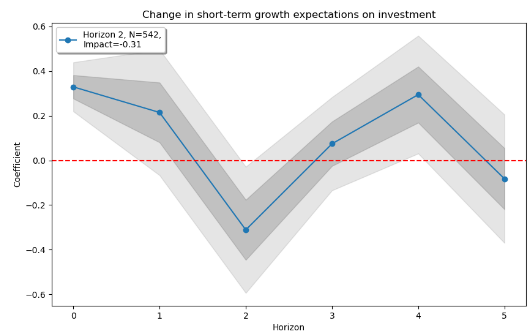
cycle” shock identified in Angeletos, Collard, and Dellas (2020), which can result from fluctuations in short-term economic confidence (Angeletos, Collard, and Dellas 2018).



(a) Long-term expectations: Investment



(b) Long-term expectations: GDP



(c) Short-term expectations: Investment

(d) Short-term expectations: GDP

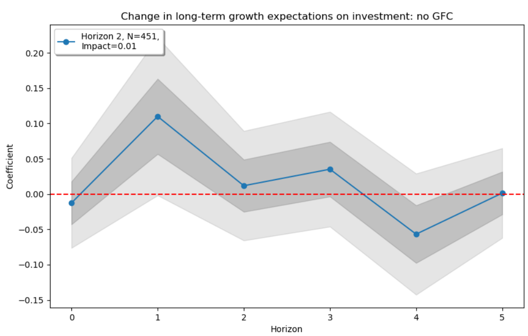
Figure 8: Regression coefficients from $y_{i,t+h} = \alpha + \beta_1 \Delta_1 \mathbb{E}_t(g_{i,t+j}) + \beta_2 \mathbb{E}_t(y_t) + \beta_3 \mathbf{X}^*_{i,t} + f_{i,t} + \epsilon_{i,t}$ for different horizon-forecasts g. The top-row uses the one-year change in six-to-ten year ahead average annual GDP growth expectations; the bottom-row uses the one-year change in average annual GDP growth expectations for the year of the forecast and the following two years. The dependent variable on the LHS column is investment, and on the RHS is GDP. Country-variable fixed effects are included and standard errors are Driscoll-Kraay with country-variable groupings.

4.1.1 No GFC

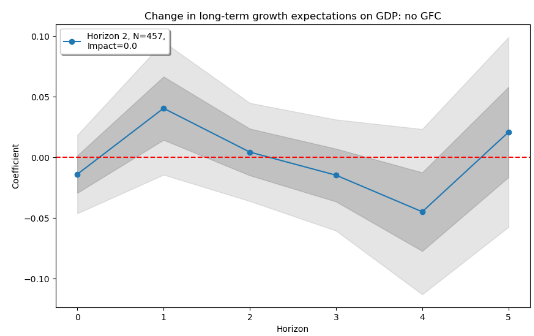
We next show that the ”bust” portion of these results is fairly reliant on including the great financial crisis (GFC) in the sample. Figure (9) shows the exact same type of results, except with one-year expectation changes in 2005, 2006, and 2007 removed from the sample.

The top-row shows that, once the GFC is excluded, there continues to be no significant association between changes in long-term growth expectations and booms and busts. The bottom-row shows that upward revisions in short-term growth expectations continue to be associated with booms in investment and growth — even within the current-year where the forecast is controlled for — consistent with the finding that average short-term forecasts under-revise. However, the association between upward movements in short-term growth expectations and two-years later busts is significantly dampened: the two-year horizon coefficients for both investment growth and GDP growth are about $1/4 - 1/3$ of their magnitude when the GFC is included, and they no longer lie outside the 95% CI.

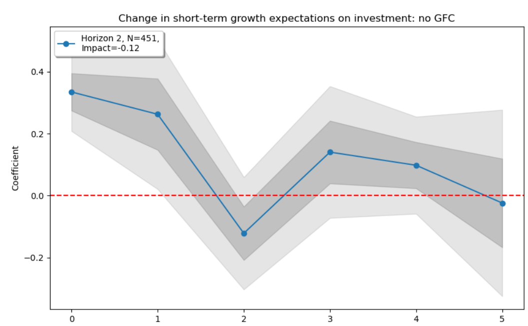
The attenuation of the bust results when the GFC is excluded is not necessarily decisive evidence *against* the boom-bust mechanism: after all, arguably the GFC is a prime example of exactly the sort of boom-bust dynamics consistent with an animal spirits story. Given that the non-GFC results point in the same direction, even if insignificant, we prefer to interpret our results as showing that changes in short-term growth expectations have a tendency to create boom-bust dynamics, but by no means always necessitate subsequent booms and busts.



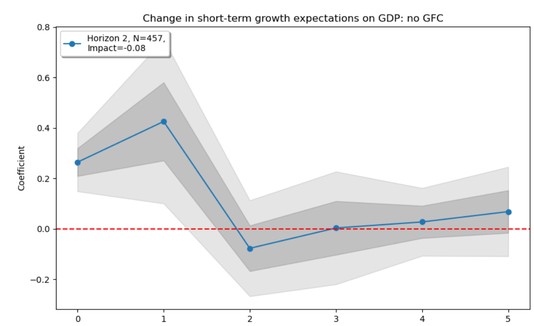
(a) No GFC, LT expectations: Investment



(b) No GFC, LT expectations: GDP



(c) No GFC, ST expectations: Investment



(d) No GFC, ST expectations: GDP

Figure 9: No GFC: Regression coefficients from $y_{i,t+h} = \alpha + \beta_1 \Delta_1 \mathbb{E}_t(g_{i,t+j}) + \beta_2 \mathbb{E}_t(y_t) + \beta_3 \mathbf{X}^*_{i,t} + f_{i,t} + \epsilon_{i,t}$ for different horizon-forecasts g. The top-row uses the one-year change in six-to-ten year ahead average annual GDP growth expectations; the bottom-row uses the one-year change in average annual GDP growth expectations for the year of the forecast and the following two years. Changes in expectations from 2005, 2006, and 2007 are removed from the sample. The dependent variable on the LHS column is investment, and on the RHS is GDP. Country-variable fixed effects are included and standard errors are Driscoll-Kraay with country-variable groupings.

Table (3) presents the regression results for horizons zero through four years later, for both the with and without GFC samples. The main entries are the β_1 coefficient estimates and the parentheses below the coefficient are (X, Y) where X is the number of observations for that regression and Y is the (Driscoll-Kraay) standard error of the β_1 coefficient.[6]

Table 3: Changes in growth expectations on actual investment and growth

	Horizon				
Variable	h=0	h=1	h=2	h=3	h=4
LT Growth → GDP	-0.02	0.03	-0.07	-0.01	0.03
	$(713, 0.02)$	$(630, 0.04)$	$(548, 0.06)$	$(486, 0.03)$	$(444, 0.07)$
LT Growth → Investment	-0.02	0.08	-0.05	0.03	-0.00
	$(710, 0.03)$	$(624, 0.07)$	$(542, 0.07)$	$(481, 0.04)$	$(440, 0.05)$
ST Growth → GDP	0.28^{***}	0.34^{*}	-0.32^{*}	-0.08	0.20
	$(715, 0.06)$	$(631, 0.18)$	$(548, 0.17)$	$(486, 0.12)$	$(444, 0.14)$
ST Growth → Investment	0.33^{***}	0.22	-0.31^{**}	0.08	0.30^{**}
	$(712, 0.06)$	$(625, 0.15)$	$(542, 0.15)$	$(481, 0.11)$	$(440, 0.14)$
No GFC					
LT Growth → GDP	-0.01	0.04	0.00	-0.01	-0.05
	$(622, 0.02)$	$(539, 0.03)$	$(457, 0.02)$	$(395, 0.02)$	$(353, 0.04)$
LT Growth → Investment	-0.01	0.11^{*}	0.01	0.04	-0.06
	$(619, 0.03)$	$(533, 0.06)$	$(451, 0.04)$	$(390, 0.04)$	$(349, 0.04)$
ST Growth → GDP	0.26^{***}	0.42^{**}	-0.08	$3.35e-03$	0.03
	$(624, 0.06)$	$(540, 0.16)$	$(457, 0.10)$	$(395, 0.11)$	$(353, 0.07)$
ST Growth → Investment	0.33^{***}	0.26^{**}	-0.12	0.14	0.10
	$(621, 0.06)$	$(534, 0.12)$	$(451, 0.09)$	$(390, 0.11)$	$(349, 0.08)$

Notes: Values in parentheses show (N, SE) where N is the number of observations and SE is the standard error. *** p<0.01, ** p<0.05, * p<0.1

5 Stock Return Predictability

In a different paper from the one just discussed, Bordalo, Gennaioli, Porta, and Shleifer (2024) show that the same index of "long-term" expected earnings growth predicts negative stock-market returns on five-year horizons in the US better than short-term expectations, and they claim this is evidence that overreacting long-term expectations help explain a number of stock-market puzzles.

In this section, we show that the relationship between *GDP* growth forecasts and subsequent stock-market returns is similar in the US to the findings of Bordalo, Gennaioli,

[6] Due to all the control variables used, the sample is significantly smaller here. Appendix figure (19) shows that you use *no* control variables other than the current-year forecast you still get a similar pattern of results; however, we do not think the quantitative part of those results should be paid much attention.

Porta, and Shleifer (2024), but differs around the world. In the rest of the world, high *short-term* GDP growth forecasts are most predictive of subsequent local country stock-market returns, due to their association with short-term *weak* returns. This result is consistent with our previous section that changes in short-term expectations are most strongly associated with booms and busts in the business cycle.

5.1 USA

First, we show that we get a similar pattern of results as Bordalo, Gennaioli, Porta, and Shleifer (2024) when looking at just US data. The US data runs from April 1990 to July 2023.[7] Stock market returns for the US are real cum-dividend returns on the S&P 500. Nominal returns are deflated by the World Bank's World Development Indicators CPI measure, for the US and for all other countries.

At each date at which we have a new set of consensus expectations, we construct the one-year, three-year, five-year, and the one-year-four-year-forward (1y4y) real cum-dividend return. The latter is the return from $t+1$ to $t+5$ where $t+1$ begins one year from the date of the forecast.

Table (4) examines return predictability at these different horizons relative to three different expectations. The first-panel is ”short-term” growth expectations: average annual GDP growth expectations in the year of the forecast and the subsequent two-years The second-panel is average annual GDP growth expectations over the next ten years. The third-panel is average annual GDP growth over the years six-to-ten years from the forecast.

The reason we include both 0-10 year and 6-10 year average forecasts is because we believe 6-10 year average forecasts are the right measure of ”long-term” growth expectations, but 0-10 year expectations are more comparable to the expectations measure used in Bordalo, Gennaioli, Porta, and Shleifer (2024).

The key result in the table is that, within the US, the true *long-term* expectations — the six-to-ten year ahead average growth expectations — are the strongest return predictors for horizons beyond one-year. For 3-year, 5-year, and 1y4y year returns the regression using 6-10 year average GDP growth has the largest R^2 and the largest coefficients, in absolute value. Since all variables are z-scored the coefficients are directly comparable across independent variables.

[7] The results we present in this section include 2020, unlike our results from previous sections, since the rapid stock market recovery in the wake of its Covid crash makes it less likely that Covid distorts this data. Appendix tables (7) and (8) show that all the results in this section retain the same pattern when Covid is excluded from the sample.

Table 4: United States: Return Predictability Regressions

	Return Horizon			
	1-year	3-year	5-year	1y4y
2-year avg GDP growth	-0.33^{***}	-0.21^{*}	-0.19	-0.14
	(0.10)	(0.13)	(0.14)	(0.13)
	[13.5%]	[3.7%]	[3.2%]	[1.7%]
10-year avg GDP growth	-0.33^{***}	-0.33^{***}	-0.23^{**}	-0.13
	(0.09)	(0.10)	(0.11)	(0.11)
	[14.2%]	[14.1%]	[6.6%]	[2.1%]
6-10 year avg GDP growth	-0.26^{***}	-0.39^{***}	-0.27^{***}	-0.17^{**}
	(0.09)	(0.08)	(0.08)	(0.08)
	[10.5%]	[23.8%]	[11.3%]	[4.7%]

Notes: Standard errors in parentheses, R^2 in square brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

As in Bordalo, Gennaioli, Porta, and Shleifer (2024), there is a strong pattern of high-growth expectations predicting subsequently low returns. This holds across all return and expectation horizons, but is strongest for the long-term growth expectations, which predict low returns on both a one-year horizon and one-year-four-year-forward horizon. The level of six-to-ten year ahead GDP growth expectations explains 24% of the variance of the next three-year's returns and 11% of the variance of the next five-year's returns.

5.2 All countries

We now turn to the results when using a broad sample of 34 countries for which we have both expectations and stock-market return data. All stock-market return data are cum-dividend returns and come from WRDS's "Daily World Indices" database. Returns are constructed relative to the exact date of the forecast, the same as they were in the US, and are deflated by the WDI's CPI measure. The US remains part of this sample.

Table 5: 34 countries: Return Predictability Regressions

	Return Horizon			
	1-year	3-year	5-year	1y4y
2-year avg GDP growth	-0.30^{***}	-0.14	-0.22^{***}	-0.03
	(0.05)	(0.09)	(0.06)	(0.06)
	[9.4%]	[1.4%]	[3.8%]	[0.1%]
10-year avg GDP growth	-0.31^{***}	-0.13	-0.14^{**}	0.08
	(0.08)	(0.10)	(0.06)	(0.06)
	[6.9%]	[1.1%]	[1.5%]	[0.5%]
6-10 year avg GDP growth	-0.15^{*}	-0.10	-0.04	0.12
	(0.09)	(0.08)	(0.07)	(0.08)
	[1.5%]	[0.7%]	[0.1%]	[1.1%]

Notes: Standard errors in parentheses, R^2 in square brackets. *** p<0.01, ** p<0.05, * p<0.1

Table (5) shows the results. The key result is that, unlike in the US, the long-term 6-10 year later GDP growth expectations no longer significantly predict returns at *any* horizon, while the short-term 0-2 year growth forecasts are the strongest return predictor. As before, return predictability comes from high growth expectations predicting subsequently *low* returns, but now the relationship is strongest between short-term growth expectations and weak subsequent returns.

The regression results also reveal two other interesting findings: 1) the fact that short-term expectations predict weak returns is driven by the relationship between high short-term growth expectations and immediate weak returns in the following year. There is no meaningful relationship between short-term growth expectations and 1y4y forward returns. 2) Unlike in the US, there is a *positive* relationship between long-term growth expectations and 1y4y returns. It is not significant, but the fact that it is positive while one-year returns are negatively associated with long-term growth expectations explains why there is no longer horizon return predictability.

These results suggest that long-term expectations are not a systematic and universal explainer of stock-market puzzles, but rather, that US long-term growth expectations have happened to be off in a way that strongly predicted subsequent returns. That could still very well be due to the sort of mechanisms outlined by Bordalo, Gennaioli, Porta, and Shleifer (2024), but these results show that more work is done to need to understand *when* and *why* those sorts of mechanisms kick-in, and whether there are features of the US stock market that makes it uniquely responsive to long-term growth expectations.

One important caveat is it could be that in the US, firms earnings growth expectations are more connected with aggregate GDP growth expectations than in other countries, making it so that long-term *earnings* growth expectations are a consistent explainer of return anomalies, but that our data is unable to capture such a relationship around the world.[8] Appendix tables (9) and (10) show that our results are not significantly different across advanced and emerging market economies, but more work needs to be done to fully examine this hypothesis.

Our finding that short-term growth expectations explain more of subsequent stock-market returns than long-term expectations is consistent with that of De la O and Myers (2024), though we emphasize that here we use GDP growth expectations, while they focus on firms' earnings-growth expectations.

6 Conclusion

This paper leverages Consensus Economics long-term survey — a large cross-country panel of economists' macroeconomic expectations out to ten-years in the future — to establish four facts about average macroeconomic expectations:

(i) Less than one-year ahead expectations under-revise. This is the under-revision of average expectations documented by Coibion and Gorodnichenko (2015).

(ii) Two or more year ahead expectations over-revise. There has been other evidence of over-revision at longer horizons, but we are the first to document at precisely which horizons over-revision prevails for macroeconomic expectations.

(iii) All horizon expectations tend to be too extreme. The fact that over-extremity tends to be the case at all horizons is novel.

(iv) Over-revision and over-extremity increase in the horizon of the forecast. This matches the experimental evidence about abstract AR(1) processes in Afrouzi et al. (2023), suggesting that this pattern is a general feature of human expectation formation.

These facts hold for each of four different macroeconomic variables, and across advanced and emerging economies. By adjusting for these biases in forecasts, we are able to improve out-of-sample forecasting of six-to-ten year ahead macroeconomic outcomes by over 20%.

We then show that despite long-term expectations overreacting more, it is short-term expectations which are most strongly associated with "booms and busts" in investment, GDP, and the stock market. This result stands in contrast to other recent work which has emphasized the role of long-run overreaction in explaining business cycle and stock market fluctuations.

We leave it to future work to come up with a model which integrates all these facts about expectations while matching the associations between different horizon expectations and real and financial variables.

[8] Allen et al. (2024) presents evidence that China's stock market has a different relationships with *realized* GDP growth than other countries, but the patterns across non-China countries are fairly similar.

References

Adam, Klaus, Albert Marcet, and Johannes Beutel (2017). "Stock price booms and expected capital gains". In: *American Economic Review* 107.8, pp. 2352–2408.

Afrouzi, Hassan et al. (2023). "Overreaction in expectations: Evidence and theory". In: *The Quarterly Journal of Economics* 138.3, pp. 1713–1764.

Ahn, Hie Joo and Leland Farmer (2024). "Diagreement about the Term Structure of Inflation Expectations". In.

Allen, Franklin et al. (2024). "Dissecting the Long-Term Performance of the Chinese Stock Market". In: *The Journal of Finance* 79.2, pp. 993–1054.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas (2018). "Quantifying confidence". In: *Econometrica* 86.5, pp. 1689–1726.

— (2020). "Business-cycle anatomy". In: *American Economic Review* 110.10, pp. 3030–3070.

Angeletos, George-Marios and Zhen Huo (2021). "Myopia and anchoring". In: *American Economic Review* 111.4, pp. 1166–1200.

Angeletos, George-Marios, Zhen Huo, and Karthik A Sastry (2021). "Imperfect macroeconomic expectations: Evidence and theory". In: *NBER Macroeconomics Annual* 35.1, pp. 1–86.

Angeletos, George-Marios and Chen Lian (2018). "Forward guidance without common knowledge". In: *American Economic Review* 108.9, pp. 2477–2512.

Augenblick, Ned, Eben Lazarus, and Michael Thaler (2025). "Overinference from weak signals and underinference from strong signals". In: *The Quarterly Journal of Economics* 140.1, pp. 335–401.

Ba, Cuimin, J Aislinn Bohren, and Alex Imas (2024). "Over-and underreaction to information". In: *Available at SSRN 4274617*.

Bardóczy, Bence and Joao Guerreiro (2023). "Unemployment insurance in macroeconomic stabilization with imperfect expectations". In: *Manuscript, April*.

Beaudry, Paul and Franck Portier (2014). "News-driven business cycles: Insights and challenges". In: *Journal of Economic Literature* 52.4, pp. 993–1074.

Bianchi, Francesco, Cosmin Ilut, and Hikaru Saijo (2024). "Diagnostic business cycles". In: *Review of Economic Studies* 91.1, pp. 129–162.

Bianchi, Francesco, Sydney C Ludvigson, and Sai Ma (2024). *What Hundreds of Economic News Events Say About Belief Overreaction in the Stock Market*. Tech. rep. National Bureau of Economic Research.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, et al. (2020). "Overreaction in macroeconomic expectations". In: *American Economic Review* 110.9, pp. 2748–2782.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, Matthew OBrien, et al. (2024). "Long-Term Expectations and Aggregate Fluctuations". In: *NBER Macroeconomics Annual* 38.1, pp. 311–347.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer (2024). "Belief overreaction and stock market puzzles". In: *Journal of Political Economy* 132.5, pp. 1450–1484.

Cai, Michael (2024). *Explaining the Macroeconomic Inertia Puzzle*. Tech. rep. Working Paper, Northwestern University.

Coibion, Olivier and Yuriy Gorodnichenko (2015). "Information rigidity and the expectations formation process: A simple framework and new facts". In: *American Economic Review* 105.8, pp. 2644–2678.

d'Arienzo, Daniele (2020). "Maturity increasing overreaction and bond market puzzles". In: *Available at SSRN 3733056*.

De la O, Ricardo and Sean Myers (2024). "Which Subjective Expectations Explain Asset Prices?" In: *The Review of Financial Studies* 37.6, pp. 1929–1978.

De Silva, Tim and David Thesmar (2024). "Noise in expectations: Evidence from analyst forecasts". In: *The Review of Financial Studies* 37.5, pp. 1494–1537.

Del Negro, Marco (2024). "Can Professional Forecasters Predict Uncertain Times?" In: *Federal Reserve Bank of New York, Liberty Street Economics*.

Driscoll, John C and Aart C Kraay (1998). "Consistent covariance matrix estimation with spatially dependent panel data". In: *Review of economics and statistics* 80.4, pp. 549–560.

Dupor, Bill and M Saif Mehkari (2014). "The analytics of technology news shocks". In: *Journal of Economic Theory* 153, pp. 392–427.

Eva, Kenneth and Fabian Winkler (2023). "A Comprehensive Empirical Evaluation of Biases in Expectation Formation". In.

Farmer, Leland E, Emi Nakamura, and Jón Steinsson (2024). "Learning about the long run". In: *Journal of Political Economy* 132.10, pp. 3334–3377.

Gormsen, Niels Joachim and Eben Lazarus (2023). "Duration-driven returns". In: *The Journal of Finance* 78.3, pp. 1393–1447.

H Décaire, Paul and Marius Guenzel (2023). "What Drives Very Long-Run Cash Flow Expectations?" In: *Marius, What Drives Very Long-Run Cash Flow Expectations*.

Kohlhas, Alexandre N and Ansgar Walther (2021). "Asymmetric attention". In: *American Economic Review* 111.9, pp. 2879–2925.

L'Huillier, Jean-Paul, Sanjay R Singh, and Donghoon Yoo (2024). "Incorporating diagnostic expectations into the New Keynesian framework". In: *Review of Economic Studies* 91.5, pp. 3013–3046.

McCarthy, Odhrain and Sebastian Hillenbrand (2021). "Heterogeneous beliefs and stock market fluctuations". In: *Available at SSRN 3944887*.

Mei, Pierfrancesco and Lingxuan Wu (2024). "Thinking about the Economy, Deep or Shallow?" In.

Patton, Andrew J and Allan Timmermann (2010). "Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion". In: *Journal of Monetary Economics* 57.7, pp. 803–820.

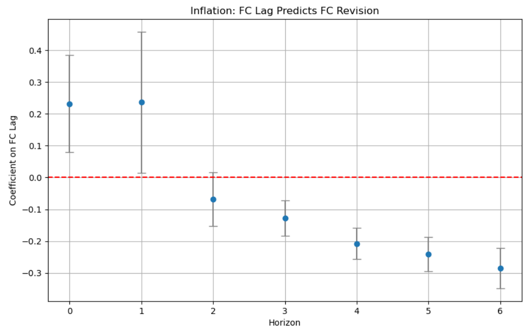
Sias, Richard, Laura Starks, and Harry Turtle (2024). "Long-term Beliefs and Financial Choices". In.

Sims, Christopher A (2003). "Implications of rational inattention". In: *Journal of monetary Economics* 50.3, pp. 665–690.

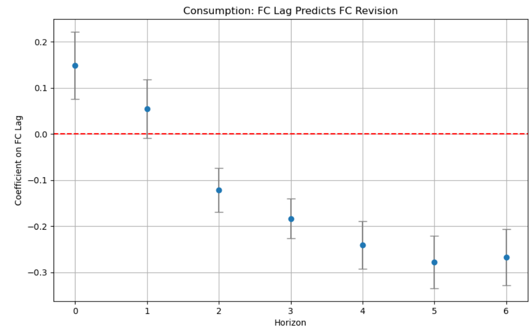
Winkler, Fabian (2020). “The role of learning for asset prices and business cycles”. In: *Journal of Monetary Economics* 114, pp. 42–58.
Woodford, Michael (2001). *Imperfect common knowledge and the effects of monetary policy*.

A Appendix

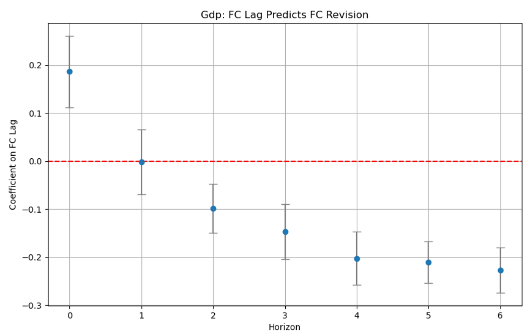
A.1 Additional Figures



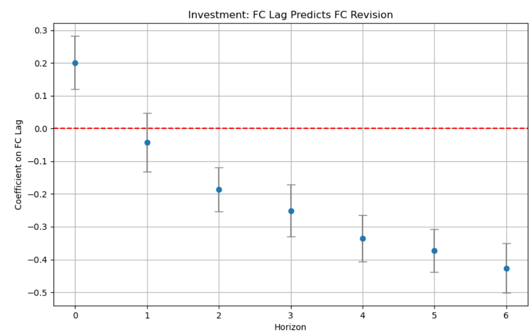
(a) Inflation



(b) Consumption

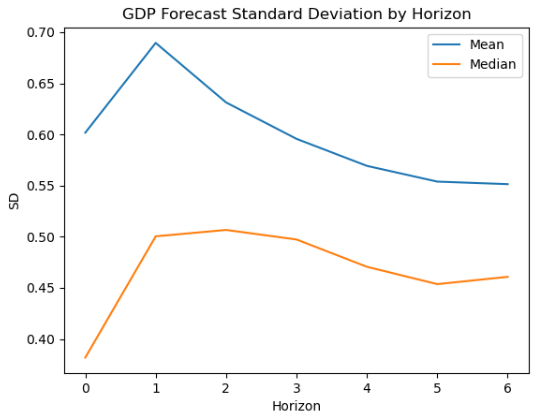


(c) GDP

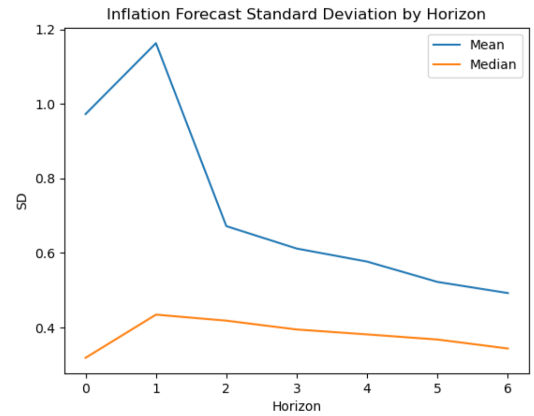


(d) Investment

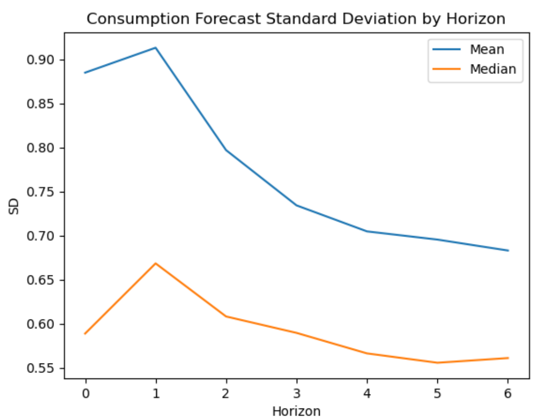
Figure 10: Regression coefficients from $\Delta \mathbb{E}_t[x_{i,t+j}] = \alpha + \beta \mathbb{E}_{t-1}[x_{i,t+j}] + \varepsilon_{i,t}$ for different forecast variables x.



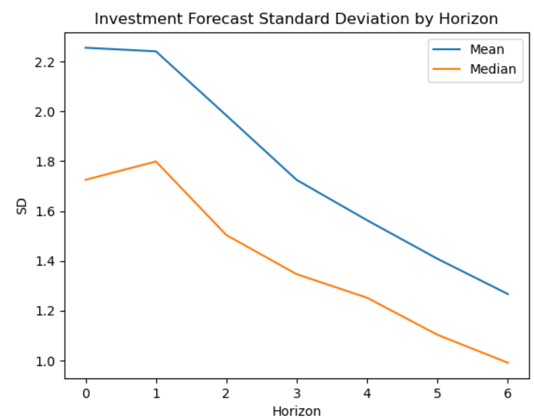
(a) GDP



(b) Inflation



(c) Consumption



(d) Investment

Figure 11: These plots show the mean and median standard deviation of forecasts by horizon and variable.

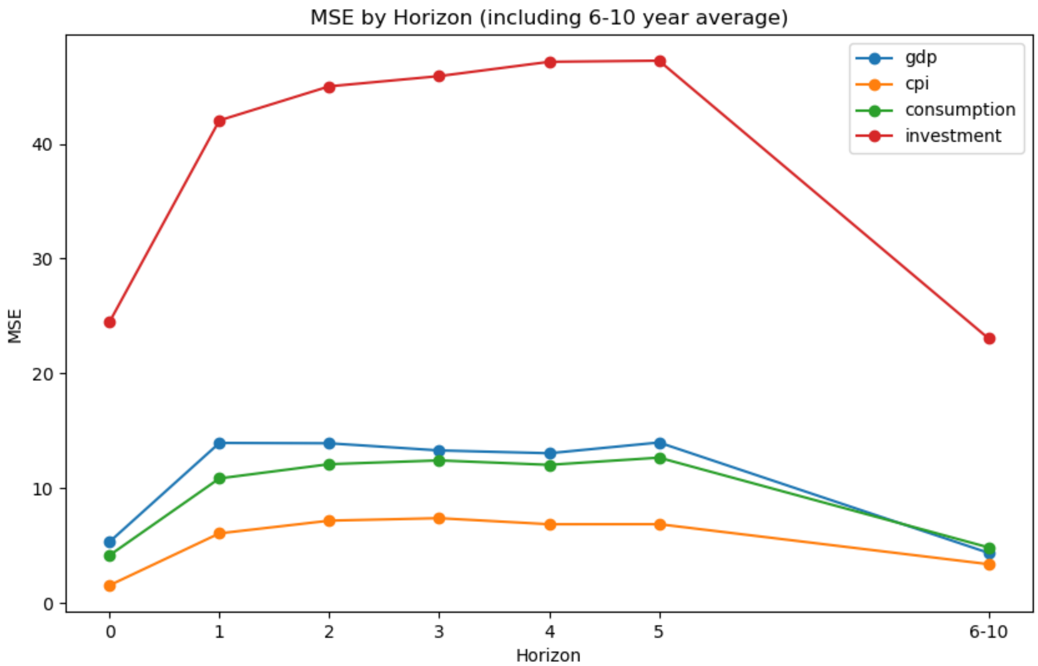


Figure 12: MSE by Horizon

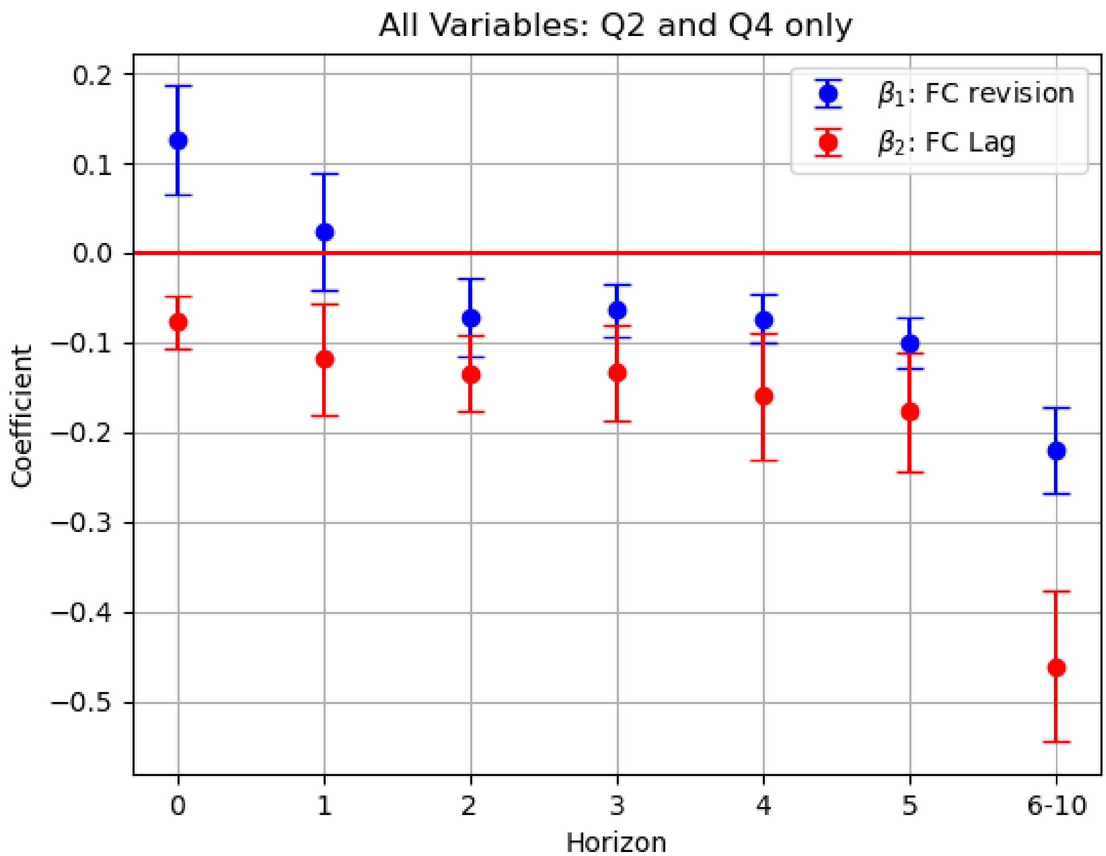


Figure 13: **Q2 and Q4 data only:** The figure plots the blue β_1 and red β_2 coefficients from the regression $e_{i,t+j} = \alpha + \beta_1 \Delta\, \mathbb{E}_t(x_{i,t+j}) + \beta_2\, \mathbb{E}_{t-1}(x_{i,t+j}) + f_{i,t,x} + \epsilon_{i,t}$, where $e_{i,t+j}$ are forecast errors, $\Delta\, \mathbb{E}_t(x_{i,t+j})$ are consecutive survey forecast revisions, and $\mathbb{E}_{t-1}(x_{i,t+j})$ is the previous survey's forecast. All forecasts are pooled and are z-scored with respect to variable, country, and horizon, with expanding window z-scores for the lagged forecast. All observations where the outcome was 2020 or later are removed. Country-variable fixed effects are included and standard errors are Driscoll-Kraay with country-variable groupings.

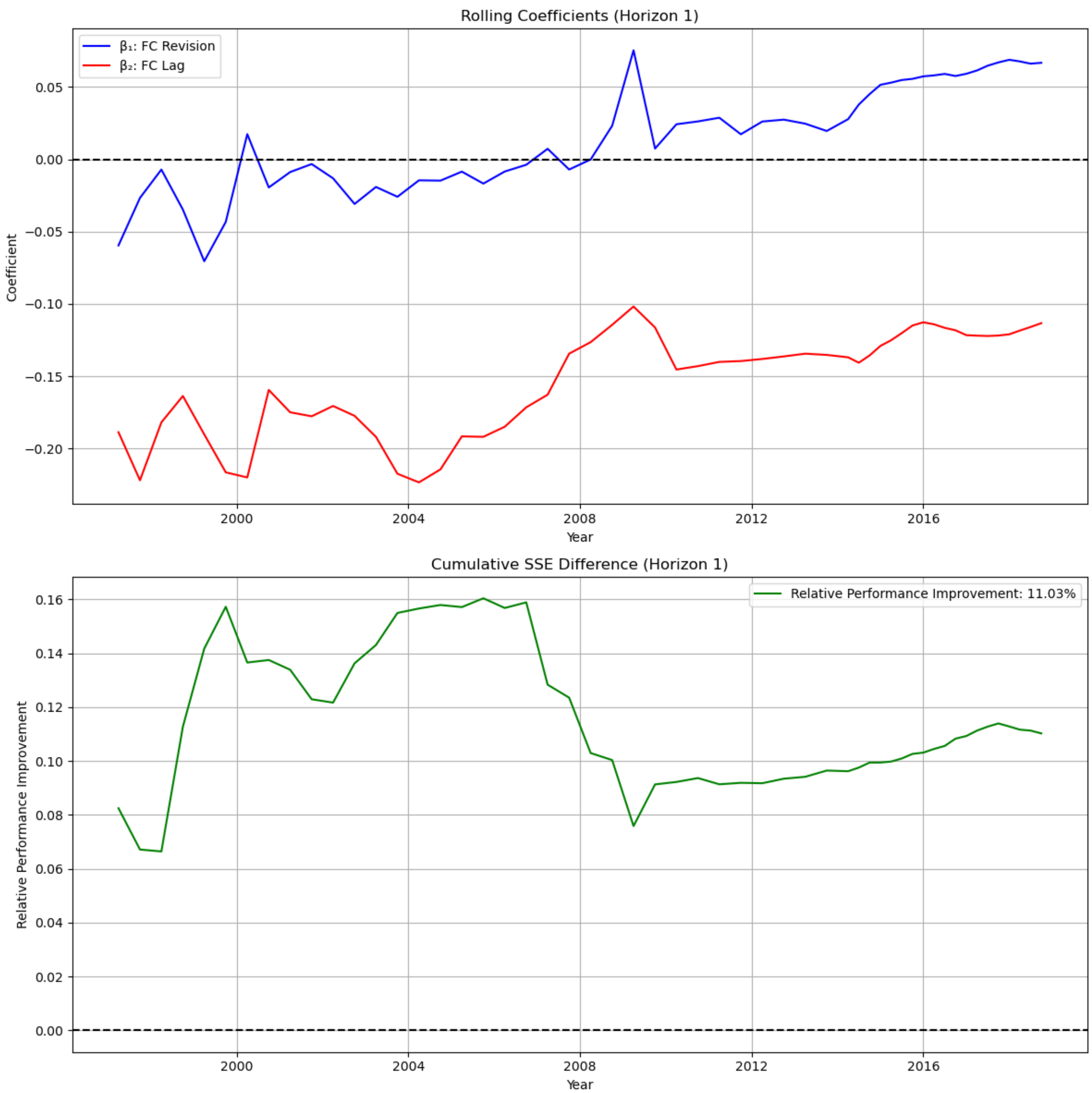


Figure 14: Horizon 1 OOS Forecasting

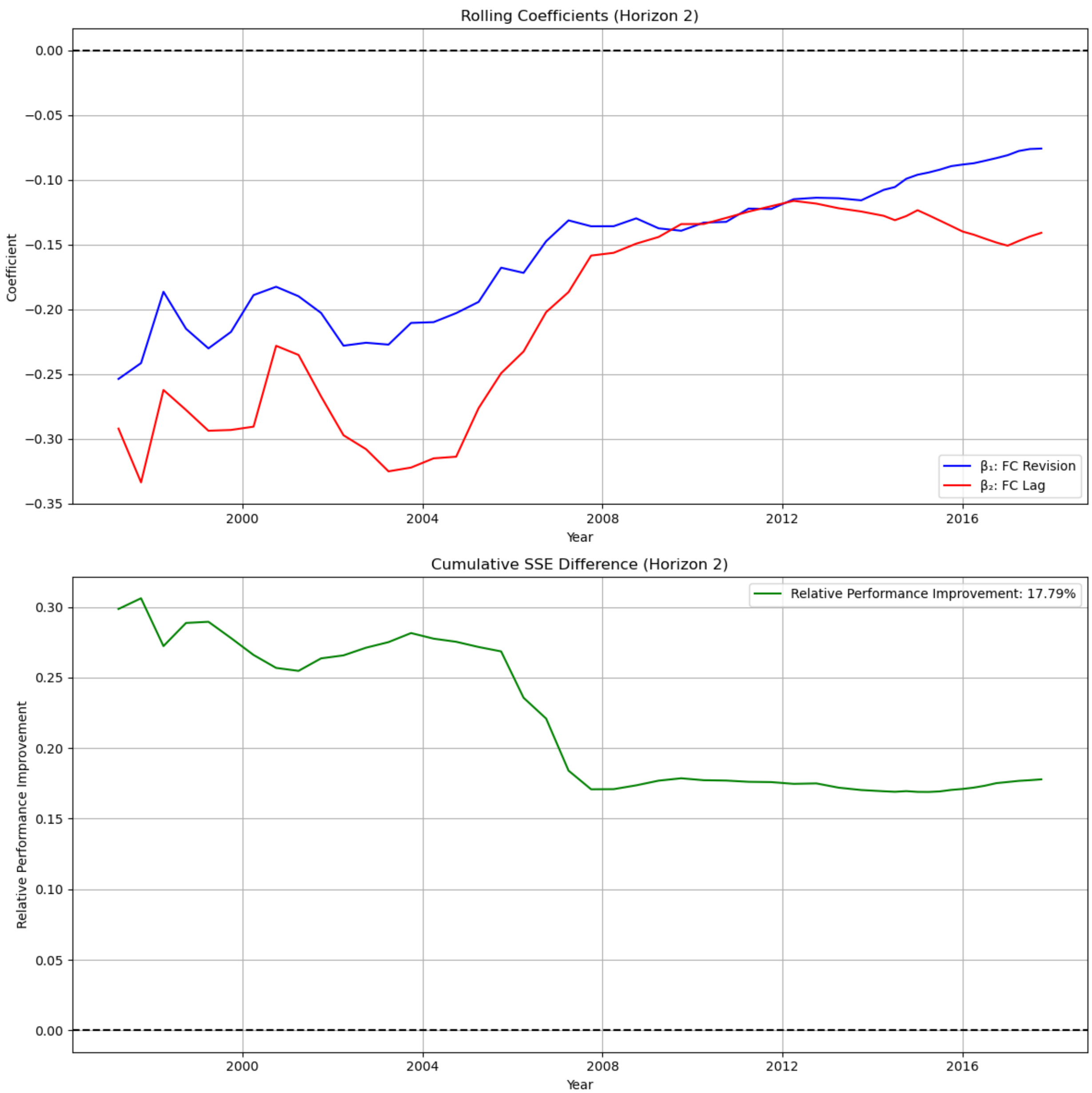


Figure 15: Horizon 2 OOS Forecasting

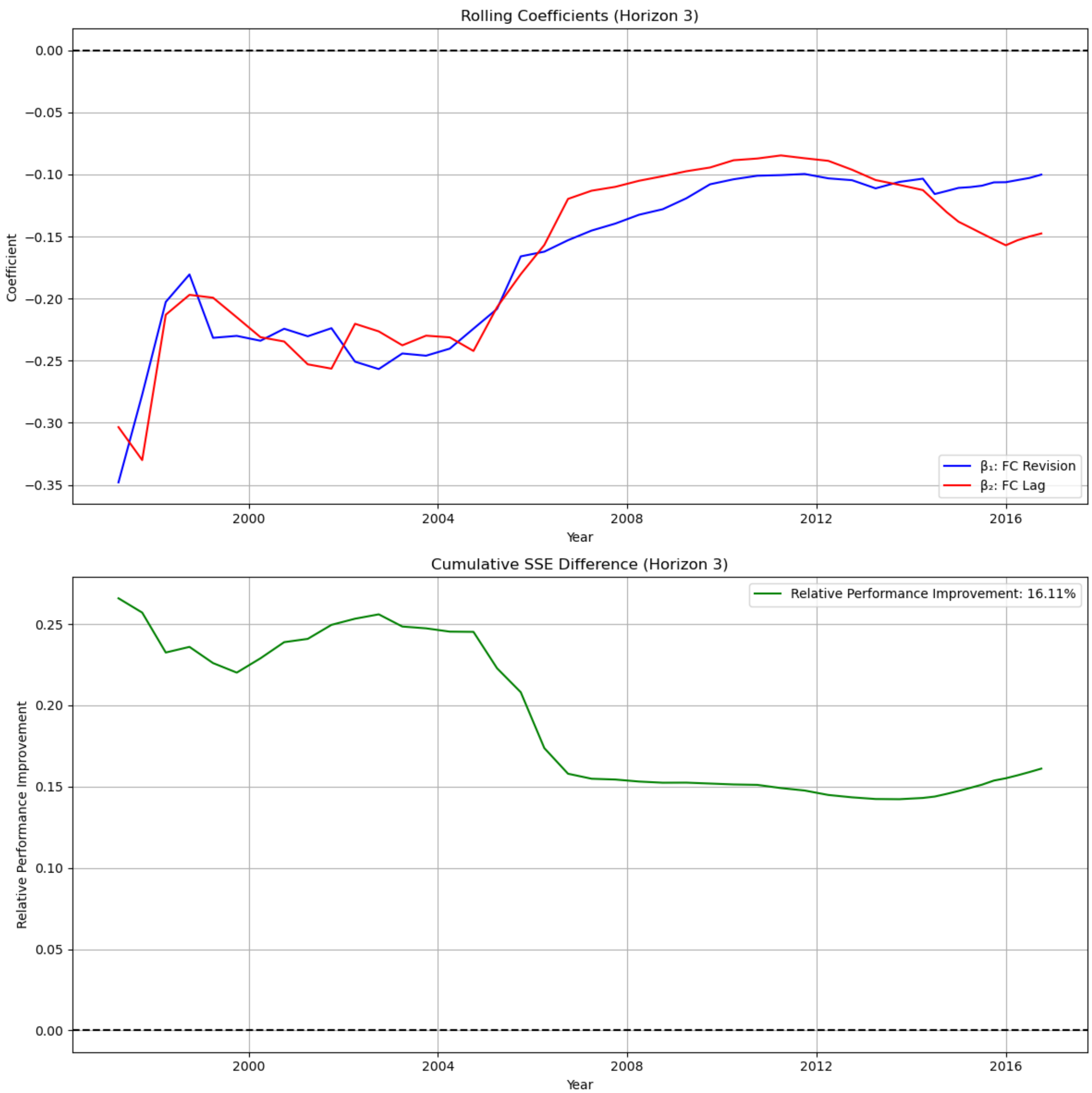


Figure 16: Horizon 3 OOS Forecasting

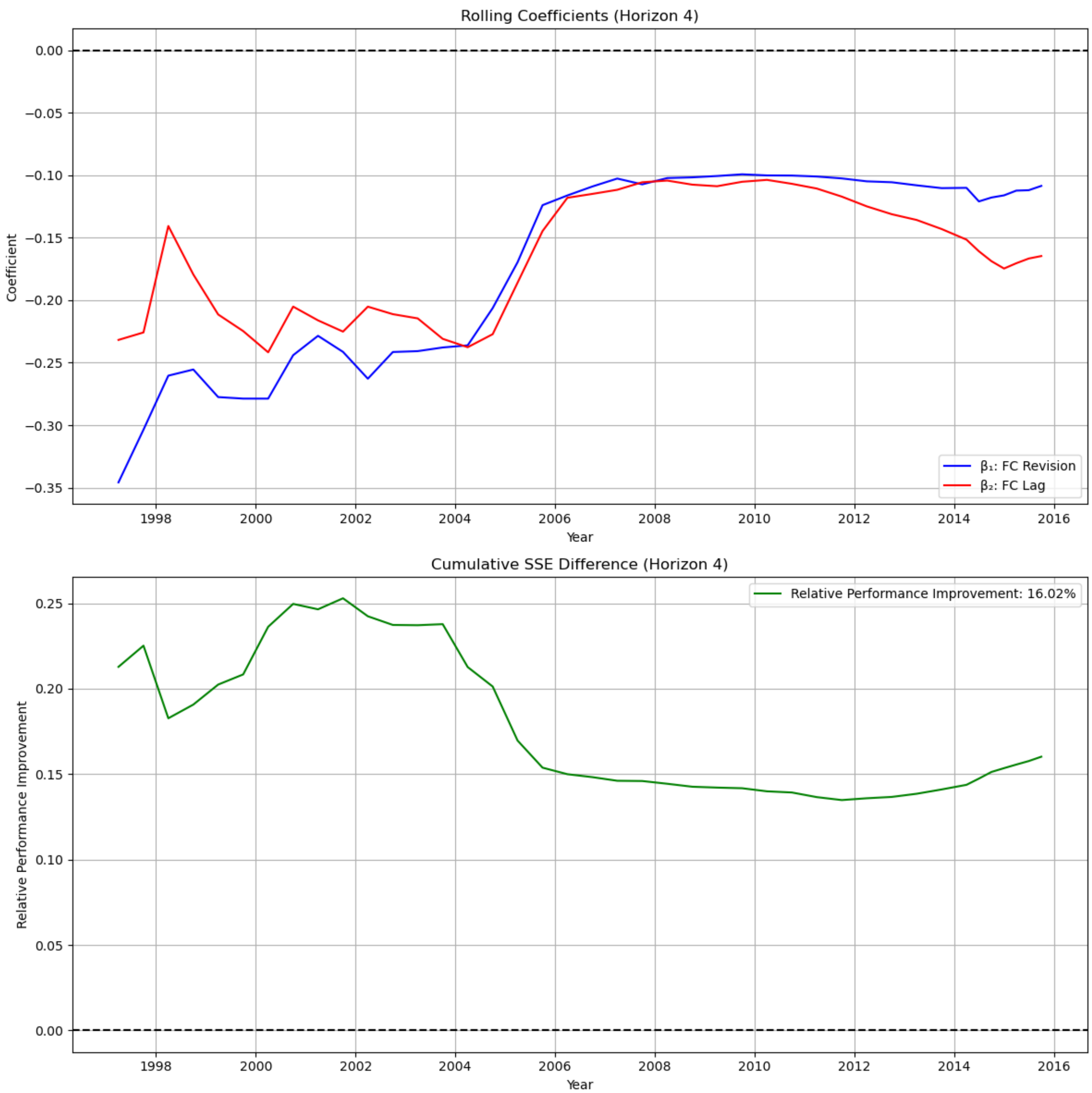


Figure 17: Horizon 4 OOS Forecasting

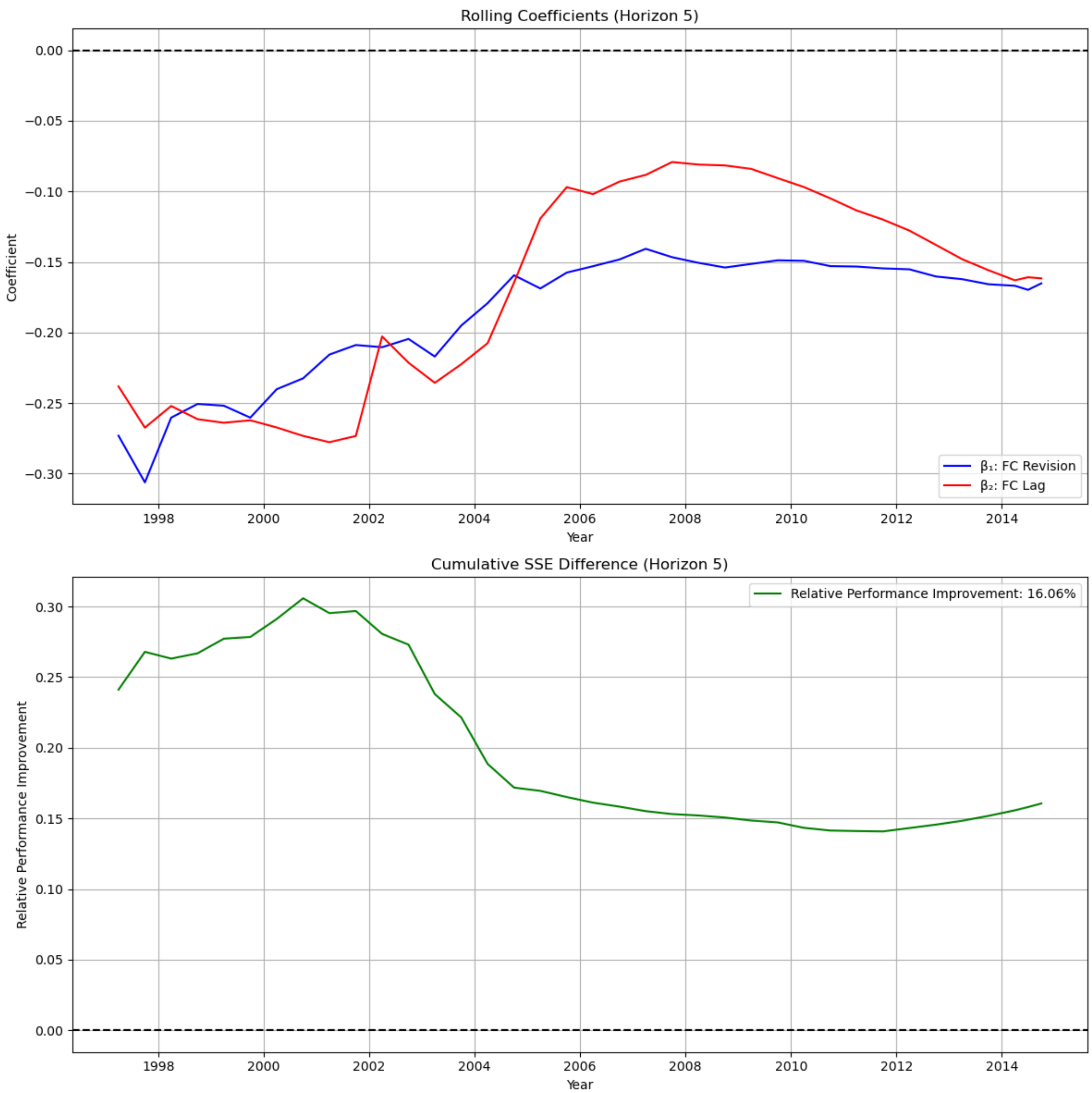
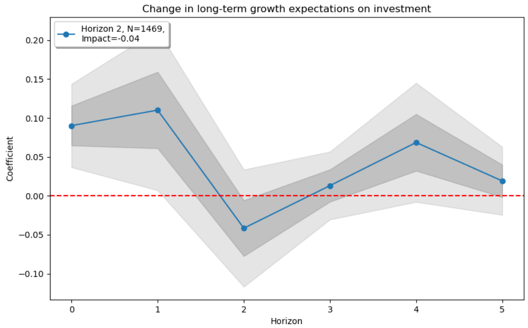
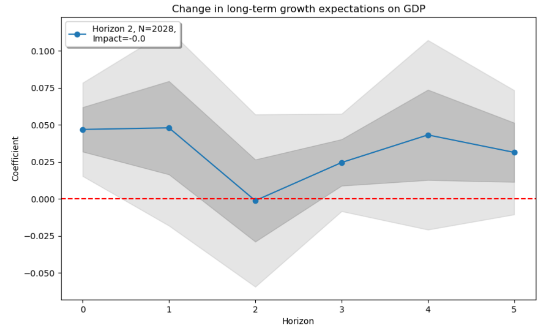


Figure 18: Horizon 5 OOS Forecasting

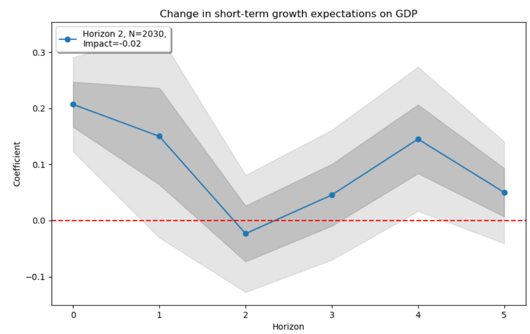


(a) Long-term expectations: Investment



(b) Long-term expectations: GDP

(c) Short-term expectations: Investment



(d) Short-term expectations: GDP

Figure 19: No control regressions: Regression coefficients from $y_{i,t+h} = \alpha + \beta_1 \Delta_1 \mathbb{E}_t(g_{i,t+j}) + \beta_2 \mathbb{E}_t(y_t) + f_{i,t} + \epsilon_{i,t}$ for different horizon-forecasts g. In this version of the regression, the only control variable is the current-year forecast $\mathbb{E}_t(y_t)$. The top-row uses the one-year change in six-to-ten year ahead average annual GDP growth expectations; the bottom-row uses the one-year change in average annual GDP growth expectations for the year of the forecast and the following two years. The dependent variable on the LHS column is investment, and on the RHS is GDP. Country-variable fixed effects are included and standard errors are Driscoll-Kraay with country-variable groupings.

A.2 Additional Tables

Table 6: Number of Observations by Country and Variable

Country	Variable			
	GDP	Inflation	Investment	Consumption
Albania	18	18	0	0
Argentina	79	78	79	79
Armenia	18	18	0	0
Australia	84	84	69	84
Austria	41	41	0	0
Azerbaijan	18	18	0	0

Continued on next page

Table 6 – continued from previous page

Country	Variable			
	GDP	Inflation	Investment	Consumption
Bangladesh	18	18	0	0
Belarus	18	18	0	0
Belgium	41	41	0	0
Bolivia	18	18	0	0
Bosnia & Herzegovina	18	18	0	0
Brazil	79	79	79	79
Bulgaria	51	51	51	51
Canada	86	86	69	86
Chile	79	79	79	79
China	76	76	45	0
Colombia	69	69	67	68
Costa Rica	18	18	0	0
Croatia	51	51	51	51
Cyprus	18	18	0	0
Czech Republic	70	70	67	70
Denmark	41	41	0	0
Dominican Republic	18	18	0	0
Ecuador	18	18	0	0
Egypt	22	22	0	0
El Salvador	18	18	0	0
Estonia	51	51	51	51
Euro zone	60	60	56	57
Finland	41	41	0	0
France	86	86	69	86
Georgia	18	18	0	0
Germany	86	86	69	86
Greece	41	41	0	0
Guatemala	18	18	0	0
Honduras	18	18	0	0
Hong Kong	73	73	69	72
Hungary	70	70	67	70
India	76	75	69	34
Indonesia	73	73	68	73
Ireland	41	41	0	0
Israel	22	22	0	0
Italy	86	86	69	86
Japan	85	85	68	85

Continued on next page

Table 6 – continued from previous page

Country	Variable			
	GDP	Inflation	Investment	Consumption
Kazakhstan	18	18	0	0
Kosovo	9	9	0	0
Latvia	51	51	51	51
Lithuania	51	51	51	51
Macedonia	18	18	0	0
Malaysia	76	76	69	75
Mexico	79	79	79	79
Moldova	18	18	0	0
Montenegro	8	8	0	0
Myanmar	18	18	0	0
Netherlands	76	76	68	76
New Zealand	76	76	69	76
Nicaragua	18	18	0	0
Nigeria	22	22	0	0
Norway	66	66	66	66
Pakistan	18	18	0	0
Panama	18	18	0	0
Paraguay	18	18	0	0
Peru	71	71	71	71
Philippines	48	48	48	47
Poland	70	70	67	70
Portugal	41	41	0	0
Romania	70	70	66	69
Russia	70	70	67	70
Saudi Arabia	22	22	0	0
Serbia	18	18	0	0
Singapore	76	76	67	74
Slovakia	70	70	67	70
Slovenia	51	51	51	51
South Africa	22	22	0	0
South Korea	76	76	69	76
Spain	76	76	69	76
Sri Lanka	18	18	0	0
Sweden	76	76	68	75
Switzerland	69	69	69	69
Taiwan	76	76	69	76
Thailand	74	73	65	70

Continued on next page

Table 6 – continued from previous page

Country	Variable			
	GDP	Inflation	Investment	Consumption
Turkey	69	69	66	69
Turkmenistan	18	18	0	0
Ukraine	70	70	48	70
United Kingdom	86	57	69	86
United States	86	86	69	86
Uruguay	18	18	0	0
Uzbekistan	18	18	0	0
Venezuela	75	72	74	74
Vietnam	18	18	15	15
Total	4240	4205	3018	3185

Table 7: USA: Return Predictability Regressions, no Covid

	Return Horizon			
	1-year	3-year	5-year	1-to-5-year
2-year avg GDP growth	-0.21	-0.21	-0.16	-0.12
	(0.19)	(0.14)	(0.14)	(0.13)
	[3.9%]	[3.2%]	[2.0%]	[1.1%]
10-year avg GDP growth	-0.44^{***}	-0.50^{***}	-0.25^{*}	-0.12
	(0.15)	(0.13)	(0.13)	(0.13)
	[16.3%]	[17.8%]	[5.0%]	[1.1%]
6-10 year avg GDP growth	-0.40^{***}	-0.53^{***}	-0.30^{***}	-0.18^{*}
	(0.11)	(0.09)	(0.09)	(0.09)
	[19.3%]	[29.4%]	[10.4%]	[3.7%]

Notes: Standard errors in parentheses, R^2 in square brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 8: 34 countries: Return Predictability Regressions, no Covid

	Return Horizon			
	1-year	3-year	5-year	1-to-5-year
2-year avg GDP growth	-0.36^{***}	-0.22^{*}	-0.28^{***}	-0.07
	(0.11)	(0.12)	(0.07)	(0.08)
	[8.2%]	[3.0%]	[6.0%]	[0.4%]
10-year avg GDP growth	-0.45^{***}	-0.29^{***}	-0.24^{***}	0.04
	(0.09)	(0.10)	(0.06)	(0.06)
	[10.5%]	[4.4%]	[3.7%]	[0.1%]
6-10 year avg GDP growth	-0.32^{***}	-0.24^{***}	-0.11^{*}	0.10
	(0.08)	(0.08)	(0.06)	(0.08)
	[5.3%]	[3.1%]	[0.7%]	[0.6%]

Notes: Standard errors in parentheses, R^2 in square brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Advanced economies: Return Predictability Regressions

	Return Horizon			
	1-year	3-year	5-year	1y4y
2-year avg GDP growth	-0.34^{***}	-0.21	-0.34^{***}	-0.14^{*}
	(0.07)	(0.14)	(0.08)	(0.08)
	[12.2%]	[3.2%]	[9.1%]	[1.5%]
10-year avg GDP growth	-0.25^{**}	-0.19	-0.28^{**}	-0.07
	(0.12)	(0.16)	(0.12)	(0.15)
	[5.8%]	[3.0%]	[5.9%]	[0.4%]
6-10 year avg GDP growth	-0.08	-0.14	-0.16	$6.75e-03$
	(0.14)	(0.17)	(0.15)	(0.18)
	[0.5%]	[1.5%]	[1.6%]	[0.0%]

Notes: Standard errors in parentheses, R^2 in square brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 10: Emerging markets: Return Predictability Regressions

	Return Horizon			
	1-year	3-year	5-year	1y4y
2-year avg GDP growth	-0.33^{***}	-0.22^{**}	-0.23^{***}	$5.91e-03$
	(0.07)	(0.10)	(0.07)	(0.10)
	[11.4%]	[4.3%]	[4.8%]	[0.0%]
10-year avg GDP growth	-0.20^{***}	-0.14	-0.15	0.03
	(0.08)	(0.11)	(0.11)	(0.13)
	[4.0%]	[1.7%]	[2.0%]	[0.1%]
6-10 year avg GDP growth	-0.03	-0.02	-0.02	0.11
	(0.10)	(0.13)	(0.14)	(0.15)
	[0.1%]	[0.0%]	[0.0%]	[0.9%]

Notes: Standard errors in parentheses, R^2 in square brackets. *** p<0.01, ** p<0.05, * p<0.1